DESIGN AND DEVELOPMENT OF AN INTERNET COLLABORATION SYSTEM TO SUPPORT GIS DATA PRODUCTION MANAGEMENT

SONGNIAN LI

May 2003

TECHNICAL REPORT NO. 219

DESIGN AND DEVELOPMENT OF AN INTERNET COLLABORATION SYSTEM TO SUPPORT DISTRIBUTED GIS DATA PRODUCTION MANAGEMENT

Songnian Li

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

May 2003

PREFACE

This technical report is an unedited reproduction of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geodesy and Geomatics Engineering, May 2002. The research was supervised by Dr. David Coleman. Support was provided by the GEOIDE Network of Centres of Excellence, WaterMark Industries, Inc., and Service New Brunswick.

Li, Songnian (2003). *Design and Development of an Internet Collaboration System to Support Distributed GIS Data Production Management.* Ph.D. dissertation, Department of Geodesy and Geomatics Engineering Technical Report No. 219, University of New Brunswick, Fredericton, New Brunswick, Canada, 339 pp.

ABSTRACT

Computer supported cooperative work (CSCW) technologies (e.g., groupware) are increasingly used to support distributed collaborative production in many areas, especially with the rapid growth of the Internet and broadband network technologies. Empirical studies of existing practices and preliminary research at the University of New Brunswick have already demonstrated the potential of applying these technologies in GIS data production work environments.

In order to achieve the improved efficiency and productivity, group work processes must be properly addressed and the design of computer systems to support them must be thoroughly studied. These usually vary from one area to another depending on a comprehensive understanding of the surroundings of the needed computer system through modeling its business environments (e.g., production processes and user requirements). However, no research using a formal modeling approach to characterize current distributed GIS data production projects has been conducted so far. In addition, little documented evidence has indicated that groupware tools other than electronic mail and FTP have been used in existing geomatics production environments.

This thesis describes the testing of the hypothesis that CSCW tools, when integrated with existing GIS installations, can provide significant efficiency and productivity improvement of distributed GIS data production operations and management over the Internet infrastructure. The reported research developed an Internet collaboration model

as the framework for implementing the required collaborative workspace. A research prototype system was then developed and tested within a data quality control process under laboratory conditions. The testing involved 1895 data files and its results are analyzed comparing with the existing manual and paper-based GIS data production projects.

The research results indicate that: (1) the GIS, Internet, groupware, and data warehousing principles can be possibly brought together and applied to real-world GIS data production environments to provide better solutions; (2) CSCW-based technologies such as workflow can be used to effectively facilitate collaborative GIS data production in a distributed work environment if a *"sound"* model is in place; and (2) the potential time saving obtained using the proposed approach is at least 60% of the total elapsed time of a data QC process using traditional production approach.

ACKNOWLEDGMENTS

Over the last four years, this thesis was completed with the efforts, guidance, patience and encouragement of many persons to whom I want to express my sincere thanks. While it would be impossible to acknowledge all of them, it is my sincere pleasure to extend special thanks to the following individuals:

First, my very special thanks go to my supervisor, Dr. David Coleman, for his continuous and patient criticism, concern and valuable scientific remarks and guidance, especially his understanding, friendship, and effort in keeping my research on the right track. I would also like to extend my sincere thanks to Dr. Brad Nickerson, Dr. Y. C. Lee, Members of my supervisory board, and Dr. Yvan Bedard, a professor at Laval University, for spending their valuable time providing me advice and guiding me through the problems I had.

Second, I thank Mr. Mark E. Doucette at WaterMark Industries, Inc. and Mr. David Finley at Service New Brunswick in Fredericton, New Brunswick, for their insights on actual production projects and support by providing me access to all kinds of information materials. I am grateful to Dr. Weichang Du, a professor of the School of Computer Science at the University of New Brunswick, for his helpful advices on UML-based process models as well as allowing me full access to the needed software tools. There were many people who provided me with their valuable technical support during my thesis research, among whom I would like to specially thank Mr. David Fraser for his patient support and friendship.

Finally, my special thanks to my wife, Jun, and my daughter, Xuanyi, who have sacrificed so much for this adventure and whose complete support and encouragement have greatly helped make this thesis a success, and to whom I owe too much to return.

The financial support provided by Dr. David J. Coleman and the GEOIDE Network of Centres of Excellence has made this research possible and it has been greatly appreciated.

TABLE of CONTENT

List of Tables

List of Figures

List of Abbreviations

ANSI:	American National Standards Institute
API:	Application Programming Interface
ASP:	Active Server Page
BFILE:	Binary File (Binary data stored in an external file)
BLOB:	Binary Large Object (an Oracle data type)
BMP:	Bitmap
CAD:	Computer-Aided Design
CAM:	Computer-Aided Manufacturing
CASE:	Computer-Aided Software Engineering
CBP:	Contract-Based Production
CGI:	Common Gateway Interface
CLOB:	Character Large Object (an Oracle data type)
CODEC:	Compression and Decompression
COM:	Common Object Model
CORBA:	Common Object Request Broker Architecture
CSCW:	Computer-Supported Cooperative Work
CW:	Collaborative Workspace
DBMS:	Database Management System
DLL:	Dynamic Link Library
DOS:	Disk Operating System
ESRI:	Environment Source Research Institute
ETB:	Enhanced Topographic Base
FGDC:	Federal Geospatial Data Committee
FTP:	File Transfer Protocol
GIF:	Graphics Interchange Format
GIS:	Geographical Information System
GUI:	Graphical User Interface
HTML:	HyperText Markup Language
HTTP:	HyperText Transfer Protocol
IDE:	Integrated Development Environment
IDEF:	Integrated computer-aided manufacturing Definition language
IICW:	Internet-based Integrated Collaborative Workspace
IIS:	Internet Information Server
IMS:	Internet Map Server
IRC:	Internet Relay Chat
ISO:	International Organization for Standardization
ISP:	Internet Server Provider
JPEG:	Joint Photographic Expert Group
LAN:	Local Area Network
LCD:	Liquid Crystal Display
NCLOB:	National Character Large Object (an Oracle data type)
NNTP:	Network News Transfer Protocol
OGC:	OpenGIS Consortium

OLAP:	Online Analytical Analysis Process
OLE:	Object Linking and Embedding
OLTP:	Online Transaction Process
OM:	Organizational Model
OMG:	Object Management Group
ORB:	Object Request Broker
OSI-TP:	Open Systems Interconnection distributed Transaction Processing
POP:	Post Office Protocol
QC:	Quality Control
RTC:	Real-Time Chatting
SMTP:	Simple Mail Transfer Protocol
SNB:	Service New Brunswick
SQL:	Structured Query Language
TIFF:	Tag Image File Format
TRM:	Technology Road Map
UML:	Unified Modeling Language
UNB:	University of New Brunswick
VB:	Visual Basic
WAN:	Wide Area Network
WfMC:	Workflow Management Coalition
WFS:	Workflow System
WMFS:	Workflow Management Facilities Specification
WMS:	Workflow Management System
WPDL:	Workflow Process Definition Language
WRM:	Workflow Reference Model
WWW:	World Wide Web
WYSIWIS:	What You See Is What I See
XML:	Extensible Markup Language

Chapter 1 INTRODUCTION

The Internet-based **computer supported cooperative work** (CSCW) concepts and related technologies have been increasingly applied to support distributed production in many areas, permitting collaboration among groups of people located at various geographical locations. GIS data production is among these potential application areas, particularly when its work environments are widely dispersed and its projects are regulated by data production contracts. Applying CSCW to existing GIS data production environments benefits both data production processes – where process activities may need to be coordinated or automated – and project management requiring better communications and project information sharing.

The adoption of CSCW concepts and technologies to facilitate distributed spatial decision making processes at various levels has been focusing on integrating CSCW and GIS for the last few years. Some simple forms of informal use of CSCW, such as email and FTP of GIS data files, have also been found in GIS data production work environments. Few GIS/mapping firms, however, have formally adopted strategies of applying CSCW systematically in support of their data production workflow and related project management. This is mainly because of the lack of formal understandings of collaboration characteristics of GIS data production projects, high costs of required supporting infrastructure, and social/technical uncertainties of CSCW software implementation, together with the concern of potential reengineering of existing business processes.

This thesis will model collaboration characteristics of the existing GIS data production projects from both process and supporting CSCW-based software system perspectives. The other goal of the thesis research is to determine, based on the modeling outputs, whether the Internet-based CSCW technologies when integrated into existing GIS installations are appropriate for improving the efficiency and productivity of distributed GIS data production projects through a prototype supporting system.

1.1 Research Context

The future of the Canadian geomatics industry is driven by dynamically expanding domestic and international markets, and the convergence of continuously developing technologies all in the light of the emerging knowledge-based economy.

[Technology Road Map, Industry Canada, 1998]

Similar to the Canadian geomatics industry in general, the GIS and mapping organizations have been driven by emerging technologies and expanding markets, shifting from traditional data production techniques to digital mapping and geographical information systems [Coleman and McLaughlin, 1988]. The corresponding business goals have shifted from providing not only pure high-quality data products, but also total satisfaction to customers [Manheim, 1998]. This requires both a change of underlining work environments and the integration of new supporting technologies.

1.1.1 Emergence of Distributed GIS Data Production

The early GIS data production industry in Canada was dominated by a few large surveying and mapping companies. Each company had to maintain in-house expertise and software/hardware facilities required by the data production process, which might contain procedures from field data collection/conversion and original production through inspection, correction, initial distribution and recurrent updating. The whole process was traditionally designed in an environment where work was completed in a single location and complete sets of data materials were shipped in bulk from one unit to another, usually in the same building [Coleman and Brooks, 1996]. This "centralized" production model not only lacked an independent data quality inspector, but also contained many potential bottlenecks that might delay the final delivery of data products to customers [Coleman and McLaughlin, 1988]. One major bottleneck was that the company might not be able to process a large number of data files at the same time due to its limited production throughput.

This small number of large organizations has been gradually replaced by a collection of many small GIS/mapping firms due to the trend of deregulation, outsourcing and downsizing of organizations as well as an increasing demand for quickly-delivered and high-quality GIS data. Given their relatively limited resources, these small firms can only keep competence in a few market niches. At the same time, it has been increasingly difficult for the government mapping agencies to maintain in-house expertise and necessary resources to account for their whole responsibilities of providing timely and high-quality GIS and mapping data. It has, therefore, increasingly become common practices for government agencies to outsource GIS data production tasks and for one or more geomatics firms to contract the whole or part of the data production responsibility on a program or project basis [Sebert, 1989; Sabourin, 1994; Coleman and Brooks, 1996; Castonguay, 1999].

This contract-based production (CBP) model utilizes consortia of data production companies and government agencies. One of the notable advantages of the CBP model is the (occasional) inclusion of a separate contractor serving as data quality inspector whose responsibility is to assure all produced data satisfies specified quality standards on behalf of the customer. The contract production process involves procedures being done from different locations, either across a city or across the country. Data may be collected or produced by contractors in several different centers, checked by inspection staff in some single location, returned to the supplier(s) for correction or verification, submitted to a customer or client residing somewhere else, and ultimately distributed to third-party suppliers or end-users in various centers [Coleman, 1994b]. The materials involved in the production process (including the digital data itself) are typically shipped among production participants by courier or, most recently, via the Internet using file transfer protocols (FTP).

1.1.2 Problems Concerned with the CBP Model

GIS data production projects organized based on the CBP model may be conducted successfully in terms of quality, delivery and improved productivity if all project related data materials are transferred in a timely manner, project information and issues (e.g., specification ambiguities) are well communicated, and all production processes are under control. Due to the multi-party and multi-location nature of the CBP model, however, these assumptions are often compromised by the following issues identified from the existing data production projects:

1. Time consuming or unreliable approaches of transferring data and other production-related materials among project participants;
2. Lack of efficient communications channels for distributed participants to discuss technical and managerial problems;
3. Difficulties in efficiently and dynamically controlling project progress and tracking status of data files and production process activities; and
4. Lack of efficient management of the production-related reports, comments and correspondences, which may be used as "project memory" for future reference.

While the on-time delivery of customer specified digital data may be ensured by resolving the above issues, there are other issues which also concern the project management. These issues relate to the fact that GIS data production sometimes involves more than one round of data file submission and resubmission due to possible failures of some data files in passing quality control (QC) inspections. Therefore, CBP-based project managers often face the following three important challenges:

1. How to manage multiple production contractors including sub contractors and data quality control inspectors involved?
2. How to manage multiple contractor submissions and resubmission to ensure that all involved data files are kept with proper identifications, e.g. version controlled? and
3. How to reduce duplications of effort, wasted space for data storage, and other wasted resources such as office suppliers?

Although all above issues and challenges must be addressed to ensure the success of contract data production projects, the biggest concerns from existing practices are: (1) the communications of the project status, problems and progress; and (2) the transfer of large digital data files consisting of very large datasets – often up to 85 megabytes for graphical files and 200 megabytes for image files. Project deliveries could be delayed due to inefficient communications of specification changes or long time and risk of data loss associated with transferring data files over current postal services. More importantly, the different interpretations of the specifications and lack of common understandings of quality requirements may result in more data files failing to pass data quality inspection processes.

Measures have already been taken in various data production projects to minimize the effects of the above issues and maximize the project management capacities in handling challenges being faced. As discussed in Chapter 2, however, none of these measures provides efficient solutions to the above problems. This thesis addresses the above problems to a certain extent by integrating CSCW technologies and the Internet within existing GIS data production environments, focusing on production process control and project management tasks.

1.1.3 Internet-based CSCW as Potential Solutions

The recent progress of computer-supported cooperative work (CSCW) in supporting distributed applications has been greatly fostered by the rapid development of high-speed computer networks and related technologies. Among these developments, the Internet and Internet-based client/server technologies have already played important roles in the application areas involving high-volume materials, such as distributed production.

On one hand, CSCW has brought with it investigations into interactions of individuals in a group through coordinated actions, shared workspaces and group awareness of work related information and presence of other group members. Groupware – the technology implementation of CSCW – has been extensively developed and, until recently, many groupware software packages have been redesigned to support Internet or Web based applications (see Section 2.2). The major functions these groupware packages provide include:

1. Informal and formal (in the sense of whether electronic signature is required or not) sharing and dissemination of many types of information;

2. Collaborative viewing, editing and manipulating of shared objects (in both asynchronous and synchronous modes);
3. Real-time communications (conferencing and electronic meeting); and
4. Workflow process management.

On the other hand, the Internet, empowered by standard network protocols and software, allows "mass" access of information and services available across various computing platforms and transfer of electronic data using FTP. With the development of the client/server computing model, computer-based information and services may be hosted on one or more centralized servers and accessed via widely dispersed workstations and personal computers (PC) through standard client interface software such as web browsers or dedicated client software.

These developments have already enhanced GIS related applications through the growing use of the Internet to handle spatial information. Examples include group-based spatial decision-making with CSCW technologies, hosting web sites and sending emails, and in some cases transferring digital files using FTP programs between GIS data production companies. Introduction of integrated Internet-based CSCW solutions with existing GIS installations may hold great potential for streamlining some or all of the above issues faced by CBP-based GIS data production projects. In addition, they may open many new opportunities for managing these projects more efficiently and effectively.

Integration of CSCW technologies into GIS data production work environments forms the concept of "collaborative GIS data production" which was initially discussed in geomatics in 1994 [Coleman, 1994b]. This concept implies a distributed work environment that encompasses people, organizations, computer networks, production processes, supporting technologies, and suitable management policies. While all of these constituent components are important for the overall implementation of a collaborative data production environment, it is the virtual environment – called "collaborative workspace" in the context of this thesis – which provides required technological support.

This distributed collaborative workspace can be described as a shared project workspace that – using shared databases, a set of collaboration functions, and a collection of digital data files – permits definition of group members' roles, project status reporting and tracking, and gateways to electronic mail and other sources of data. Such a workspace should also permit the organization of correspondence, comments, reports, and other documents associated with a project or product and should support the management of multiple versions of objects [Coleman, 1999b]. From the users' perspective, the workspace should provide access to all project-related information they are authorized to share, a personal worklist containing all tasks assigned to them, and collaboration functions they can use to collaborate and communicate with other participants.

The development of such a collaborative workspace largely depends on a comprehensive understanding of the collaboration characteristics of the CBP-based GIS data production practices. In addition, hardware, software and operational constraints to online project management procedures have to be well identified to facilitate the implementation of such collaborative workspaces. More importantly, GIS data production tasks involved in the workspace have to be investigated with respect to the level of network connection speed required. To date, however, research effort in this respect has been limited to conducting feasibility studies of CSCW, preliminary investigation of existing data production projects, and evaluation of currently available groupware systems [Finley, 1997; Boettcher, 1999; Coleman and Li, 1999].

The software implementation of such a collaborative workspace may be relatively easy to realize to provide required functional support. Its ultimate adoption to support distributed GIS data production projects will remain uncertain unless it can be proven that the proposed solutions provide shortened "production float", streamlined or alternative data production procedures, and improved communications channels. In other words, the improvement of the project efficiency and productivity potentially offered must be verified and promoted.

1.2 Research Hypothesis and Objectives

Based on the previous discussion, a review of state-of-the-art developments in the similar areas, the information gained through evaluating currently available groupware systems and GIS software, and the results of the preliminary research conducted at the University of New Brunswick (UNB), it is hypothesized that:

The collaborative workspace implemented based on the proposed model can provide sound performance for improving distributed GIS production operations and management over Internet-based network infrastructure.

Understanding what is meant by "*performance for improving*" is very important for the testing of this hypothesis. Given the fact this research deals with actual production systems, factors such as cost effectiveness, time efficiency, service level (availability), and others will be carefully studied to measure the performance improvement. In each case, appropriate metrics will be defined and values which characterize "*sound*" levels of performance will be proposed and justified.

Performance will be tested by comparing differences in the above factors between the prototype and selective manual systems, such as decreases of time span and increases in cost-effectiveness. In this regard, the measurements based on the selected manual system are assumed as base values to allow identified differences to demonstrate how "*sound*" the proposed workspace improves GIS production operations and management over the Internet-based network infrastructure.

To test this hypothesis, a suitable approach is needed to design representative combinations of typical procedures with highly frequent activities executed by key roles based on associated rules as bases for measuring differences. To this extent, measurable units should be well isolated and the effecting factors well controlled. It is also necessary to design and develop a research prototype of the proposed collaborative workspace based on the established understanding of practical GIS data production environments. With this in mind, the objectives of the proposed research include:

(1) To design an Internet-based geomatics production collaboration model which includes: (a) a workflow model; (b) a workspace architecture; and (c) an implementation framework to facilitate the development and implementation of the collaborative workspace;

(2) To develop a workable collaborative workspace prototype based on the designed model by integrating functional modules or software components provided by existing GIS and groupware packages; and

(3) To test and/or simulate the performance measures of this prototyped collaborative workspace and refine the collaboration model that is feasible for applications within a real-world geomatics production environment.

1.3 Research Methodology

The research started with a reality check, focusing on a literature review, an investigation of selected groupware tools, and an examination of existing GIS data production projects and their work environment. The literature review searched relevant research papers both in CSCW domain and other domains involving applications of CSCW technologies (or groupware tools) such as manufacturing and construction fields. An evaluation of the capabilities provided by existing groupware tools was conducted based on selected packages only which were downloaded from the Internet free of charge. The existing production processes were examined in detail, relevant documents (e.g., work sheets, sample reports, official correspondence, and technical specifications) were collected, and participants and tasks were identified for supporting further analysis and modeling. The purpose of the above work was to:

1. build necessary terminology and knowledge for the research;
2. identify problems in existing production process and objects of workflow;
3. refine the collaboration requirements of geomatics production; and
4. identify the gaps between what is required and what is available

Based on the outcomes from the reality check, the collaboration model was then designed and developed. Both Unified Modeling Language (UML) and Workflow Process Definition Language (WPDL) were used to model the distributed GIS data production workflows and to develop an architectural model of the collaborative workspace. In developing workflow models, the results from examining existing GIS data production projects and their work environment played an essential role. While the architecture development, to a large extent, depended on the collaboration requirements defined in the first phase, the workflow requirements were also carefully considered to make sure that the architecture supports the integration of workflow management components.

Since the overall approach adopted in this research was a participatory, prototype-oriented paradigm (as discussed in Section 4.1), the next step was to develop a prototype collaborative workspace. At this stage, some programming work was done using Visual

Basic, PL/SQL and Scripting languages to realize the integration between different components of the workspace and to provide suitable user-friendly interfaces over the web. The prototype was used to evaluate the collaboration model and to develop an implementation framework which provides a set of guidelines to facilitate the implementation of such a collaborative workspace. Although the approach emphasized a define-design-build-evaluate process, the usability and functionality of prototype was not tested using rigorous software engineering testing methods such as alpha and beta testing.

Finally, the collaboration model was verified and the collaboration system was tested under controlled laboratory conditions. A topological structuring of the digital topographical database project (see Section 6.1 for a description) was selected for the required performance testing. The performance analysis focused on the time and cost savings obtained through the collaborative system to validate the research hypothesis.

Figure 1-1 presents an overview of the research methodology described above. Four major steps are illustrated with sub tasks in each step.

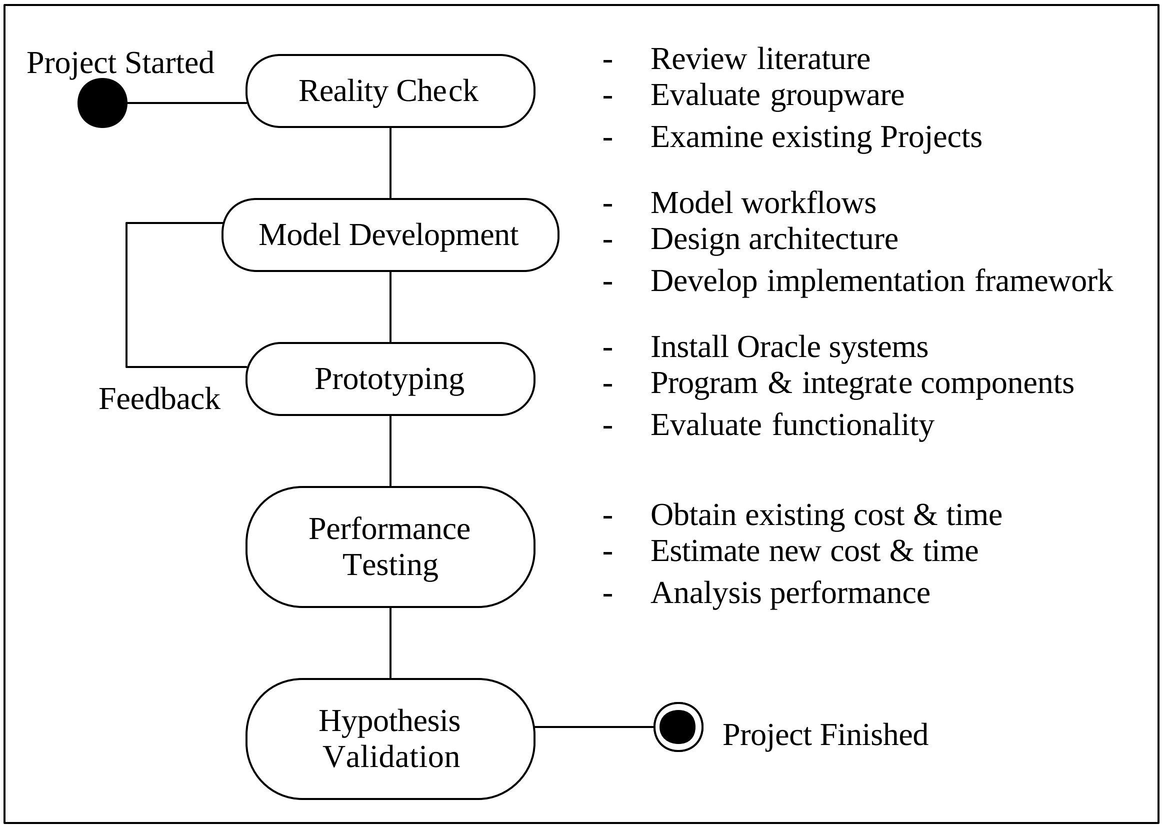


Figure 1-1 An overview of the research methodology

1.4 Significance and Contributions

The outputs of this research project could potentially have a significant impact on the way GIS data production firms perform their data production, quality control and updating activities because the research addresses alternative solutions to the problems existing within current GIS data production practices. With more and more GIS/mapping organizations outsource their data production related responsibilities to third-party companies (mostly in the private sector), the CBP model will play an increasing role in managing and operating related contracts and/or projects. If proven successful, the proposed approach will definitely provide a viable alternative to accommodate this trend.

Data collection and maintenance often comprises 60 – 80% of the total cost of an operational GIS [Bernhardsen, 1999]. In addition, numerous GIS-related applications rely on data delivered in a timely manner. Source data can now be obtained very fast from various data acquisition channels. However, the production, quality control and updating procedures are not as fast as acquiring source data due to rigorous and time-consuming procedures involved, lack of proper human and technical resources, among others. Any solutions that could improve the efficiency and productivity of these production processes will make significant contributions to the field, including the one developed in this thesis.

This research makes the following contributions to the overall field of knowledge in this area:

- The thesis presents the first effort of modeling the workflows of GIS data production processes, which provides a formal, comprehensive understanding of the characteristics, strengths and weaknesses of existing contract based production models. As well, the formal modeling approach – Unified Modeling Language (UML) – provides means of modeling GIS data handling environments other than just production ones.
- The collaboration model is developed to give users a framework with which to evaluate and implement a collaborative workspace for supporting collaborative GIS

data production under contract-based production model. Such a framework would also assist in the design and development of off-the-shelf software that is specially designed to support collaborative GIS data production.

- The developed research prototype demonstrates the feasibility of integrating CSCW (groupware), database and GIS capabilities at both component and system levels onto the Internet infrastructure to support GIS data project management requirements and several operational tasks.
- The results obtained from performance testing will provide initial measures of the efficiency the proposed solutions bring to GIS data production projects. The documentation and procedures of the research prototype will enable further research and testing work.

1.5 Limitations of This Research

The research reported in this dissertation is subject to several possible constraints and limitations listed as follows:

- *Software Selection*: Due to a very large stock of groupware, it is not feasible to check all existing groupware packages to review the available collaboration functions. Instead, a limited number of the most popular groupware packages were selected to represent what is available in the market.
- *Workflow Process*: The examination of existing workflow processes using the contract production model in geomatics is limited since so much of the information

is proprietary in nature. Production manuals were obtained from *DataQC Inc.*, a consortium of several small companies specialized in contract inspection and located in Fredericton, New Brunswick [WaterMark, 1995; DataQC, 1998]. Relevant documents (work sheets, sample reports, official correspondences, etc.) proved hard to collect due to lack of formal documentation and poor organization and, sometimes, reasons of confidentiality. Although the available manuals do provide detailed information on process procedures, other information describing roles and relationships, rules, policies and associated ISO 9000 documentation are hard to obtain.

- *Data Format*: "Geomatics firms rely on GIS software for many of their core business functions." [Finley and Coleman, 1999]. Data format incompatibilities not only hinder sharing data transparently, but also affect the design of the collaboration model and the implementation of the collaborative workspace. The data used in this research is provided by SNB in CARIS ™ [1] format, which is the format of the CARIS GIS. Therefore, existing GIS installations are more focused on the CARIS GIS side. However, the processes could be applied to data stored in other formats as well.
- *Software*: In developing the collaborative workspace based on the designed model, the selection of software components may be limited because of budget and licenses issues. This could potentially affect the overall performance of the prototyped workspace.

[1] CARIS is a registered trade mark of CARIS, formally Universal Systems, Inc.

- *Hardware*: The hardware configurations and availability were constrained by the research budget. No hardware alternatives were available for performance testing in different hardware environments. Therefore, the results of performance testing should be considered as being affected by certain hardware combinations.
- *Access to the Internet*: The required Internet access for the research development and testing was obtained through the local area network at University of New Brunswick (UNB). UNB networks were connected to the NBNet through a T3 connection with 45 Mbs speed. The prototype testing was completed under the normal operating conditions common to the UNB environment.

1.6 Organization of the Dissertation

This thesis is organized into seven chapters, starting with this chapter which presents the overall introduction of research context, hypothesis and objectives, research contributions, and limitations that may affect the research results.

The second chapter aims at presenting necessary background knowledge for interpreting the results described in the following chapters, especially the concept of collaborative GIS data production, CSCW and groupware implementation, and methodology considerations for developing and evaluating the collaboration model.

Chapter 3 introduces the information sources for modeling current GIS data production projects and the results from examining the existing project management and production processes, followed by requirements specification of the collaborative workspace. The chapter finishes with a review of current groupware functionality.

The fourth chapter introduces the development of the proposed collaboration model, started with the overall research methodology followed by descriptions of three model components. Chapter 5 presents the development and testing work of the research prototype system, with a detailed introduction of prototyped functions within the collaborative workspace.

The results and analysis are presented in Chapter 6 and conclusions and recommendations for further research are summarized in the final chapter.

Chapter 2 BACKGROUND

Chapter 1 presented an overall introduction of this thesis research and proposed that using an Internet-based collaborative environment (through integrating groupware tools into the existing GIS installations) can help to improve efficiency and productivity of distributed GIS data production. While the subsequent chapters will be used to help to determine this assumption, this chapter presents necessary background knowledge for interpreting the results described in the following chapters. Especially, this chapter examines: (1) how the concept of collaborative production has been merged with geomatics data production processes and their work environments; (2) information technologies needed to support collaborative production environments; and (3) methodology considerations for developing and evaluating the collaboration model.

Section 2.1 discusses the concept of collaborative production in the geomatics arena from both organizational and technological perspectives, followed by Section 2.2 that introduces computer-supported collaborative work (CSCW) and its technology implementations. In Section 2.3, the considerations of designing a collaboration model and measuring prototype system performance are discussed. Before finishing this chapter, the author reviews relevant research activities in Section 2.4.

2.1 Collaborative GIS Data Production

The concept of Collaborative Production has been applied in fields other than geomatics (see Section 2.4.2) to solve real world production and development problems in multi-participated, distributed work environments, supported by CSCW concepts and technologies. Only recently has it started to draw attention in the geomatics areas. This section examines how this concept has been merged with the change of geomatics data handling processes and work environments, current information technologies required and the need for a collaboration model to support the design and implementation of this emerging development.

2.1.1 The Shift to Distributed GIS Data Handling

Beginning from the early 1980s, the concept of distributed computing and its supporting systems, mostly local area networks (LANs) and wide area networks (WANs) have been widely adopted in the GIS community at both the enterprise and inter-corporate level [Coleman, 1999]. While very few organizations used WAN services in the 1980s to routinely link together GIS users due to relatively slow transmission rates and narrow bandwidths for bulk transfer of large graphics and image files, developments in computer hardware and broadband data communications have changed this situation. With high-speed data communication services, users in the GIS community are now able to handle GIS data at locations hundreds of miles apart yet achieve comparable levels of performance [Coleman, 1994a].

While many LAN and WAN technologies exist, most are incompatible with each other in the sense that computers on a LAN may not be able to communicate directly across the WAN that connects to the LAN because this incompatibility keeps the WAN isolated from the LAN [Comer, 1995]. For any distributed application systems in general and GIS systems in particular, this implies that only a limited number of users in an organization can directly access services provided across the wide area networks.

By 1993, the Internet became the world's largest computer network – a network of computer networks available to a wider user community than just academic and research communities. Supported by standard network protocols and software, the Internet overcomes the incompatibility problems inherent in wide area networks and allows "mass" access of information and services available across this "wide-area" network. The use of the Internet for GIS data handling has been extensively researched and documented (e.g., [Dawe, 1996]; [Plewe, 1997]; [Hardie, 1998]).

While the Internet has provided another promising network platform for supporting distributed computing concept atop of LANs and WANs, the capability of distributed computing has been further empowered with the emergence of the "client/server" computing model in which a client system makes requests of one or more server systems. The concept of client-server in the context of geographical information systems (GIS) have been widely discussed in a variety of research (e.g., [Coleman, 1994a]; [Seggern, 1994]; [Plewe, 1997]). Seggern discussed how the client-server model could be used in distributed GIS database applications to deal with issues such as data extraction and direct updating, locking strategies and "versioning" mechanisms. Plewe, among others, described various implementation scenarios of the client-server model over the Internet. The variations are mainly based on "thin" or "thick" client concept and are generally summarized into the following two loose categories:

(1) *Thin Client*: All GIS functions reside on the server side. The Web browser, in this case, acts as a pure "thin" client interface (for presentation logic only) without any plug-ins, add-ons, Java Applets and any other scripting programs. Since clients rely heavily on server's computing resources, the server load and the network traffic factors have to be well considered and balanced.

(2) *Thick Client*: Client computing power is used to perform most GIS functions (for both business and presentation logic) such as data rendering, map browsing and simple data analysis and querying. Web-based GIS functions are usually implemented by web-enabled technologies such as Java and ActiveX. Data access logic is performed on the server side due to such issues as data security and access policies.

These classifications do not intend to be exhaustive. Actually, there are always some variations where the client is between "thin" and "thick", making a "balanced" client depending on the individual design of the overall client-server architecture at hand [Plewe, 1997]. Practically, the design of a client-server architecture is very much affected by the GIS functions provided by selected Internet GIS/Map server packages. For example, having the ESRI MapObjects Internet Map Server™ on the server side, the client side cannot manipulate GIS data in its native format with ActiveX controls (containing one or more map controls) within a web browser.

While the above discussions provide a brief introduction of existing computer networking technology provisions for handling GIS data in distributed environments, the formal use of these technologies in supporting collaborative data production has not been found yet.

2.1.2 Current Geomatics Work Environments

As has happened in many other industries such as utilities, the pressure from government deregulation and privatization strategies have contributed as external forces on many large mapping organizations to outsource their mapping tasks to other qualified small mapping firms [Fry, 1999; Li et al., 1999; Meyers, 1999]. Internally, the main motivation was to complete a large amount of work more quickly and cheaply than the organizations could if they kept it in-house. The way geomatics firms do their business has also been changed due to the shift from traditional data production techniques to digital mapping and geographical information systems [Coleman and McLaughlin, 1988], and the business goal shift from providing purely high-quality data products to providing not only quality but also total satisfaction to customers [Manheim, 1998].

Compared with an earlier industry dominated by a few large companies, the current geomatics industry in Canada consists of many small firms with specialized expertise [Finley, 1997]. Given the fact that each firm has its own competence and limited resources, these firms often form into industry consortia to realize certain business objectives on a program or project basis. In dealing with large-scale projects where participating firms may be geographically dispersed, the production procedures are actually carried out in distributed work environments.

Especially since the mid-1980s, a contract-based production (CBP) model, carried out in distributed GIS data production environments, has increasingly replaced the traditionally centralized production model [Coleman and Brooks, 1996; Finley, 1997]. The practical adoption of this model has been seen in Canada to handle provincial mapping programs in British Columbia, Alberta, Ontario, and Quebec, as well as at the federal level through Geomatics Canada [Sebert, 1989; Sabourin, 1994]. Outside Canada, instances of adopting CBP model were found at least in various mapping organizations in the United States [Brelsford, 2000] and China if not evident in any other countries.

More recently, as an example, Service New Brunswick (SNB) in Canada applied this CBP model in various spatial data production projects of preparing digital topographical maps covering the whole province and creating orthophoto maps for coastal zone management purposes [Castonguay, 1999]. Currently, SNB manages its contracts in-house and outsources all production and quality control activities to private industry.

With a contract-based production model, there are four major types of companies that may participate: (a) client who initiates contracts; (b) project/program manager who manages overall project/program; (c) production contractor who performs actual production tasks; and (d) sometimes an independent inspector who conducts data quality control. In some cases, the production contractor may sub-contract a partial production contract to one or more sub-contractors. While the organizational structure among the participating firms varies depending on the actual projects/programs, a generic structure is shown in Figure 2-1.

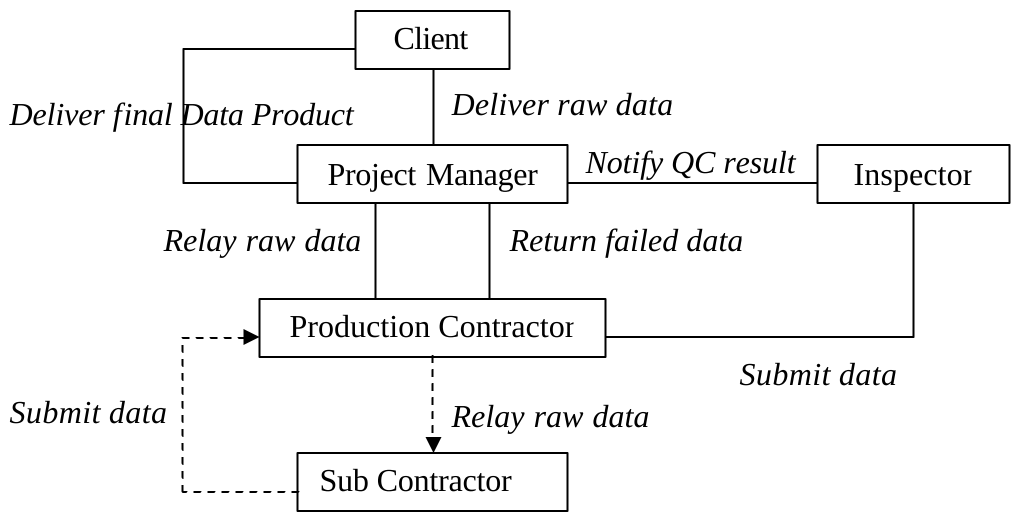


Figure 2-1 A general organization structure of contract production

In distributed work environments, the GIS data production process involves procedures being done at various locations, either across a city or across the country [Coleman and Brookes, 1996]. These procedures may include data production (e.g., collection, conversion and structuring), data inspection, returned data correction/verification, and initial and final distribution. While data materials involved are usually shipped by either couriers or, most recently, via telecommunication networks such as using FTP mechanism among production participants, the overall workflow process is still sequential in nature.

This changed paradigm does provide new horizons for geomatics organizations to: (a) focus on their core competence, (b) complement each other's specialty, and (c) increase capability in gaining market share. It also imposes challenges and problems at both management and operation levels, especially for GIS data production.

2.1.3 Discussion on Challenges and Problems

The multi-party nature of distributed GIS data production creates many new challenges for project or program management teams. The fact that each participating firm has its own polices and procedure in managing production operations and related information makes it very difficult to control overall project progress and information flow. While in general the success of any distributed production project relies heavily on timely transfer of project related information [Rojas and Songer, 1999], current approaches based on manual reporting and distribution are not considered satisfactory in terms of the time elapsed in preparing and delivering this information.

GIS data specifications as defined by production contracts are often complex with many technical details. Over the project life cycle, problems often arise when different parties have different interpretations of the same specifications or because they are not kept informed of the most recent specification updates in a timely manner due to the lack of efficient communications [Finley, 1997], causing the delay of project deliveries. In practice, some pre-production procedures related to technical specification validations and pilot of initial data production are adopted to ensure mutual understandings of contract specifications between participating parties [Castonguay and Doucette, 2000]. However, they do not ensure an informed updating process of specifications.

The communications of the project status, problems and progress are currently based on timely project reporting via either postal or electronic mail services. In many cases, log books and log databases are used as a repository for all project related information [Hastings and White, 1997; WaterMark, 1995]. The use of ArcView GIS to facilitate project management, allowing visualization of project status information has also been found [McConnell and White, 1999]. These approaches, however, are either time-consuming and unreliable or lacking of centralized management for information sharing purpose among distributed participants.

Currently, GIS data files may be transferred between distributed locations by either couriers or via the Internet (e.g., Internet-based data submission for quality control [Roberts, 2002]). However, the reliability, speed and efficiency may still be a concern because of the nature of spatial digital files and usually low speed “Dial-Up” Internet connections possessed by small geomatics firms. For example, digital map graphic files from SNB may range from 20 to 85MB in size and up to 225MB for color orthophoto image files. In addition, the number of data files involved in a single production project is usually very large; for example, there were 1895 sets of windowed digital map data files in SNB's topographic database restructuring project [Castonguay, 1999]. The lack of project-wide file management and storage mechanisms resulted in significant duplication of effort, waste of resources and difficulty of tracking different data submissions in current production practices [Finley, 1997], especially when multiple submissions occurred.

Beyond the transfer of project information and data management, an inefficient production process may be the biggest hurdle for the production throughput. While production procedures (e.g., QC procedures) are sometimes predefined and tested, and batch-processing tools are written for automating certain production tasks [WaterMark, 1995], they do not fundamentally improve the production process because of following two reasons:

(1) Not all parties accept predefined, tested procedures and tools simply because they prefer using their own procedures and tools [Finley, 1995]. To make the situation worse, all parties have to make appropriate changes if project-wide procedures and tools are adopted and have been updated.

(2) Currently all these procedures, no matter whether they are project-wide or not, are enforced manually and batch-processing tools are run by individual workers [WaterMark, 1995]. There is no formal automation mechanism used in this respect.

Problems existing at both management and operation levels are documented in several publications and research reports. For example, Finley [1997] identified several problems particularly associated with a particular contract-based project. Freiden [1997] described the importance of process improvement, while Brelsford [2000] discussed three reasons that caused project fails including poor communications, "scope creep" and lack of clear specifications. Obviously, these observations coincide with the above discussions (see problems summarized in Table 2-1).

Table 2-1 Summary of production management and operation problems

Type **Problem**	**Communication**	**Collaboration**	**Coordination**
Manage project progress	R	SR	SR
Solve production issues	R	SR	UR
Exchange project metadata	SR	UR	UR
Access data specifications	SR	UR	UR
Share QC procedures	SR	R	UR
Notify project updates	R	UR	SR
Transfer data materials	SR	UR	R
Track status of data files	R	UR	SR
Control version of data file	WR	SR	UR
Manage reporting activities	R	UR	SR
Legend: SR – Strongly Relevant R – Relevant WR – Weakly Relevant UR – Unlikely Relevant			

Generally speaking, the production management teams of GIS data production project or program in distributed environments are now facing the following challenges:

(1) how to efficiently manage multiple production contractors and multiple data submissions;

(2) how to share project and production information - including specifications, contract revisions, changes in procedures, solutions to production problems, and status tracking information - among project participants in a timely fashion;
(3) how to effectively reduce duplications of effort and waste of resources such as disk storage; and
(4) how to efficiently control and coordinate production processes and project information flow to be able to shorten "production float" and keep track of project schedules.

While there are certainly many organizational and social issues involved in handling these challenges and problems, technology will definitely play a very important role. It is on this technical perspective that this thesis focuses.

2.1.4 Collaborative Data Production

The concept of "collaborative production" has been applied in many areas to support various distributed work environments involving multi-parties, such as collaborative manufacturing [Poltrock and Engelbeck, 1997], collaborative CAD/CAM design [Kao and Lin, 1998] and collaborative product development [Bruce et. al, 1995], among others. The information technologies involved in supporting these collaborative production efforts include computer-supported cooperative work (CSCW) technologies, computer networking (especially the Internet in the last few years) and, certainly, domain-specialized software packages such as AutoCAD.

In geomatics, the first formal discussion on this concept was from an internal workshop, called "ChartNet Workshop on Collaborative Production", which gathered many experts from Canadian geomatics area to provide inputs for the Canadian Hydrographic Service and Nautical Data International (Ottawa, Canada) in support of a CANARIE-funded project on application of broadband networks to electronic chart production [Coleman, 1994b]. Although no final definition was derived from that workshop, the workshop participants did provide valuable insights on required supporting technologies for and needed capabilities by collaborative map and chart production environments.

While human-computer interaction principles are applied as fundamentals for interface design, both artificial intelligence and digital library techniques are also considered as keys for implementing a "smart" yet information-rich collaboration environment [Favreau and Mills, 1996]. As such, "collaborative production" has been considered not just a label but a distributed work environment that encompasses people, organizations, computer networks, production processes, supporting technologies, and suitable management policies.

A computer-supported virtual environment provides technological support for collaborative production. In the context of collaborative GIS data production, while networking trends and distributed computing technologies were discussed in Section 2.1.1, both people and organization aspects were briefly addressed in the above parts of this section.

2.1.4.1 Information Technology Need

Geomatics firms rely heavily on GIS and image processing software for many of their core business functions [Finley and Coleman, 1999]. Network-capable GIS software such as ArcInfo ™ and CARIS for Windows ™ has been available for many years to allow users to share the same geographic data and GIS functions. However, this often requires the same operating systems and platforms on all networked computer terminals or workstations. Recent development of Internet-based GIS technologies and standards (e.g., Open Geospatial Data Interface - OGDI from Global Geomatics Inc. and OpenGIS Abstract Specification from Open GIS Consortium) help overcome this system configuration incompatibility problem to some extent. However, when project participants in data production environments adopt different GIS software and computer platforms, GIS incompatibility issues will cause problems at the data quality control phase [Finley and Coleman, 1999].

Issues regarding how CSCW technologies can be formally applied to solve GIS-related problems, i.e., to support GIS communications, collaboration and coordination (3C) needs have not been well addressed in GIS communities. Some early efforts investigating the use of CSCW technologies to support collaborative GIS data production revealed the need of GIS-compatible CSCW tools and the need to apply formal workflow modeling and management tools in improving data production efficiency and productivity [Coleman and Li, 1999]. Two examples of this are:

(1) *Capability to simultaneously view and handle the same GIS data*: This would be extremely helpful in data quality control operations and provide production technicians on-the-spot assistance [Coleman, 1994b], where current practices rely on back and forth delivery of marked-up hardcopy maps or electronic images. In this case, functions that permit adding annotations and marking up certain features are assumed.

(2) *Workflow modeling*: In collaborative production environments, workflow becomes characterized by a greater number of operations completed in parallel to one another, hence need to be carefully modeled, if not redesigned [Coleman and Li, 1999]. The most effective adoption of any workflow modeling and management technology would require that the new workflow model should include certain predefined instances to help tackle basic workflow management problems. Efforts at detailed documentation of workflow components, processes, and tolerances required for ISO 9000 certification (among others) should help facilitate this process.

Technology needs to support collaboration in general - collaborative production in particular - have been widely discussed by many authors as mentioned above. Especially, Favreau and Mills [1996] discussed several key technologies for supporting a global collaboration infrastructure and Finley [1997] identified information technology needs for geomatics firms. In summary, the technology needs to support collaborative GIS data production are as follows:

(1) *High-performance Networking Infrastructure*: High-speed networking provides the underlying infrastructure that brings participating firms together in a restructured fashion [Finley, 1997] and enables project information and GIS data materials to move quickly and reliably across collaborative production environments.

(2) *Database Technology*: Databases serve as data warehouses for both project information (e.g., project metadata) and GIS data management. Online databases techniques such as online analytical analysis process (OLAP) and online transaction process (OLTP) may provide necessary support [Chaudhuri and Doyal 1997; Bedard et al., 2001; Rivest and Bedard, 2001]. In the long run, when GIS data in production is stored and managed in spatial databases, the spatial data warehousing and spatial OLAP principles may be required.

(3) *GIS and Mapping Techniques*: GIS software and mapping tools are necessary to perform actual data acquisition, conversion, editing and quality control tasks.

(4) *CSCW Technology*: This technology supports both formal and informal collaborations among participants. It can also provide support for controlling and automating production procedures, as well as coordinating project activities to control the overall project progress.

(5) *Programming Language*: Both programming languages and scripting languages are necessary for the implementation of technology integration.

With the support of these technologies, it is possible to develop a virtual collaborative environment that supports collaborative data production efforts. While the collaborative data production environments contain both organizational and technical aspects, it is its technical side on which this thesis focuses.

2.1.4.2 Collaborative Workspace

Several synonyms such as "shared workspace", "electronic workspace", "virtual workspace", and "distributed workspace", have been used in various areas to present the similar concept – a computer environment in supporting collaborative work. "Collaborative Workspace" was adopted in this thesis to denote a networked computer environment that supports collaborative GIS data production.

While high-speed networking is one of the key technologies required by collaborative work environments in general, Finley [1997] suggested that the Internet holds great potentials for supporting distributed geomatics projects, especially the Extranet-based model to be the most appropriate. The adoption of the Internet and more specifically the Extranet model allows the collaborative workspace to have an open platform that greatly facilitates project information exchange. However, there are two issues that must be properly considered in implementing a collaborative workspace:

(1) *Performance*: Although current broadband communications networks provide necessary performance for GIS functions and data management [Coleman, 1994a], the actual performance over the uncontrolled Internet is still not determined.

(2) *Interactivity and Compatibility of WWW*: Forms and CGI-scripts provide basic support for Web Interactivity. Although Java, JavaScript, VBScript and ActiveX Control all provide extra interactivities to web access, not all browsers support all of these scripts/languages. For example, Netscape Navigator does not support ActiveX controls. It is not likely that this problem will be resolved in the near future.

The implementation of a collaborative workspace will require system integration of the various components to provide an environment for total collaboration [Nunamaker, 1997]. While most required software components may already exist in some forms [Coleman and Li, 1999], the level of integration is still low in the sense that either the communications between these components are difficult or their interfaces are not compatible. In addition, the collaborative workspace has to offer more than just connectivity and integrated capabilities, tools for structuring processes and developing results should also be presented [Nunamaker, 1997].

To summarize, the collaborative production approach has the potential to become a viable option in the geomatics industry due to technology shift and a change of work environments. In order to realize this paradigm, a collaborative workspace must be in place to provide a networked (in case of this study, Internet-based) computer environment for a virtual team consisting of geomatics organizations and data production firms.

While most technology components are already in place to support this kind of collaborative workspace, the actual implementation is still problematic because of the complexity of software integration and lack of workflow process modeling. Therefore, a model must be developed to address these problems and help the industry easily implement the needed collaborative workspace. Before moving on to the discussion of collaboration model design, collaborative workspace prototyping and performance analysis, the next section provides an in-depth look at CSCW principles and technology implementations.

2.2 Computer-Supported Collaborative Work (CSCW)

Although the term CSCW is sometimes used interchangeably with "groupware", there is an essential difference between the two. While CSCW refers to the field that studies the design, adoption and use of technologies that supports team work, groupware refers to the pure technology implementation of CSCW.

The computer-based coordination and collaboration mechanism was initially created to support information sharing among groups and/or members in working teams for better and fast decision-making (e.g., group decision support systems) and routing work-related documents in a controlled way (e.g., workflow), mostly in a centralized homogenous environment [Vinze, 1997]. Over the years, systems based on this mechanism have been spanned to incorporate more complicated functionality over widely distributed computer networks.

This section provides an overview of some technical details by introducing basic concepts of CSCW and, more specifically, groupware, followed by an examination of implications of applying them to solve geomatics related problems. The final part of this section presents a summary of discussions.

2.2.1 Basic Concepts

Computer-supported collaborative work (CSCW): This term was first coined to describe an identifiable research field focused on the role of the computer in group-oriented work [Greif (editor), 1998]. Although many synonyms have been used since then, the role of CSCW as an umbrella collecting researchers from a variety of disciplines to contribute different perspectives and methodologies for both acquiring knowledge of group and suggesting how group's work can be supported remains dominant in the area [Greenberg, 1991].

While intensive discussions have taken place on defining the term – CSCW may be found in many books and journal papers (e.g., [Greenberg, 1991]; [Khoshafin and Buckiewicz, 1995]; [Coleman and Kbanna, 1995]) – no single definition is widely accepted yet. However, some commonly shared senses can be drawn from these diverse definitions including CSCW as a research field, associated with technical, human and social aspects, and as a framework for technology implementations. Therefore, in this research, CSCW is loosely defined as "a research field of studying the design, adoption and use of groupware technologies".

Groupware: As pure technology implementation of CSCW, groupware has been around since the time even before the term CSCW was coined (the term – "groupware" started to appear in early 1980s). Although various definitions have been given or adopted by researchers under different circumstances [Coleman and Kbanna, 1995; Khoshafin and Buckiewicz, 1995], groupware is essentially the software or software systems which support and augment group work. For the purpose of this research, the following definition is used:

> *Groupware is a set of various software tools and technologies facilitating computer-mediated communications, cooperation and coordination that increase the productivity or functionality of human-to-human and human-to-computer processes. [Coleman, 1995]*

2.2.2 Groupware Components

Several classification schemes have been researched and used to categorize groupware functions. Among these efforts, Coleman and Kbanna [1995] and Khoshafin and Buckiewicz [1995] provided substantial discussions in this respect. The schemes they discussed are based on organizational functions, groupware products and time-place dimensions, respectively. Each classification scheme categorizes groupware functions in different perspectives in its own right. While the "time-place" scheme provides a high-level classification of groupware functions into four option spaces, the product-oriented scheme focused on groupware product functionality to classify. Both schemes have gained the wider popularity in the sense that their uses have been found in a wider variety of literature. Figure 2-2 illustrates how groupware functions are categorized into "asynchronous" or "synchronous" groups with respect to time and place.

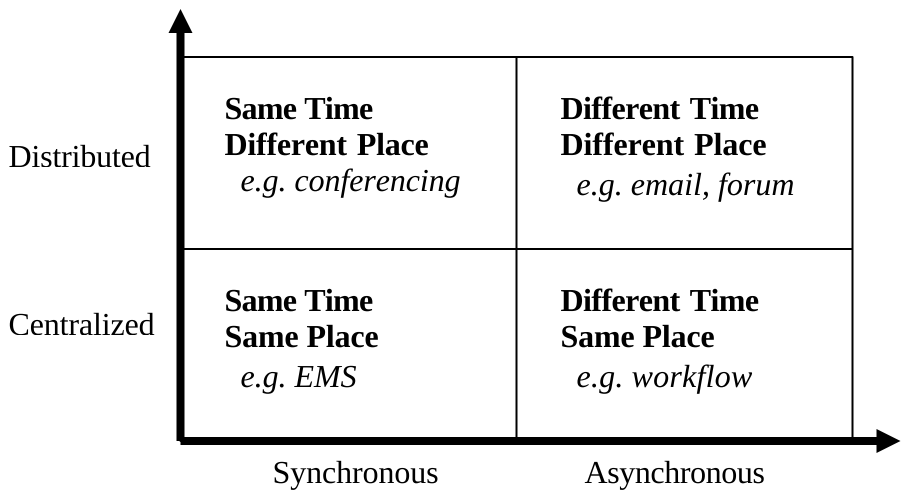


Figure 2-2 Classification of groupware based on time and place dimensions

Khoshafian and Buckiewicz [1995] stated that there is always overlap between groupware categorized according to groupware product-based scheme. This is also true for other classification schemes. In case the "time-place" scheme is used, one groupware function may fall into one or more groups depending on how the function is designed and implemented in the overall application systems. For example, a *"collaborative editing/viewing"* function may be categorized in the "same-time and different place" group if it is designed to allow users at different locations to simultaneously edit or view the same content. It may also fall into "different-time and different-place" group if synchronization effort is not required.

Detailed taxonomies of groupware components have been discussed in many publications (e.g., [Coleman and Kbanna, 1995]; [Khoshafin and Buckiewicz, 1995]; [Finley, 1997]; [Podgorny et al., 1998]). Boettcher [1999] and Li et al. [1998] also examined the market as of 1997, 1998, and 1999 based on selected groupware vendors. However, little evidence has been found that a formal scientific approach for building this kind of taxonomy was well researched. The product-oriented scheme was adopted to describe the groupware function taxonomy to be used for this research. Various groupware vendors' profiles were considered in this summarizing effort. Table 2-2 lists all groupware components identified from these efforts and maps them to a "time-place" classification scheme.

The list of groupware components in Table 2-2 is not exhaustive. Instead, it provides an overview of major groupware components that may be found from existing groupware packages. As seen from this classification, many functions are overlapping across one or more "time-place" spaces, in which case decisions have to be made at the design stage to decide how the functions behave in the overall collaborating systems. While functions of the above groupware components have been extensively discussed in CSCW-related and computer literature, more comprehensive descriptions may be found from Coleman and Kbanna [1995] and Khoshafin and Buckiewicz [1995]. To build a basic understanding of these groupware components, they are briefly explained as follows:

Table 2-2 Groupware components classified with respect to "time-place"

Groupware Component	Synchronous		Asynchronous	
	ST – SP	ST - DP	DT - SP	DT - DP
Electronic Mail and Messaging				
▪ Email			√	√
▪ Real-time Chatting (RTC)	√	√		
▪ Discussion Group/Forum			√	√
Group Decision Support				
▪ Electronic meeting	√	√		
▪ Whiteboarding – shared board	√	√		
▪ Voting	√			√
▪ Audio/Video Conferencing		√		
Collaborative Document Handling				
▪ Electronic file transfer		√		√
▪ Collaborative editing/viewing	√	√	√	√
▪ Version control (multiple-access)			√	√
▪ Check-in/out (multiple-access)			√	√
▪ Cataloging/indexing/filtering	√	√	√	√
Workflow				
▪ Task scheduling/routing				√
▪ Information routing				√
▪ Application invoking				√
▪ Process modeling				√
Group Calendar and Scheduling				
▪ Calendar creation			√	√
▪ Resources scheduling			√	√
▪ Conflicts checking	√	√		
▪ Event Notification				√
▪ Group calendar sharing		√	√	√

Electronic Mail and Messaging

Familiar to most people now, electronic mail and messaging functions provide simple mechanisms for sharing mostly unstructured information among groups of people in an uncontrolled, sometimes informal way. To control and make the conveyed information useful knowledge, rules and message templates have to be applied, which very often compromise the flexibility of using these mechanisms.

1. *Email*: Considered as the first successful groupware application, email is supported by all network models from LAN to the Internet. Email relies on transport services, directory services and email database services on the back-end via operating system and email APIs. With persistent information about senders, groups, recipients, messages, routing information/rules, and so forth, email applications can potentially enhance the corporate memories. There are two ways to incorporate email functions into a collaborating system: stand-alone application or email-enabled application [Khoshafin and Buckiewicz, 1995], where, in the second case, email is either a separate option or a "*printer-option*" of an application.

2. *Real-time Chatting (RTC)*: Allows two or more participants engaged loosely into a communication session to express their ideas on any *ad-hoc* topics. However, its requirement of maintaining a synchronous session makes it less efficient approach that can be adopted by collaborative production systems if the system serves users located in different time zones.

3. *Discussion Group/Forum*: An alternative mechanism that has been supported for many years by public or private email-based systems allows messages being posted on a bulletin board or virtual forum rather than routed to individuals [Khoshafin and Buckiewicz, 1995]. Tracking message threads is required at both message storage

and representation (a *"response hierarchy"* as described by [Khoshafin and Buckiewicz, 1995]) level. While the majority of use now is for informal information and idea exchange especially over the Internet, adopting this functional component within a project group or corporate boundary helps in several aspects. Some significant benefits are: (1) exchanging technical solutions to reduce duplicate efforts; (2) discussing problems to find better solutions; and (3) enriching the project or corporate knowledge base if this mechanism is backed up by a well-designed message database.

Group Decision Support

Group decision-making functions use proper techniques, software and technology designed to focus and enhance the communication, deliberations and decision-making of groups [Nunamaker, 1997]. Generally speaking, group decision support includes efforts such as electronic brainstorming, polling of consensus, and evaluation of alternatives. Nunamaker presented an overall review of this field and addressed general issues regarding the needs, research and challenges faced by both users and researchers.

1. *Electronic Meeting*: Essentially as a set of networked personal computers or terminals with proper provision of facilities, the electronic meeting can conduct locally or in a distributed mode. While a LCD projector is used as facilitating equipment in a local meeting room, software is normally used as a substitute for LCD projectors for distributed meetings [Coleman and Kbanna, 1995]. The development of electronic

meeting systems is now at the point where virtual reality technology plays a very important role to form so-called distributed virtual meeting systems [Nunamaker, 1997]. Yet, many issues such as anonymity, equal participation, reduced domination, and group memory are of concern. The following two factors should especially be considered when planning a distributed electronic meeting:

- Participants may reside in different time zones making it difficult to schedule the meeting; and
- It is hard to have participants hold their attention long enough when they have so many distractions such as talking to office mates and checking emails.

2. *Whiteboarding – shared board*: An electronic whiteboard is a virtual space shared by people in dispersed locations. Traditionally several LCD displays were set up at various locations and connected through network connections to allow people to have synchronous views of displaying objects (e.g., text, graphics or marking features). Recently, especially with the adventure of the Internet, whiteboarding sessions can be set up among distant users by having a whiteboard window on each one's desktop screen. While most whiteboarding tools only support limited file formats such as text and images (GIF, JPEG and BMP), sharing word processing documents and other domain specific documents such as CAD files within shared whiteboard are possible with some tools. Examples may be found in Netscape Communicator ™ and Microsoft NetMeeting ™ systems.

3. *Voting*: While voting used to be used to make simple decisions within corporate or private group boundaries with an electronic meeting system, electronic voting is now commonly practiced for “electronic democracy” [Khoshafin and Buckiewicz, 1995] over the Internet, gathering either public or group opinions on specific issues. The front-end interface of voting systems normally consists of "Yes/No", numeric, or other selective options to allow participants express their opinions. On the back-end, results are collected and tallied to generate various outcomes. Sometimes the voting functions may be embedded in a workflow management system (e.g., Oracle Workflow ™).

4. *Audio/Video Conferencing*: Both audio and video conferencing provides two-way interactivity for collaborations among distant participants. Since the major hurdle in audio and video conferencing is overloaded information to be transmitted, especially for video conferencing, CODEC technology remains the heart of this kind of system [Khoshafin and Buckiewicz, 1995].

Collaborative Work Document Handling

Some *de facto* interoperability standards such as OpenDoc (http://www-4.ibm.com/software/ad/opendoc/) favour the concept of taking each piece of “work” material as an object. This was further elaborated in a collaboration object model in [Khoshafin and Buckiewicz, 1995]. The collaboration object could be a document people work on when performing certain assigned tasks, the results a specific collaborative work process generates, or a decision being made by a group. In this thesis, "object" will be used to denote any work materials wherever appropriate.

1. *Electronic file transfer*: The capability based on FTP protocol to transfer files from one computer to another, usually between client machine and server. For large image and graphics files, compression and decompression (CODEC) algorithms may need to be applied at both the client and server ends. In addition, FTP tools must be able to resume an interrupted file transfer due to network errors.

2. *Collaborative editing/annotating/viewing*: This type of functions allow multiple users to edit, annotate or view the same document in any time-place space as shown in Figure 2-2 [IGD, 1998], depending on how functions are setup and how the group of users wants to use them. Collaborative editing/viewing implies the concept of "*what you see is what I see*" (WYSIWIS). It is not unusual to see the actual implementation of the functionality in whiteboarding tools if synchronous collaboration is favored. To keep track of edited documents, version control mechanisms have to be in place.

3. *Version control (multiple-access)*: Version control allows tracing back through the history of collaborative objects to, for example, undo most recent changes. It can be realized by applying many theoretical concepts such as a "version management tree" [Khoshafin and Buckiewicz, 1995] or "Timewarp" [Edwards and Mynatt, 1997]. Version control mechanisms help the collaborating system in building a corporate memory.

4. *Check-in/checkout (multiple-access)*: Check-in/out functions are widely used in systems where one or more central repositories exist. The objects stored in the repository can be checked out, possibly locked (either optionally or enforced by the system), examined/updated, and then checked in and unlocked. Usually only a copy of the object is checked out. The original copy remains in the repository and may be checked out by others, even updated, depending on the "lock" status imposed by the people who previously checked out the same object and have not checked it in yet. Check-in/out functions are normally integrated with version control functions. In most cases, the object is a single file being checked in/out individually (e.g., Oracle Internet File System *9iFS* ™). It may also be a directory that contains many files and sub directories [Khoshafin and Buckiewicz, 1995].

5. *Cataloging/indexing/filtering*: The ultimate purpose of having these functions in any collaborating system is to find information and/or documents quickly. While filtering focuses on search and presents partial contents of the document that satisfies user-defined criteria, cataloging and indexing allows a set of unique attributes to be defined and associated with a document for later search and retrieval. Instead of using a dynamic scheme defined by users, cataloging and indexing is based on predefined structures [Coleman and Kbanna, 1995].

Group Calendar and Scheduling

Originated from the concept of “time management” to manage personal calendar and contacts [Coleman and Kbanna, 1995], this type of application now allows networked users to connect to the same calendar database where both personal and group based schedule, contact and resources information may be stored. By including task and resources scheduling capability, networked users are able to obtain combined view of both personal and group calendars and to efficiently schedule personal and group activities as well as resources (very often limited) such as office equipment and space.

Usually calendar entries are set up with different levels of access and importance. While the access is generally classified into three levels: personal, group and enterprise, the implementation varies from one application to another, for instances, Netscape Calendar server allows four distinct levels (normal, confidential, personal, and public). Importance level is another dimension of calendar entries that labels the entry as having highest, high, normal, low, or lowest priority.

One significant benefit of using group or enterprise based calendar/scheduling application is to help build human resources allocation and other resources usage information into project or corporate memory, which is then used for future planning.

Because of various access restrictions to entries of a non-personal calendar, scheduling effort definitely needs support of the following two features, one way or another:

(1) *Free time searching*: Networked users can search free time slots of any specific person or the group as a whole before scheduling an event or task. They can search for free facility resources to make sure it is available.

(2) *Automatic conflicts checking*: When an entry is to be added into a calendar, it will be automatically checked against all scheduled events and resources. Any conflicts will be noted and, optionally, alternatives provided.

Workflow

Workflow was originally identified and applied with document imaging applications and office procedure automation where their steps and tasks in the process had been clearly defined or well structured [Coleman and Kbanna, 1995; Khoshafin and Buckiewicz, 1995]. In the past few years, it has been split into a separate technology with the market reaching one billion US dollars in North America and Europe in 1997 and more in 1998 [Felice, 1996; Dykeman, 1998]. At the same time, workflow is entering enterprise-wide mainstreams and becoming part of the standard IT landscape [Felice, 1996; GIG, 1997; WfMC, 1998].

The motivation of utilizing workflow in any work process is to enhance productivity and coordinate process activities and resources (including people involved). Workflow can be a tool to facilitate the following [Felice, 1996; GIG, 1997]:

(1) streamlining processes for better delivery of customer satisfaction;

(2) eliminating redundant and unnecessary work;

(3) empowering users for flexibility and responsiveness;

(4) improving coordination across and among organizations;

(5) redesigning business processes incrementally; and

(6) integrating applications, especially external applications.

Because workflow technology plays an important role in this research, Section 2.2.3 in this chapter is dedicated to the discussion of its technical details relevant to the research.

2.2.3 Workflow

2.2.3.1 Basic Definitions

Although there are variations, the following definitions are given based on the definitions of Workflow Management Coalition [WfMC, 1998]:

Business Process: A business process is "a set of one or more linked procedures or activities which collectively realize a business objective or policy goal, normally within the context of an organizational structure defining functional roles and relationships". Activity is defined here as a description of a piece of work that forms one logical step within a process. It may be a manual or workflow (automated) activity. Some commonly used synonyms of activity include step, procedure, and task.

Workflow: A workflow is "the automation of a business process during which documents and tasks are passed among participants according to a set of procedural rules and assigned roles".

Workflow Management System: A system that defines, creates, and manages the execution of workflow through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke appropriate IT tools and applications.

To fully understand workflow technology, it is very important to distinguish these concepts from each other. The discussions in the following sections under Section 2.2.3 will help to further clarify these distinctions.

2.2.3.2 Workflow Standards

Formed in 1993, the Workflow Management Coalition (WfMC) is the main international standard body dealing with workflow related standard efforts. One of the major artifacts produced by WfMC is the Workflow Reference Model (WRM), which specifies both architecture and software components of workflow management systems supported by five interface specifications dealing with workflow interoperability. While the key issues of this standard are briefly described as follows, more details may be found from WfMC's web site (http://www.wfmc.org/).

Workflow Reference Model

Architectural components of workflow management system vary from one system to another, making easy interaction between software components from different vendors very difficult. This problem has been well addressed by the Workflow Management Coalition [WfMC, 1995]. The purpose of having the workflow reference model (see Figure 2-3, from Workflow Management Coalition) is to ensure that all components of any workflow management system do not have to be provided by the same vendor. However, this objective has not yet been achieved in practice.

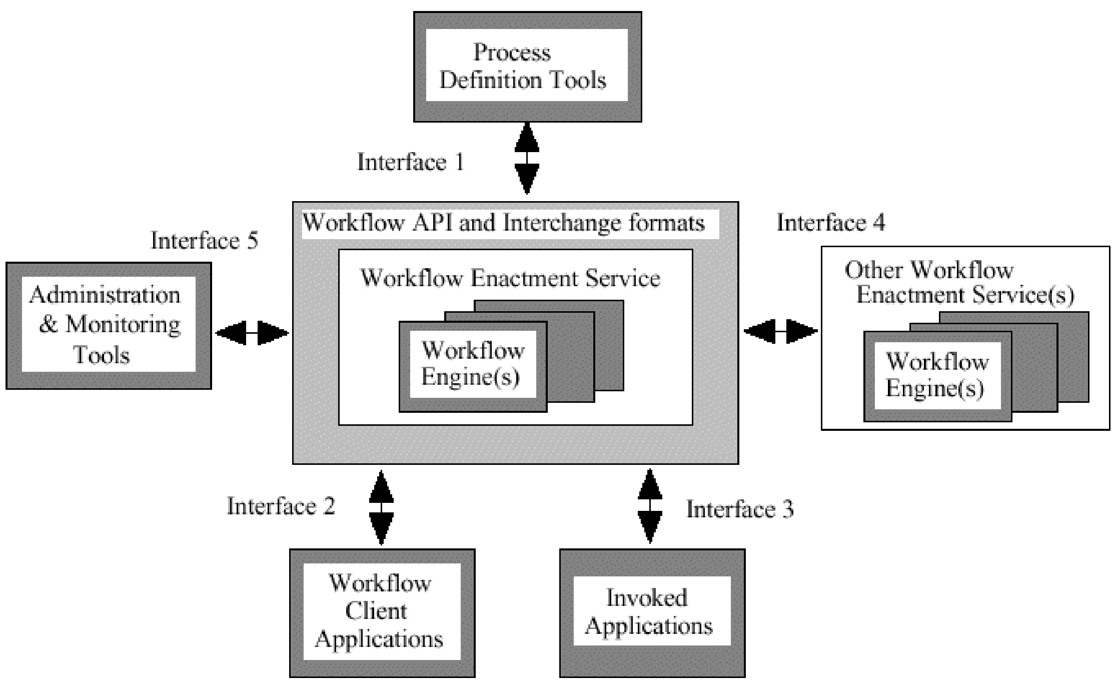


Figure 2-3 Workflow Reference Model
(Source: [Workflow Management Coalition, 1995])

The model was originally released in 1995 to facilitate workflow system (WFS) vendor's interoperability efforts. Supporting this model, WfMC also released a series of interface specifications that may be used by venders and workflow application developers to develop WRM compliant WFS components. The model describes the architecture of a workflow management system with emphasis on the interfaces between system components. While components are briefly described in Table 2-3, the workflow reference model illustrated in Figure 2-3 shows relationships between these components and where each interface specification fits.

Table 2-3 Architectural components of the workflow reference model

Workflow Enactment Service	Consists of one or more workflow engines and is responsible for creating, managing and executing workflow instances based on workflow process definitions.
Process Definition Tools	Are used to analyze, model, describe and document a business process
Workflow Client Applications	Provide a worklist handler that interacts between client-users and workflow engines when human interaction occurs along workflow execution. The worklist handler may be written by users to customize client application interfaces.
Workflow Invoked Applications	Allow invocation of external applications through identified interfaces including local process call, shell script, ORB call, remote execution call, message passing (e.g., X400), and transaction (e.g., OSI-TP).
Workflow Administration and Monitoring Tools	Manage workflow users and roles, control resources, and monitor workflow execution
Other Workflow Enactment Service	Present only when multiple workflow enactment services are required

Interface Specifications

While the Workflow Reference Model provides an overview of the WFS architecture and relationships to the five interface specifications, it is the first interface specification (Interface 1 - Definition Interchange) that provides guidelines instructing how to model a workflow so that interoperability can be realized. Since this Interface will be used for workflow modeling discussed in Chapter 3, it will be discussed here in detail (Note: the diagrams illustrated below are all from Workflow Management Coalition - Interface 1).

The Definition Interchange Interface defines three artifacts at different levels. They are, in the order from high to low, workflow model, workflow process model and workflow process definitions (meta-model). At the highest level, this specification defines all the entities a workflow should have and how these entities relate to each other (see Figure 2-4), which form the workflow model (OM stands for organizational model).

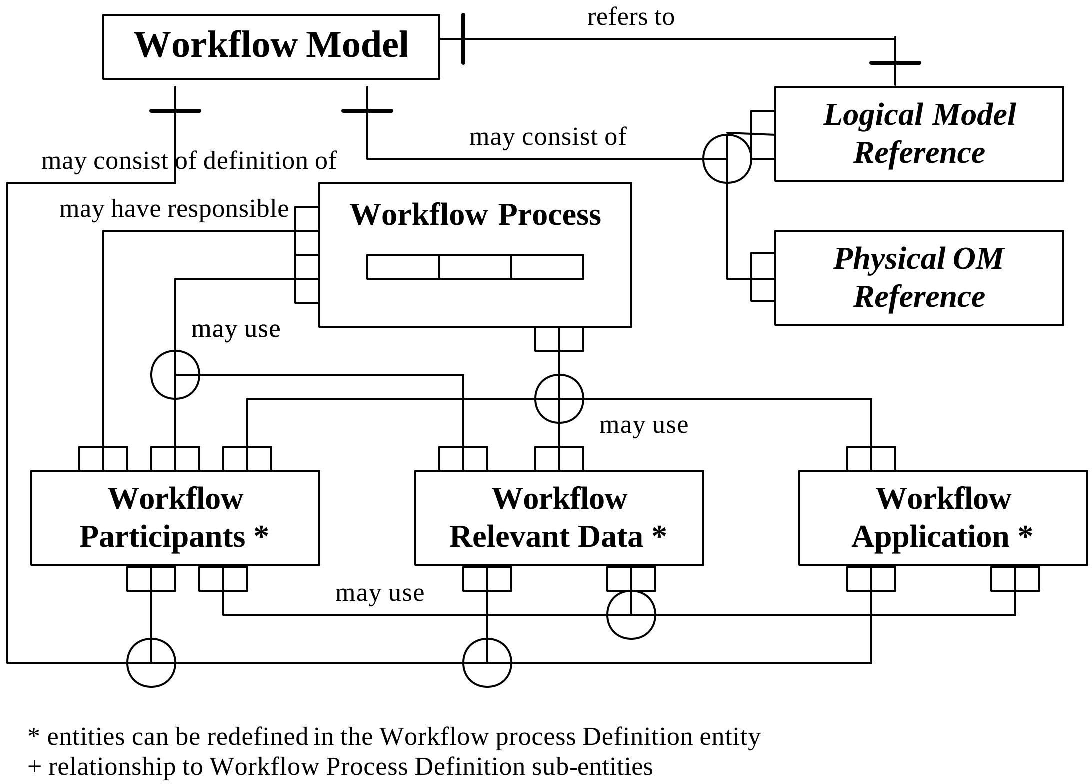


Figure 2-4 Workflow Model Entities

(Source: [Workflow Management Coalition, 1999])

Because workflow is about the automation of a set of procedures, the workflow process model (see Figure 2-5) remains the core of the workflow model. Each workflow process model may contain one or more process definitions that follow a meta-model for process definition (see Figure 2-6). The process meta-model defines top-level entities including process objects, relationships between objects and attributes associated with the process and its objects.

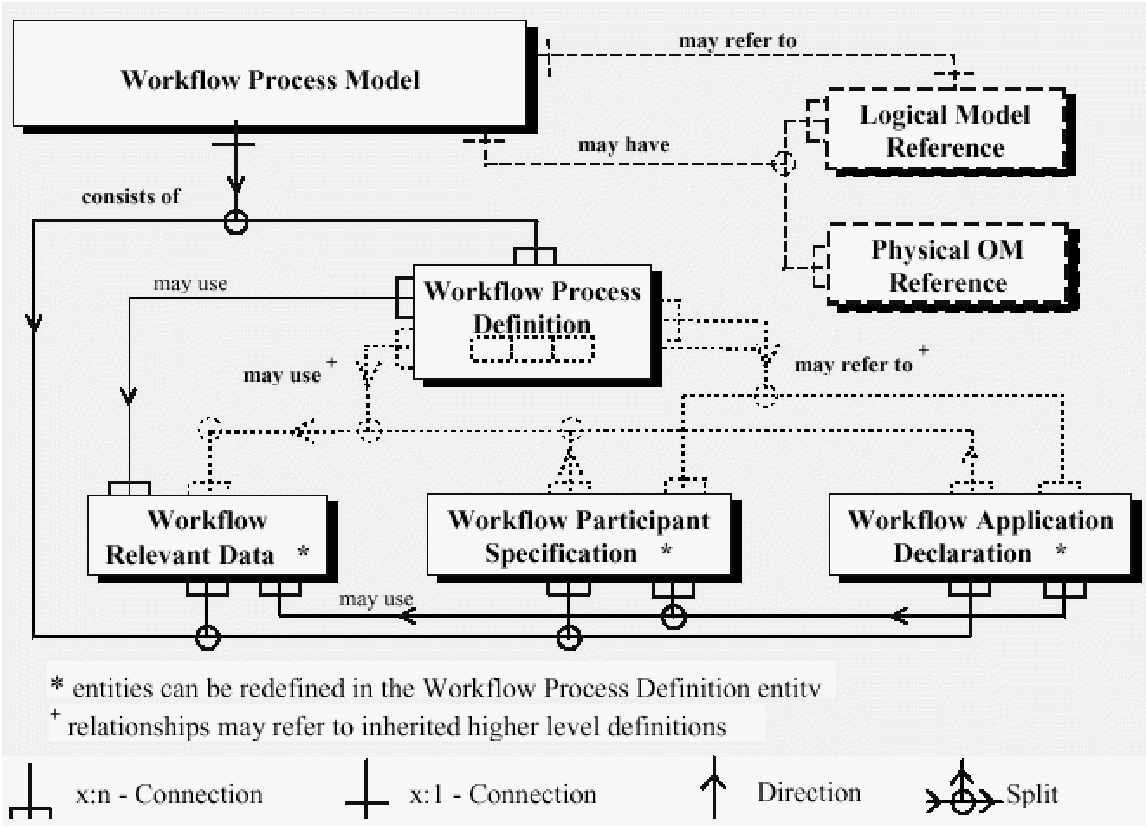


Figure 2-5 Workflow Process Model Entities

(Source: [Workflow Management Coalition, 1999])

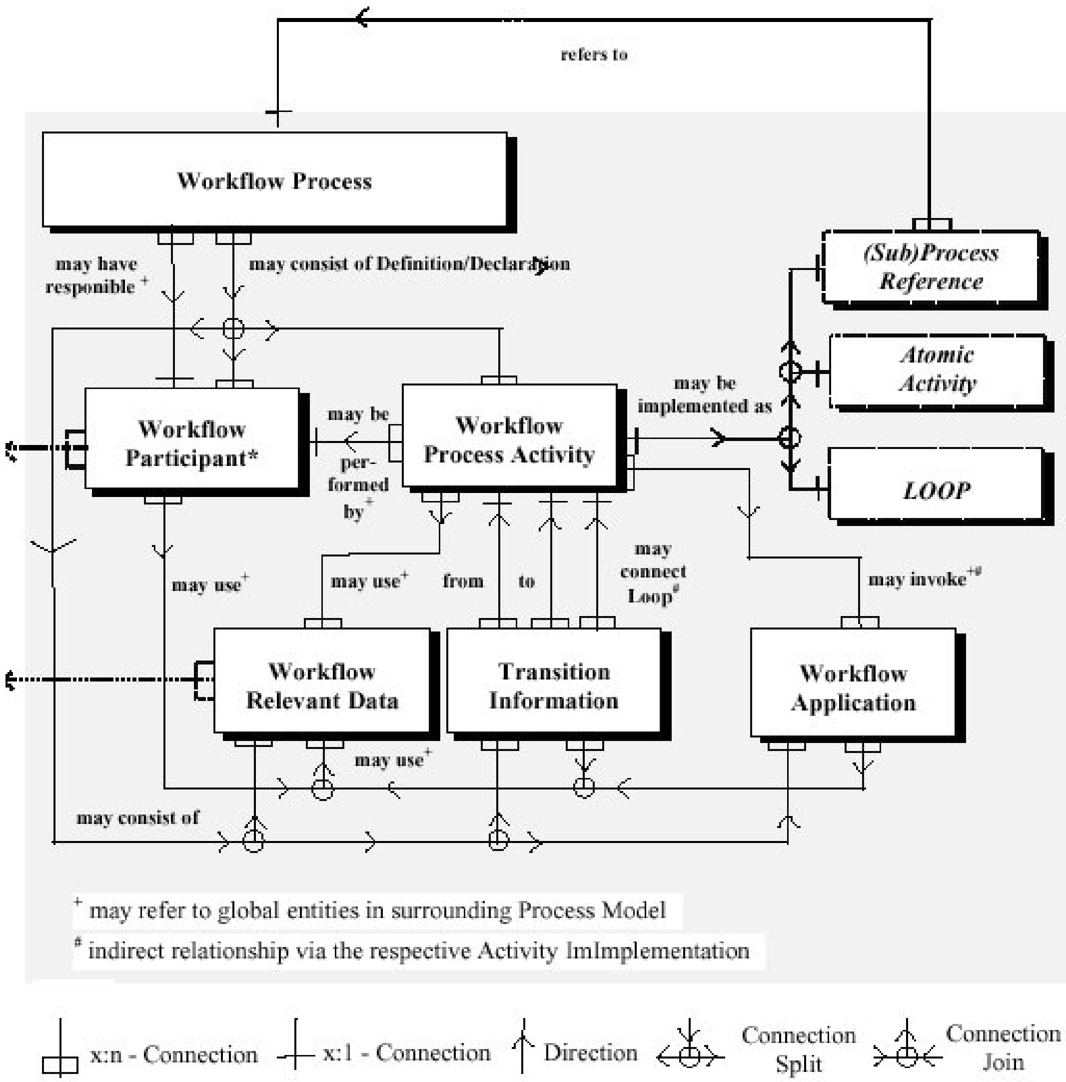


Figure 2-6 Meta-model Top Level Entities
(Source: [Workflow Management Coalition, 1999])

The object-based hierarchical organization, with the workflow model at the top and process definition at the bottom, allows the lower-level model to redefine those entities (e.g., workflow relevant data and workflow applications) common to the higher-level model. Yet all process details such as activities and transitions are only defined at the process definition level. While the workflow process model has a very similar structure as the workflow model, there are some essential differences between these two:

(1) For each specific domain problem, the workflow model may have one or more workflow process models. Each of them may consist of more than one business process definition. Each process definition corresponds to one business process.
(2) While any workflow entities defined at the workflow model level such as workflow relevant data and workflow applications are visible inside each process model, the process model can only use entities defined inside its scope.

Discussion

Other relevant standard efforts include Workflow Management Facility Specification (WMFS) provided by the Object Management Group (OMG) [OMG, 2000]. OMG is a non-profit organization promoting CORBA related technologies. Its Workflow Management Facility Specification is an effort to integrate workflow management facilities within its object management architecture. Since the focus of OMG WMFS is on software integration rather than workflow modeling, it was not selected in this research.

Although some WFS vendors adopted modular architecture in their system developments [Coleman and Kbanna, 1995], most workflow management systems in the market are still proprietary in nature in the following two senses [Action, 97; Plexus, 98; Chang et al.]:

(1) The workflow engine remains the core of any workflow management system – meaning that WFS modules may be installed separately and optionally, but they have to interact with a specific workflow engine that supports these modules.

(2) The efficient mechanisms that allow workflow functional components, especially workflow engine component, to be easily embedded within or integrated with other applications are not there yet.

By comparison, the WfMC standards clearly favor component-based software developments and supports interoperable operations among workflow software components from various workflow management systems.

As a picture can say more than a thousand words, having a graphic presentation of workflow process definitions can greatly enhance communications between business analysts and domain experts. The WfMC decided not to include a graphical standard for process definitions due to the great diversity of tools already in the market. Therefore, a language-based standard, named as Workflow Process Definition Language (WPDL), was provided to describe entities specified on diagrams in Figure 2-4, Figure 2-5, and Figure 2-6. This modeling language will be further elaborated in the next section.

2.2.3.3 Workflow Modeling

Modeling a workflow has a different focus from modeling a business process. While business process modeling answers the questions beginning with "What" (higher level), "Why", "What-else" and "What-if", the workflow modeling focuses on answering questions beginning with "What" (detailed level), "How" and "Who" ([Aalst and Hee, 1995] and [Amberg, 1996]). Many modeling languages have been used to model business processes and workflow in the past [Darntons, 1997; Van der Aalst, 1998; Cichocki et al., 1998; Eriksson and Penker, 1999]. In addition, almost every workflow management system has a workflow process definition tool that may use its own or selected modeling approach. While deferring in semantics, presentation, documentation, and so forth, most of these modeling languages fall in one of the following groups:

1) Flow-charting based (e.g., Petri-Net, ANSI Process Charts, and IDEF Diagrams)
2) Object-oriented analysis (e.g., Unified Modeling Language)
3) Text based (e.g., Workflow Process Definition Language)

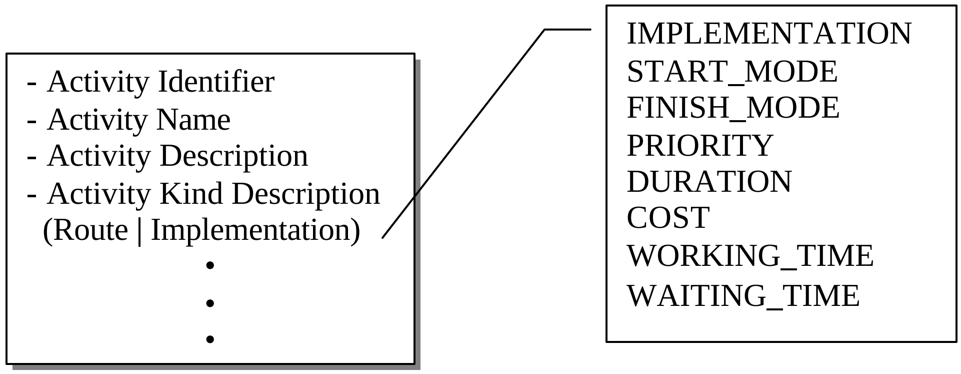
An important reason for using a standard language is to build a common design repository of domain workflow process definitions that can be accessed by a number of various tools and run-time systems [WfMC, 1999]. Given the great diversity of modeling languages and workflow process definition tools, it is not possible to examine all of them in this thesis. Instead, only the Unified Modeling Language (UML) and Workflow Process Definition Language (WPDL) will be examined here. Using a combination of them can present workflow models both graphically and textually, yet comply with the existing standards (relevant standards bodies have already adopted these two modeling languages).

While detailed technical information about these two modeling languages may be found through extensive literature (e.g., [Booch et al., 1999]; [WfMC, 1999]) and appropriate standard bodies (OMG and WfMC), key ideas are briefly discussed below.

Workflow Process Definition Language (WPDL)

As discussed in Section 2.2.3.2, the Workflow Process Definition Language (WPDL) is used to describe meta-model entities (objects, relationships and attributes) that build up workflow models with predefined and extendable constructs. The description of each entity is a sequence of an entity keyword, followed by an identifier, an attribute list, and an end-entity keyword (see example descriptions for workflow activity illustrated in Figure 2-7). While WPDL provides minimum sets of predefined attributes by meta-model (including both mandatory and optional ones), it also provides a generic symbol for defining extended attributes, as listed below, to allow individual users to satisfy their own set of attributes.

```
<extended attribute list> ::= EXTENDED_ATTRIBUTE
                                <attribute id>
                                <attribute type>
                                <attribute value>
                                [<description>]
                                [<extended attribute list>]
     <attribute id> ::= <identifier>
     <attribute type> ::= <complex data type>
     <attribute value> ::= <initial> | <function access>
     <description> ::= <string>
```




```
ACTIVITY                    'Data_Inspection'
        NAME                "Processed Data Inspection"
        DESCRIPTION         "A sub flow deals for data quality control"
        IMPLEMEMTATION      WORKFLOW SYNCHR
        XOR JOIN            T0_2, T0_5
        XOR SPLIT           T0_3, T0_4
END_ACTIVITY
```

Figure 2-7 Example of WPDL Entity Description

The list below shows the overall description structure of a complete workflow model, where the entities in square brackets are optional. Apparently, omitting too many optional entities does not really present a useful model, especially with the "Workflow Process Definition" entity [WfMC, 1999]. Normally, the "Workflow Process Definition" entity contains more than one process definition, which is also enclosed by a pair of keywords: WORKFLOW and END_WORKFLOW.

```
MODEL <model id>
     <Workflow Model Definition Header>
     [<conformance class declaration>]
     [<extended library declaration>]
     [<external model declaration>]
     [<Workflow Participant Specification>]
     [<Workflow Application List>]
     [<Workflow Relevant Type Declaration List >]
     [<Workflow Relevant Data List>]
     [<Workflow Process Definition>]
     [<extended attribute list>]
END_MODEL
```

__Unified Modeling Language (UML)__

UML is a language and notation, adopted by the Object Management Group (OMG) in 1997 as a standard, to specify, construct, visualize and document software-extensive systems. Recently, it has been applied in the business analysis world to model business processes and workflows [Dewalt, 1999; Eriksson and Penker, 1999].

The overall benefits of using UML to model software systems ranging from enterprise information systems to web-based applications have been extensively discussed at length in, for example, [Conallen, 2000] and [Booch et al., 1999], among others. By modeling systems with various UML building blocks (Things, Relationships and Diagrams - see Figure 2-8, more details may be found from [Booch et al., 1999]), system designers and domain experts have better communication facilities yet retain rich documentation. With extensibility mechanisms such as *Stereotype*, *Tagged Values and Constraints*, UML enables expressions of many possible nuances across various domains.

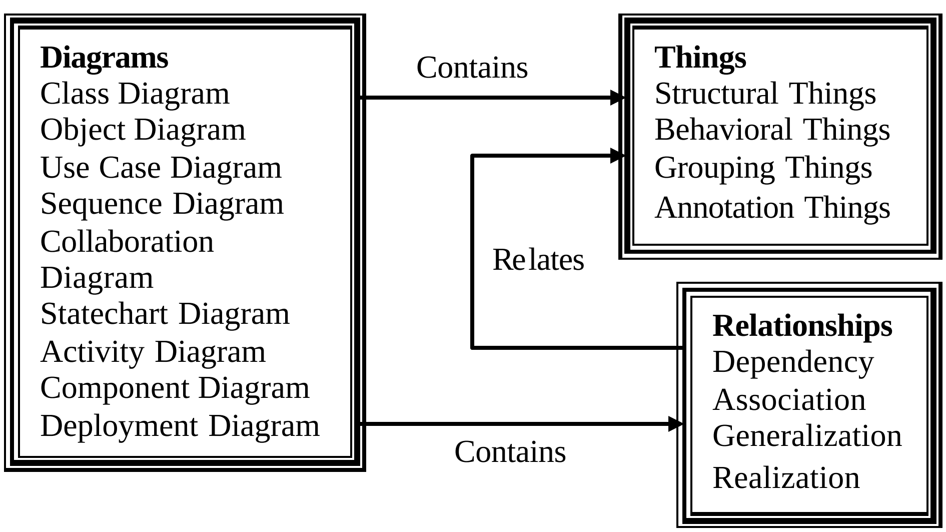


Figure 2-8 Three building blocks and their relationships

Originated from object-oriented modeling and analyzing, UML was originally designed to model software systems. There is no inherent support for modeling business processes. However, several features are potentially useful for doing so, such as activity diagrams and sequence diagrams. For example, using the activity diagram and its "swimlane" feature, a business process may be described as a series of activities with each activity fallen into a "swimlane" specific to a business performer. In addition, several research efforts in the last three years have used UML extensibility mechanisms to design a kind of "UML Extensions for Business Modeling" [Eriksson and Penker, 1999]. As an example, Figure 2-9 shows the symbols used in Rational Rose TM [2] CASE tool to present this kind of extensions.

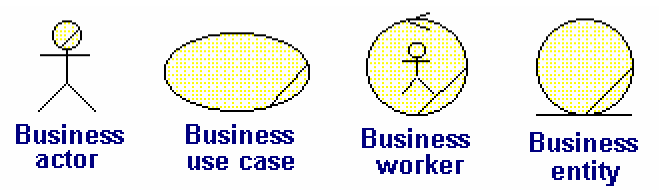


Figure 2-9 Symbols used in Rational Rose to present business components

UML is largely process-independent, meaning that it can be used with a number of software engineering processes [Booch et al., 1998]. The Rational Unified Process is one of these processes that have been applied to many domains such as building web applications [Conallen, 2000] and project management [Cantor, 1998]. Although these development processes were originally designed for software system development,

[2] Rational Rose is the trade mark of Rational Software Corporation, 18880 Homestead Rd., Cupertino, CA 95014, USA, http://www.rational.com/

almost all of them involve steps of modeling business surroundings of the underlying systems, providing useful hints in using UML to model business process or workflow.

Comparison

While UML has strong graphical support to allow visual presentation of models built based on its underlying semantics and notations, WPDL was intentionally designed as a text-based language to allow descriptions of models in an interchangeable format. The fact that WPDL workflow models comply with international workflow standards may offset its disadvantage of not having graphical representations. The trade-off of using the graphical capability provided by vendor's workflow process definition tools to compensate this defect is that all workflow models may be vendor specific.

WPDL allows modeling such business performance aspects as priority, cost, duration, waiting time and working time. While UML may be able to describe similar performance parameters by using stereotypes and tagged values, it inherently does not support capturing business performance measures [Dewalt, 1999]. By including too many stereotypes and tagged values, any UML-based modeler (software tool) may have difficulties in interpreting these extensions simply because the UML model itself has been modified.

Differences between UML and WPDL in modeling business process and workflow at the detailed description level have not been well studied. Currently no solutions have been proposed to reconcile these differences. Based on the literature review, there is also no software currently available in the market to allow the direct mapping between UML-based workflow models to WPDL-based models.

2.2.4 Groupware with the Internet and WWW

For the last three decades, groupware developments focused on providing tools to support proprietary systems within LAN or WAN environments [Coleman and Kbanna, 1995; Khoshafin and Buckiewicz, 1995; Finley, 1997]. Support for Internet-based network models, such as Intranet and Extranet, has not been evident until recently when many groupware vendors began to redesign or extend their products as Internet-enabled, or -based, applications.

In particular, significant developments of web-enabled or web-based workflow management systems have been seen that promise to facilitate workflow management within distributed work environments [Action, 1997; Plexus, 1998; Grather et al., 1997; Khoshafian, 1998]. In a web-based workflow, "*the Internet (usually web browsers) becomes both the way to initiate a transaction and the trigger for the process that will service it*" [WfMC, 1998]. Web browsers play the role of client interface and interact with the "back office" consisting of workflow engine, database server and web server, as well as other supporting applications. The initial web based workflow was mainly form-based and lacked the ability to manage tasks and administrative functions [Plexus, 1998]. However, using Java and other web technologies (e.g., XML) as development tools is showing the way to the next generation of web-enabled workflow [Plexus, 1998; and Koch, 1999].

While LAN still plays the major role for supporting internal collaborations, Intranet- and Extranet -based collaboration models have increasingly become popular alternatives due to the WWW. On top of the corporate LANs or WANs, these models use the Internet as their network infrastructure to connect remotely located, collaborating group members. However, these web-based network models had, and still have, the following problems faced by application developers.

(1) *Interactivity*: Existing interactivity capabilities of the web constrain the development of web-based application systems. Web browsers are only able to access files using HTML specifications. Interactivities of the web used to be supported only by forms, plug-ins, and CGI-scripts. Although Java, JavaScript, and ActiveX Controls all provide extra interactivities to web applications, not all browsers support the same scripts/languages. While both Netscape Navigator ™ and Internet Explorer ™ can support Java, functions based on JavaScript and ActiveX Control, for example, are more proprietary in nature. It is not likely that this problem will be resolved in the near future.

(2) *Performance*: The major drawback of using Internet-based applications is in the potential for performance issues [Action, 1997]. Although the Internet is ideal for application integration environments, the system performance can be questionable if too many different applications are integrated into this environment. The performance

of any individual component will be affected by others, hence, constraining the performance of the whole collaborative system. The results of research from LUTCHI Research Center [Shah et al., 1998] show that, given certain workstation processing power, the performance is greatly degraded if more than one synchronous groupware application are dependent on the Internet. In addition, the working time of the day also affects the processing speed because of the usage volume of the Internet (there are certain times when the Internet is extremely slow), which varies from country to country. Performance problems stem from two sources: overloading of the server and lack of the network bandwidth [Coleman, 1994a; Action, 1997].

(3) *Time Differences*: Common to any synchronous collaboration activities (whether over the Internet or WAN) where the participating parties are located at different time zones, difficulties in scheduling collaboration sessions affect both practical usage and efficiency of groupware tools that are used in synchronous mode.

Section 2.2 describes technical details that are essential for any application developments involving groupware components. Some issues mentioned, with respect to adopting, implementing, and evaluating these groupware components, are valuable for the collaboration model development later in this thesis. Although there are still problems encountered in terms of performance, time difference, and user interface interactivity, the practical use of the Internet-based groupware tools to facilitate collaboration activities are possible using currently available hardware and software tools and network infrastructure [Shah et al., 1998].

2.3 Considerations of Model Design and Performance Measure

As will be discussed in Chapter 4, the collaboration model for supporting collaborative GIS data production encompasses a workflow model, an architecture model of the collaborative workspace, and an implementation framework. While the methodology considerations regarding workflow modeling were discussed in Section 2.2.3.3 in detail, this section serves to examine issues regarding the system design of collaborative workspaces. Some performance measurement considerations are also discussed.

2.3.1 Approach to CSCW System Design

Principles and approaches used for designing groupware have been extensively discussed by many researchers (e.g., [Greenberg, 1991]; [Cockburn and Jones, 1995]; [Nunamaker, 1997]), among others. While a full discussion is beyond the scope of this research, the author briefly discusses some major issues that are relevant to the collaboration model development and collaborative workspace prototyping covered in Chapter 4 and Chapter 5.

Designing groupware systems is difficult and complex due to (a) the involvement of multiple disciplines and multiple methodologies in nature; and (b) the lack of high-quality theories synthesized so far in the field because of the lack of actual research experiences [Nunamaker, 1997]. Although traditional experimental and observational methodologies employed to study other human-computer interaction applications are often inadequate for designing CSCW applications [Greenberg, 1991], they have still been widely used in practice.

Recent research shows that "multidisciplinary", "interdisciplinary" and "participatory" are among the most important and promising principles characterizing many modern groupware designs [Muller and Kuhn, 1993; Cockburn and Jones, 1995; Nunamaker, 1997]. While methodologies based on these principles emphasize the role each related discipline plays and the input from end-users, traditional software engineering still plays an essential role in CSCW software development.

2.3.1.1 Software Development Process

The software life-cycle paradigm has been a fundamental software development process for many years. While its variations (e.g., waterfall model, prototyping, and object-oriented) have been applied in various software systems development, the fundamental process consists of a series of phases including requirements analysis, requirements definition, architecture and component design, implementation, system testing, and operation and maintenance. With the careful selection of methods and tools that can be used in order to achieve the goal of each phase, the paradigm ensures the sequential accomplishment of the goals of all phases through a set of activities in each phase [Bischofberger and Pomberger, 1992].

For the last two decades, a prototype-oriented software development approach has been widely adopted in developing software systems [Bischofberger and Pomberger, 1992]. With prototyping, the process allows incremental feedback from users and has the potential to overcome such problems as lack of iteration between phases that occurred in traditional software life-cycle approach. The approaches to prototyping are normally classified into three categories:

(1) Exploratory prototyping;

(2) Experimental prototyping; and

(3) Evolutionary prototyping.

While exploratory prototyping approaches focus on obtaining a set of as complete as possible requirement definitions, evolutionary prototyping approaches seek an incremental development mechanism to evolve the prototype into a final software (or system) product. Experimental prototyping is an approach that supports system and component design by achieving a concise specification of the components that form the system architecture [Bischofberger and Pomberger, 1992]. None of these prototyping approaches are fundamentally different from the sequential software life-cycle approach. In fact, it is the software life-cycle approach that provides the basis for prototyping approaches.

Figure 2-10 shows how each prototyping approach fits into the sequential software life-cycle paradigm by taking over the rest phases after obtaining initial requirement definitions. For prototyping paradigm, the prototype may be constructed as (a) complete prototype, (b) an incomplete prototype, (c) a throwaway prototype, or (d) a reusable prototype, depending on project scope, schedule and details required. The experimental prototyping approach focuses on reusable prototypes, design and implementation information, which is of great concern for the proposed collaboration model development. Therefore, this approach will provide the necessary base knowledge for developing the research methodology adopted for later work of this thesis.

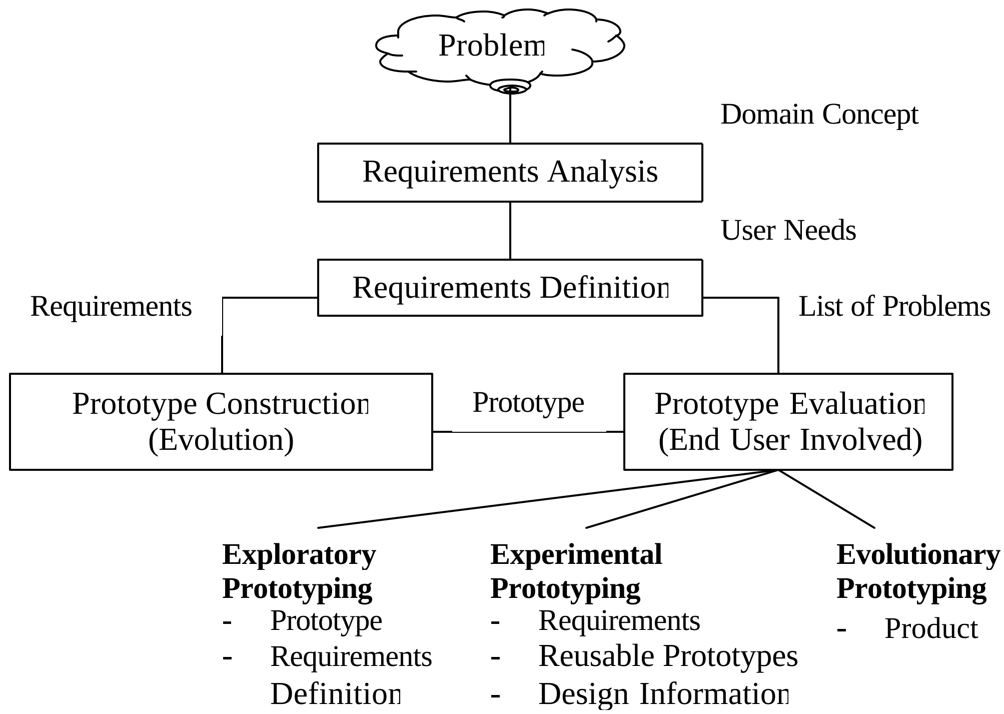


Figure 2-10 Software development using a prototyping paradigm (Source: [Bischofberger and Pomberger, 1992])

2.3.1.2 User's Impact on the Design

Beyond methodologies adopted, both social and individual impacts have to be taken into considerations when designing groupware systems. Cockburn and Jones [1995] argued, "*Society's rejection of groupware is driven by an accumulation of individual rejections*." Based on this observation, they analyzed various causes of groupware failure in reality and presented four principles of groupware design to help reduce these impacts, including: (a) maximize personal acceptance; (b) minimize requirements; (c) minimize constraints; and (d) maximize external integration.

While causes of failure in CSCW applications may be at the system-use, system-design, or system-evaluation level [Grudin, 1988 and 1994], it is those causes at the system-use level that indicate the importance of end user involvement in CSCW application design process. Some outstanding causes of groupware failure at the system-use level are summarized as follows [Cockburn and Jones, 1995; Grudin, 1988 and 1994]):

(1) *Additional effort required from users to support groupware functionality*: Proper and sufficient guidance should be provided to allow reduction of this effort, e.g., effort of entering structured information or following structured processes.
(2) *Lack of flexibility*: When applied with inefficient, restricted and inflexible working practices, the system is not very adaptable.

(3) *Effort imposed by lacking integration*: Inadequate integration between groupware and other computer tools often requires extra effort from users to make up the work environment.

(4) *Benefit and benefit-lag*: A long period during which the effort put into mastering and using a system out-weighs the benefit received.

(5) *Benefit Disparity*: Disparity exists between the person who does the work and the person who gets the benefits.

Considering the importance of end-user involvement in designing the CSCW application systems, this research used a participatory, prototyping approach to develop the proposed collaboration model, as well as the prototyped collaborative workspace called *GeoPM* (see Chapter 5 for details).

2.3.2 Performance Evaluation

The process of evaluating CSCW systems involves effort at two levels. On one hand, groupware tools developed or utilized as software or software components have to be evaluated from a software engineering perspective to determine if the interfaces, usability and performance follow the predefined software specifications. On the other hand, the "collaboration" introduced into organizations as a result of adopting these software tools has to be evaluated based on the business objectives, predefined business requirements and other generic expectations. While methodologies for evaluating groupware software or software components have been well developed in the fields such as software engineering and human-computer interaction (HCI), evaluating collaboration still remains a big hurdle for CSCW evaluators due to the diverse fields involved and business variations [Ross and Rogers, 1995].

Many researchers have principally evaluated CSCW collaborations on the basis of collaboration outcomes as well as factors affecting these collaboration outcomes (e.g., [Ross and Rogers, 1995]; [Bruce et al., 1995]; [Baeza-Yates and Pino, 1997]). However, the perceived positive outcomes – benefits – vary from one case to another. Baeza-Yates and Pino [1997] presented three broad categories of benefits that may serve as a framework to synthesize positive outcomes as follows:

(1) *Improved Outcomes*: better results through collaboration such as documents, designs, products, or decisions. The improvements should be identified based on comparison with outcomes from non-CSCW supported systems.

(2) *Individual and Group Gains*: benefits received by individual members of the group and by the group as a whole.

(3) *Efficiency*: reduced wasted time, resources and/or duplicated efforts for people to effectively contribute to a joint effort.

An important aspect of evaluating collaboration is to define "success", delimiting the boundary between successful and unsuccessful collaborations. However, it is notoriously difficult to make uniform definitions of success in collaboration supported by CSCW systems [Dodgson, 1993]. In practice, "success" is very often defined in terms of whether the collaboration has met its original objectives [Bruce, et al., 1995]. The objectives may contain technical measurement metrics such as efficiency and productivity increments.

When process automation – workflow – is involved in collaboration efforts, process performance has also to be measured. Process performance evaluations have long been conducted in the business analysis domain and well documented in many publications (e.g., [Darntons, 1997]). While measurements may be at various levels such as normal, ordinal, interval and ratio, the general metrics remain the same as efficiency and effectiveness. While efficiencies are more likely evaluated "technically" based on quantitative values and monetary values, effectiveness is often indicated by the resource inputs needed to produce a level of the enterprise or project objective.

The complexity of evaluating CSCW-based collaborations sometimes may be further amplified by contradictory outcomes of evaluations, as illustrated in the following two examples:

(1) Nunamaker [1997] summarized in his paper that "Teams using group support systems have reduced their labor costs for a project by an average of 50% and the number of calendar days required for a project by an average of 90%."
(2) Bruce et al. [1995] found that "over 40% of respondents expressed the view that collaboration made product development more costly, more complicated, less efficient, more time consuming and more difficult to control and manage." This observation was based on a survey (mail questionnaire to 300 UK companies with

106 responding) and case studies of eight companies in information and communication technology sector.

Given the rationale discussed above and time constraints, the performance evaluation focused on increased outcomes (productivity) and improved efficiencies compared with the existing GIS data production systems (see Chapter 5 and Chapter 6 for details). While increased outcomes may be evaluated with hard numbers (e.g., overall number of failed data files in first submission), they are generally hard to quantify. However, the efficiency can be measured quantitatively by two kinds of metric: cost and time.

Although difficult to quantify some costs (for instance, savings that may be realized in maintenance [Finley and Coleman, 1999]), most of the incurred costs are measurable. Examples of measurable costs include money spent on hardware, software, workspace setup and maintenance. Time efficiency is more difficult to quantify than cost efficiency because it is affected by some unforeseen factors such as randomly increased network traffic. In order to assess time efficiency, the process should be divided into *n* ($n>1$) small measurable units that are task-based [Baeza-Yates and Pino, 1997]. The time elapsed on each individual unit and the total time span on the whole process should be significantly decreased comparing with the time spans of existing production systems.

2.4 Review of Closely-Related Research

Interests in applying formal CSCW into the geomatics arena fall into two major different but inter-related areas since 1993. On one hand, collaborative efforts have been directed toward either GIS-based or groupware-based spatial decision-making into planning, environment management and other decision-making tasks, and public participation into the planning process. Examples include the work by Faber [1994; 1997; 1998], Jankowski [1995]; Densham et al. [1995], Klein [1998], and Finley et al. [1998a; 1998b]. On the other hand, other efforts, as documented in Finley [1997], Finley and Coleman [1999], Churcher et al. [1996; 1997], Li et al. [1998], and Coleman and Li [1999], among others, have applied some of the same tools and concepts to geomatics production management and editing applications.

Since 1994, geomatics organizations around the world have shown increasing interest in the Internet. Many of them either maintain their own web sites, selling their digital spatial data and products, or disseminate shared spatial information [Dawe, 1996; Hardie, 1998]. Internet-based spatial data warehouses and online analytical processing (OLAP) have become increasingly prevalent among large organizations [Bedard, 2001]. However, for most geomatics production organizations, the major Internet-based applications supporting production processes continue to be email and file transfer services [Finley, 1997]. There is little evidence these organizations have adopted more sophisticated Intranet or Extranet services to support routine geomatics production or project management.

Although only limited research efforts have been found in geomatics, similar research projects are involved in applying CSCW concepts to facilitate collaborative manufacturing, managing construction projects and inspecting buildings. Since the outcomes from the research and methodologies they adopted are closely relevant to this research, they are also briefly discussed.

2.4.1 Research in Geomatics

Commercial efforts such as TerraShare ™ from *Z/I Imaging* (http://www.ziimaging), GIS Design Server ™ and Project Point ™ from *AutoDesk* (http://www.autodesk.com/) have emerged during the latter period of this thesis research. TerraShare serves as a modular, client-server system designed to address image management and distribution needs of geo-imaging producers and distributors. ProjectPoint allows the project team to better collaborate on a project through sharing project documents and communications centralized on one secure location, while GIS Design Server enables a collaborative spatial data warehousing solution.

These more recent efforts aside, the results of an extensive early literature review indicated few documented research efforts dealing with collaborative GIS data production in a distributed environment where multiple participants are involved. The previous work done at the University of New Brunswick (UNB) focused on needs analysis of collaborative GIS data production to define collaboration requirements, comparing different network models and examining functionality of existing groupware tools [Coleman and Brookes, 1996; Finley, 1997; Li, et al., 1998; Finley and Coleman, 1999; Boettcher, 1999]. These research activities are further elaborated as follows:

1. Examining and defining collaboration requirements when operating and managing GIS data production projects within a distributed work environment based on certain contracts.
2. Examining selected groupware packages/systems in the market to identify available supporting functions or functional components to establish a possible mapping between requirements and functions.
3. Pilot testing with a specifically selected groupware package.

Through an extensive examination of a specific GIS data production project (ETB'95 – Service New Brunswick) carried out in a distributed environment, Finley [1997] identified several major project problems and addressed potential bottlenecks and issues regarding business process redesign. These problems were then mapped to a CSCW framework defined in his work to form the discussions on design considerations of a collaborative workspace. However, the actual design and implementation of the proposed collaborative workspace was developed to ease only information dissemination and exchange for the New Brunswick Costal Zone Steering Committee, which was a quite different environment than those of data production projects. In addition, the prototyped collaborative workspace barely touched groupware techniques other than discussion forum as described in his work.

In addressing business redesign strategies, Finley found that current distributed projects were largely sequential in nature and concurrent processing may be able to improve production efficiencies through proper process redesign. However, these findings came from an examination of mostly a specific project without using any process modeling principles that is considered critical to business redesign. Although Finley's work was more or less connected to a specific project, it still provided valuable insights into the nature of collaborative data production for any follow-up research efforts.

Compared with Finley's work, Boettcher [1999] tackled similar research from a different angle. Instead of being technologically driven, Boettcher and his colleagues examined the collaboration requirements of collaborative GIS data production environments based on an initial user needs analysis and mapped them to a list of functions identified from some groupware packages available in the market. Selected groupware tools were then reviewed to determine if current groupware technology was sufficient to support collaborative GIS production. Boettcher also addressed the design and implementation issues of the collaborative workspace. However, his discussions were constrained with a set of selected groupware tools (mainly from Netscape SuiteSpot 3.0 – 3.5).

UNB Research on Collaborative GIS Data Production

In 1995, the Geographical Engineering Group at the University of New Brunswick initiated a series of investigations into the feasibility and the implications of using CSCW concepts and technologies to support GIS data production processes in distributed work environments. The original objectives of initial research efforts included:

(1) To conduct preliminary research which: (a) examines the characteristics, strengths and weaknesses of existing sequential production models; (b) develops a prototype collaborative production model (or models) for digital map and chart production in self-contained and distributed operational environments; and (c) identifies hardware-, software- and operational constraints to collaborative production;
(2) To develop specification and prototype software that enables collaborative production, inspection and correction of digitized map and chart files in a wide-area network environment;
(3) To test performance of these software packages across a broadband, wide-area network service (in comparison with stand-alone and LAN-based systems) to begin to identify optional approaches to collaborative data production and delivery;
(4) To identify and classify collaborative production tasks which (a) absolutely require broadband connection to be carried out; (b) may be acceptably completed across lower-speed services now enjoying widespread usage; and/or (c) those which may be temporarily redesigned to be handled on lower-speed links; and
(5) To demonstrate and share results through a combination of prototype demonstrations, high-level presentations and published papers and reports.

As discussed above, the initial research efforts by David Finley, Robert Boettcher and Songnian Li under the supervision of Dr. David Coleman addressed the first, second and last objectives listed above. Especially for the prototyping software, it was tested with both selected software packages (e.g., Netscape SuiteSpots ^{TM}) and selected applications (e.g., New Brunswick Coastal Zone Steering Committee Prototype). Results demonstrated the lack of a CSCW model for developing and implementing collaborative workspace to help:

(1) To determine if the Internet-based groupware tools can facilitate the improvement of efficiency and productivity of collaborative GIS data production; and more specifically

(2) To fulfill the third and fourth objectives described above.

Starting in 1998, the Geographical Engineering Group at UNB initiated further investigations into the development of an Internet-based CSCW model to support distributed GIS data production management and operations. The original objectives of this research were:

(1) To further define functional requirements needed by the collaborative workspace in supporting collaborative GIS data production;

(2) To design an Internet-based geomatics production collaboration model which includes: (a) a workflow model; (b) a workspace architecture; and (c) an implementation framework to facilitate the development and implementation of the collaborative workspace;

(3) To prototype a workable collaborative workspace based on the designed model by integrating functional modules or software components provided by existing GIS and groupware packages; and

(4) To test and/or simulate the performance measures of the prototyped collaborative workspace and refine the collaboration model that is feasible for being applied within a real-world geomatics production environment.

The research was under the supervision of Dr. David Coleman, funded by the GEOIDE Network of Centres of Excellence[3]. The project obtained substantial in-kind support from outside organizations such as Service New Brunswick, DataQC and WaterMark Industries Inc. in New Brunswick, Canada.

2.4.2 Research in Other Related Areas

Several research efforts using Internet-based groupware technologies in production areas can be found from the literature (e.g., [Newton et al., 1995]; [Kao and Lin, 1998]; [Shin et al., 1996]; [Rojas, 1999]). However, they involve solving different production problems in the domain of manufacture, architecture and construction, medical imaging, and project management, among others. These research efforts are briefly summarized in this section.

[3] Part of GEOIDE project DEC#2 – Designing the technological foundations of geospatial decision making with the World Wide Web (http://sirs.scg.ulaval.ca/geodem/).

One area is in managing architecture, construction, and other projects as well as sharing their related information over a wide-area network infrastructure. Rojas [1999] presents a development model that is illustrated by a web-centric system that supports construction inspection, called "Field Inspection Reporting System". The model concentrates on a system development procedure rather than modeling the actual workflow to answer, "How the process does it". The system adopted a simple web-based client/server structure with the server application containing HTML pages, CGI scripts, C^{++} applications, and a relational database. There is no systematic process control and automation mechanism applied. The research focus was on project information sharing. While Shin et al. [1996] present their research on monitoring and controlling business processes through integrating workflow technology and project management techniques, their research focused on using existing workflow management systems and project management software by integrating them through "object mapper" and a kind of "internal control". The working team uses the native interfaces of the integrated software to interface with the collaborating system.

Others [Newton et al., 1995; Kao and Lin, 1998] study the use of Internet-based groupware tools, especially electronic whiteboarding or electronic conferencing with a shared view, in supporting collaborative engineering design and manufacture, such as collaborative CAD/CAM, and medical imaging. Kao and Lin [1998] developed a collaborative CAD/CAM system for CAD-geometry co-editing, design, and manufacture, which extends a traditional single-location CAD/CAM application to be operable over the Internet by two geographically dispersed CAD/CAM users. The system structure is flexible in the sense that it can either use API with network library routings from the existing CAD/CAM systems or an external network module to fulfill the lack of network communication channels of the existing CAD/CAM systems. Kohli [1995] studied a number of the broadband communication services opportunities for the medical industry. The focus was on how current broadband network infrastructure can be used to support medical imagery sharing and group-based diagnosis discussion according to images captured by various existing medical imaging equipment.

Although all these research efforts are relevant to this thesis research, obviously, none of them is directly related to collaboration activities in geomatics. Little has been reported about the practical use of existing groupware tools. In addition, the workflow of the design process within distributed environments is not well considered in these research efforts. However, these research results will provide very valuable inputs for this thesis research.

Development of any collaboration model requires complete understanding of the work environment involved, underlying business processes and supporting technologies. While this chapter was designed to help the reader to establish understandings regarding supporting technologies and general business environments, the next chapter will look into details of the GIS data production processes and collaboration requirements, which forms the basis for the model development, prototyping, and performance simulating and analysis discussed in Chapter 4 and Chapter 5.

Chapter 3 GIS DATA PRODUCTION: DEFINING REQUIREMENTS AND MODELING EXISTING PROCESSES

The success of designing and deploying an Internet-based collaborative workspace to support distributed GIS data production project management and production operations largely depends on a comprehensive understanding of the following:

(1) the data production work environments surrounding the collaborative workspace – including the underlying management and production processes, resources involved, and, most importantly, the goals of conducting this business; and

(2) the requirements (both functional and non-functional) identified for the collaborative workspace to be able to support actual project and/or production processes.

To reach this understanding, it is necessary to examine current GIS data production projects using formal business modeling techniques. While the original purpose of conducting this type of business modeling was to provide substantial understandings about the existing distributed GIS data production practices, the fundamental understanding necessary for developing the proposed collaboration model may be established by:

(1) building necessary terminology and knowledge for the research;

(2) examining existing management and production processes and changes that could be made of these processes for the successful implementation of the collaborative workspace, which provides fundamental inputs for developing the workflow model in supporting the process automation and control;

(3) identifying necessary collaboration and information support, as well as key functional and non-functional requirements of the supporting collaborative workspace, which formed the essential basis for its architectural design; and

(4) reviewing functionality available from existing groupware packages in the market and identifying the gaps between the functionality required by collaborative GIS data production and the functionality available in the market.

The following sections introduce the information sources for modeling current GIS data production projects and describe the results from examining the existing project management and production processes. Considering production process modeling as the centre of the whole business modeling [Eriksson and Penker, 1999], the effort of modeling individual behaviors of important resources and processes and how they interact with each other with respect to the production process are described. The author then discusses how the requirements specification of the collaborative workspace was identified based on the results from these previous activities, followed by a review study of current groupware functionality.

3.1 Information Sources

Inputs for modeling current distributed GIS data production practices were obtained from the geomatics data production industry through empirical studies based on interview, workshop, and interactive presentation techniques with intended users (as defined shortly), site visits and demonstrations of existing CSCW software capabilities and behaviors. Work materials were also collected (e.g., technical specifications, production manuals, procedures, work sheets, sample reports, and official correspondences), and some samples of these are included in Appendix A.

Participation of potential users is important to any software-related development. The following criteria were used to select such a group of people for the purpose of collecting information, discussing existing production practices, and obtaining feedback:

(1) people who have obtained previous experience through working with any distributed GIS data production projects at both management and operational levels;
(2) people who were currently working on similar production projects; and
(3) people who have had experience of conducting similar research dealing with applying CSCW technologies in supporting geomatics data production.

One might argue about this last criterion because these people do not really belong to the end-user community. However, the author believes that feedback from this group of people is very important for the research development. This is especially true when considering the fact that CSCW is still not a mature discipline in terms of its research methodology [Nunamaker, 1997] and the fact that little evidence of adopting formal CSCW technology in geomatics industry has been found so far.

Because this thesis research focused on developing a "*generic*" collaboration model that is not bounded with any specific production projects, the intention was to model various distributed data production projects to obtain more general understandings of current business situations. With this in mind, several organizations – including three private companies and one government agency in Canada – were selected to gather the necessary information. More detailed information was then obtained from Service New Brunswick (SNB) and WaterMark Industries, Inc. because of access to some information concerning several multi-participant, distributed data production projects (contract-based) conducted in the last few years.

Table 3-1 shows a list of these selected organizations and companies. The nature of each organization, the major role played in distributed data production projects, and example projects involved are briefly described. This described information shows strong relevance of these organizations to this research.

Table 3-1 Organizations and companies contacted

Organization	Relevance to this Research
Service New Brunswick Fredericton, NB	A government agency responsible for the mapping of the whole province, involved in ETB'96, ETB'98 and orthophoto map projects as ***client***.
WaterMark Industries Inc. Fredericton, NB	A private company specialized in geomatics project management and data quality control, involved in ETB'96, ETB'98 and orthophoto map projects as ***project manager*** and ***QC inspector***.
InterMap Technologies, Inc. Ottawa, ON	A private company specialized in mapping by photogrammetry and remote sensing, involved in many mapping contracts as ***production contractor***.
Terra Surveys Ltd. Ottawa, ON (Now Triathlon Mapping Corporation)	A private company specialized in surveying and mapping, involved contracted mapping project as primary ***production contractor*** connecting with sub contractors in UK.

Preliminary studies conducted by the Geographical Engineering Group at the University of New Brunswick identified some initial requirements for collaborative GIS data production systems from different angles (see Section 2.4 for details). This information was also considered as important sources for this thesis research.

3.2 Existing Production Processes

A "business process" consists of a set of activities that associate with certain resources and may change the state of these resources where resources are objects within the underlying business such as people, work materials, information and products [Eriksson and Penker, 1999]. Objects from resources are related to each other based on business structures. To better model GIS data production project processes, all resources involved and their structural behaviors have to be examined.

This research specially examined the underlying organizational structures, business workers involved, work materials passed along the management and production processes, and project-related information required to run these processes. While organizational structures and the associated business workers are discussed in a separate section that follows, other resources are described along with the process modeling.

3.2.1 Organizational Structures

Unless a client has a small-scale data production project – or opts to only outsource the production part and uses its in-house expertise to perform both project management and data quality control tasks – contract-based data production projects usually involve organizations and companies acting as client, project manager, production contractor, and QC inspector, respectively. In some cases, a production contractor may further contract part of their jobs to third-party companies with consent of the client. These third-party companies act as sub-contractors in the contract. The following gives brief descriptions of each of these functional roles:

1. *Client*: represents a functional role that outsources GIS data production, quality control and project management tasks. The client prepares and delivers source data materials and technical specifications to contractors for data production and receives the quality-assured data products at the end of the contract.
2. *Project Manager*: represents a functional role of managing the overall production project to control the project schedule, monitor data production progress, resolve production problems involving multiple parties, and manage data and QC specification updates. In most cases, the project manager is responsible for getting source data materials, ensuring data product quality, and delivering the final data products.
3. *Production Contractor (called primary contractor when subcontractor involved)*: represents a functional role responsible for processing source data materials based on

data specifications and ensuring that all processed data passes through a data quality control (Q/C) process. In cases where subcontractors are involved, the production contractor is responsible for delivering source data materials to and receiving processed data from sub-contractors, as well as resolving data correction issues with sub-contractors.

4. *QC Inspector*: represents a functional role responsible for ensuring that all processed data materials meet the contract specifications by conducting quality control procedures and reporting QC inspection results to the project manager.
5. *Subcontractor*: represents a functional role to which part of work from the production contractor has been sublet, to do the actual data processing job. Subcontractors communicate with the respective production contractor regarding data delivery and submission, as well as data corrections if it fails QC inspection both at the production contractor level and the QC inspector level.

Currently, there is no one-to-one mapping between these functional roles and the participating parties, meaning that each party may play more than one role. The number of participating parties and the number of functional roles each party assumes vary from one project to another, depending on factors such as the nature of the data production project and the timeline required to deliver the final data products, among others. Figure 3-1 illustrates the four principal organization structures found from the existing distributed GIS data production projects.

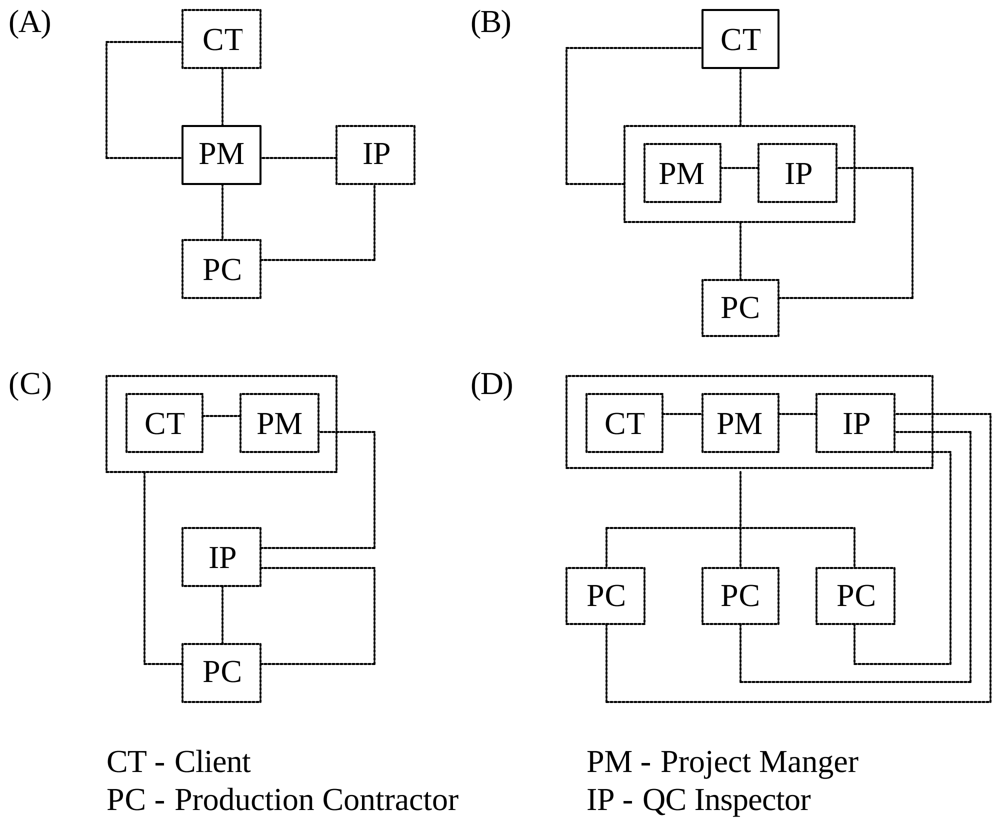


Figure 3-1 Organizational structures in distributed GIS data production

In Figure 3-1, labeled rectangles denote functional roles. Solid lines with arrowheads represent the major data flow along with necessary communications between functional roles. Dashed lines indicate internal communications and dashed rectangles denote organization boundaries. In all cases, subcontractors may be connected to the production contractor. Clearly, when more than one functional role is assumed by one participating organization, the overall production process is simplified.

Among these structures, the one in Figure 3-1 (A) was adopted in many GIS data production projects to accomplish SNB contracts. Compared with others, this structure allows the project manager to have more control of the flow of data materials, which is perceived to be very important by project managers. Yet the structure presents a more general organizational view of the existing production projects. Since this structure was used in modeling collaborative GIS data production projects, an extended organizational view with sub-contractors is described with an UML object diagram illustrated in Figure 3-2. Each participating organization is specified as a structural object of organization type, with the assumed functional roles described as attributes and several responsibilities described as operations of the object.

Since internal organizational structures vary from one company to another, the internal business roles may be defined very differently. For example, one production company may have such business roles as shipment, production, internal inspection, and production manager. Another company may only have shipment, production and production manager. These internal business roles are usually associated with the internal business processes. Since this research focused mainly on interactions between project participants to realize the overall project objectives, detailed internal structures were not modeled. However, this will be addressed with respect to how to handle internal processes in modeling the overall project process in the next section.

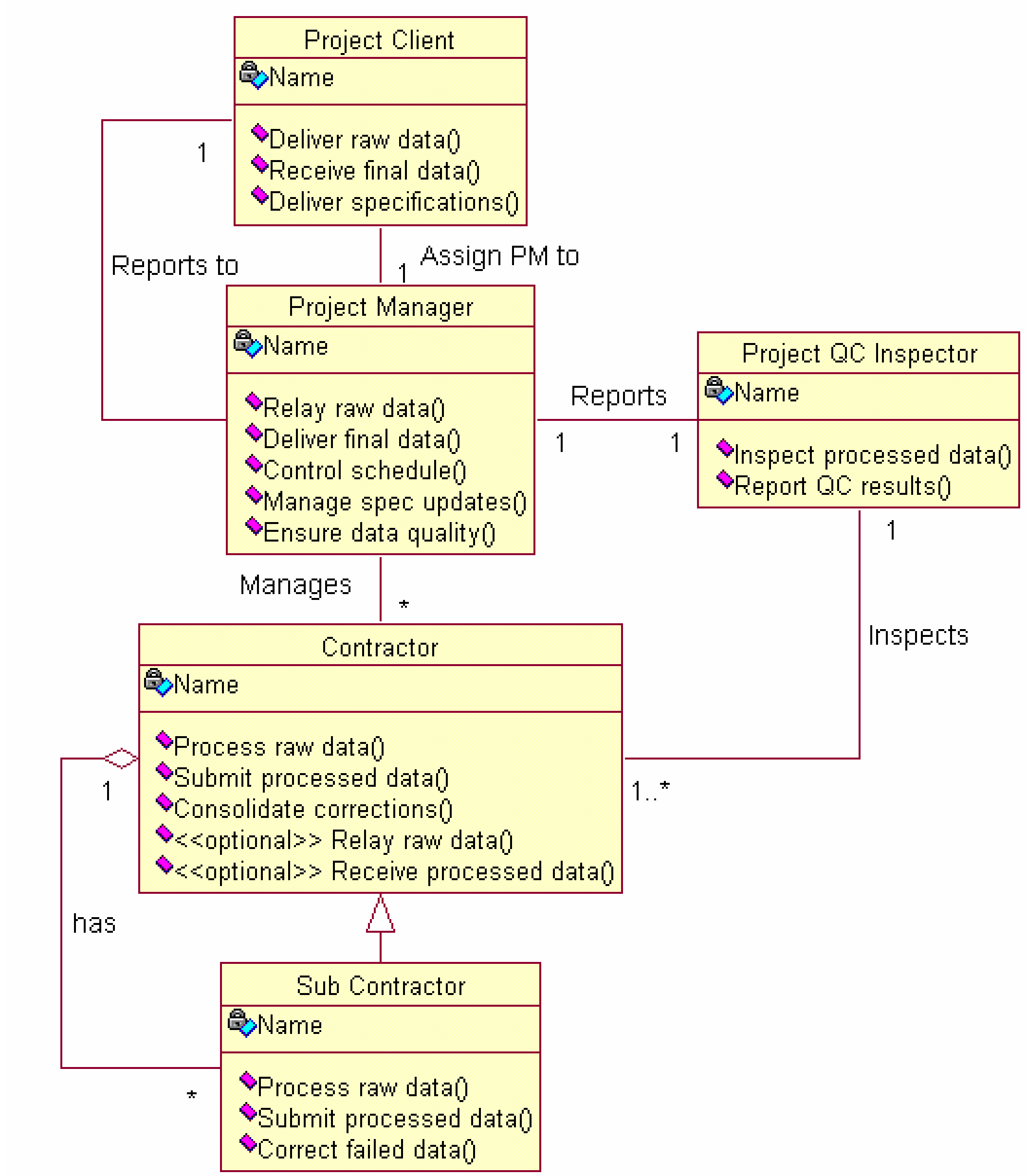


Figure 3-2 An organizational model for a distributed GIS data production project

3.2.2 Production Process Model

Generally speaking, a GIS data production life cycle includes such processes as raw data materials acquisition, data production, data quality control (QC) and initial data distribution. While raw data may be acquired from field data collection, aerial photos, or remotely sensed images, the actual acquisition processes were not considered in this research.

Given this assumption, the rest of this section discusses the existing processes that constitute the overall GIS data production project process in distributed work environments, how these processes fit into the organizational structure discussed in the previous section, and the respective nature and behaviors of resources involved in these processes. The ultimate goal of modeling existing production process is twofold:

(1) to facilitate the identification of collaborative workspace requirements from the process perspective to ensure that the identified requirements satisfy the needs of process execution; and

(2) to provide essential blueprints for modeling workflow of collaborative GIS data production - for example, obtaining the insights on which process should be fully automated or semi-automated, identifying bottlenecks, and determining how participants communicate with each other during the process to solve both operation and management problems.

Identification of Processes and Resources Involved

GIS data production projects, similar to other projects in general [Shih et al., 1996], consist of two types of processes: project management processes and production processes. ***Project management processes*** include those that encompass activities on project management and/or coordination aspects and whose objectives are to satisfy ad-hoc client requirements. ***Production processes*** include those repeatable, well-established processes that encompass activities on GIS data handling aspects. These production processes at a higher level are categorized according to two criteria: (a) each process should be well isolated in such a way that only one functional role is responsible for it; and (b) the process should be the essential part of the overall data production process. As such, the production processes include source materials preparation, source materials dispatching, data production and correction, data inspection, and final data delivery. Table 3-2 provides a typical mapping between these production processes and functional roles in contract-based production projects.

Table 3-2 Mapping functional roles to processes

Process	Functional Role
Source materials preparation	Client
Source materials dispatching	Project Manager
Data production and correction	Production Contractor
Data inspection	QC Inspector
Final data delivery	Project Manager

* A subcontractor only becomes involved when the production contractor sub-lets some of the production tasks.

While each process consists of a set of activities or sub-processes at a lower level, it also requires various resources as input and output, to control and to be supplied for the process. To capture all these resources, the UML Business Extension from Eriksson-Penker [1999] was used with UML class diagrams[4]. With this extension, the "process" is in the centre to capture required resources including work objects to be processed, work objects processed, information required to process work objects, business roles to control process, and business roles to process work objects. Figure 3-3 shows the major required resources captured for each identified process. Process activities will be described shortly in the ***Process Description and Structural Behaviors*** section.

The contract-related documents such as contracts, technical specifications, and QC procedures did not flow from one process to another since all participating organizations already had a copy of them prior to commencement of the production life cycle. While specification updates were usually made over the life of the project, they were distributed to all the participating parties as separate documents in different routines, usually by surface mail or courier.

[4] UML activity diagrams allow placing UML things such as objects in the diagrams to express dependency relationships between activities and things, where a process can be specified as an activity stereotyped as "*process*". However, the Rational Rose software does not support the expression of these relations.

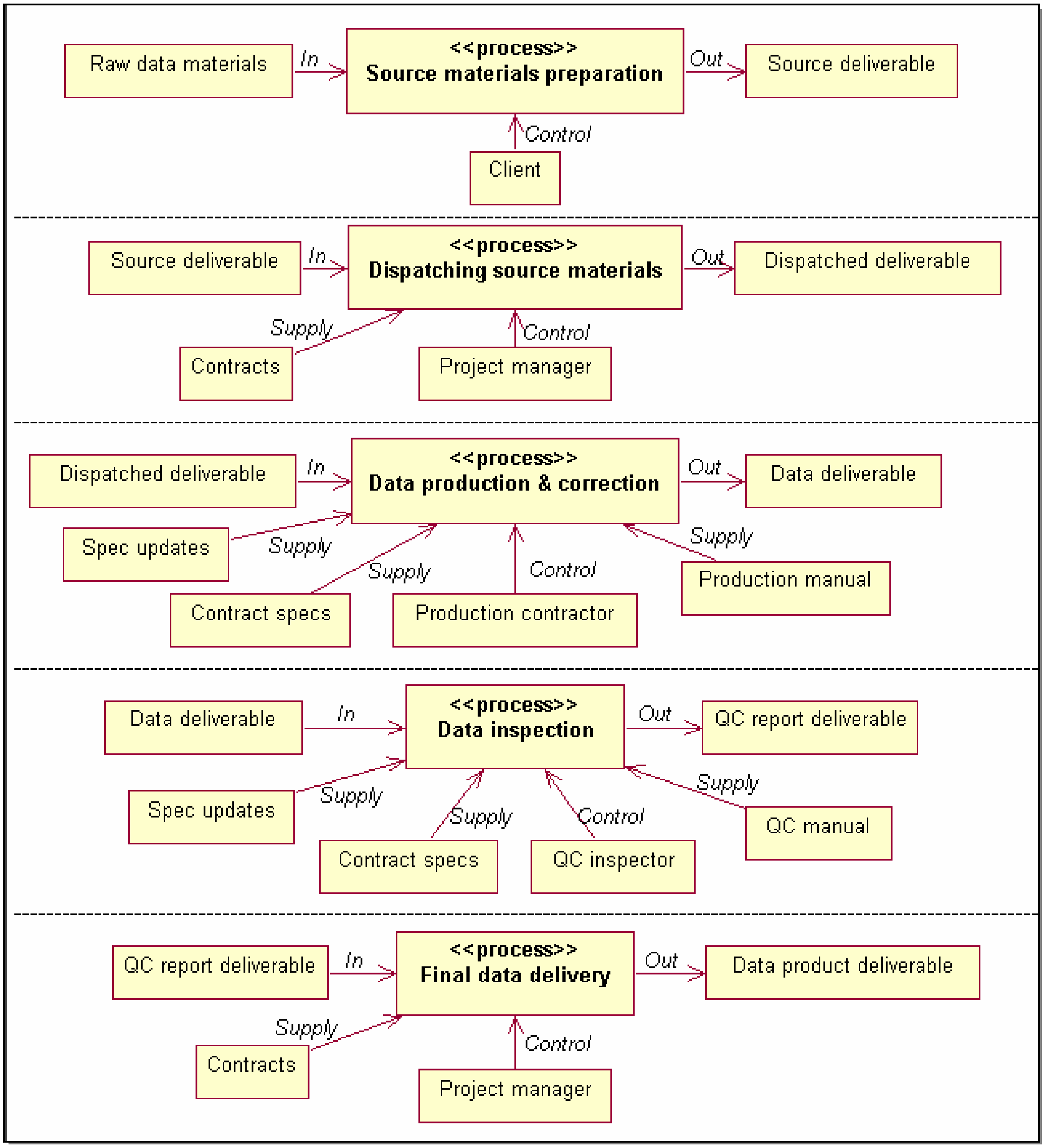


Figure 3-3 Five identified processes and major resources involved

The work materials passed on between processes were usually packed into various deliverables, which are passed as input resources for the next process except for the first and last processes. No matter what forms (source data, processed data or final data), GIS data comprised the bulk of these deliverables and were usually stored in digital files with each file covering a certain geographical area at a specified scale. For example, SNB projects windowed source data into 1:10000 windows stored in several files. In this sense, the processes identified from the past data production projects were really "*data-centric*". Mostly, deliverables were shipped in several shipments due to the large number of data files handled by the project. The typical contents of these deliverables are described as follows:

(1) *Source Deliverable*: including source data files to be processed, documents (e.g., hardcopy maps and tabular coordinates), and source data specifications;

(2) *Data Deliverable*: including processed data files, hardcopy plots of processed data, and production reports (include internal QC results), as well as source materials supplied by the client;

(3) *QC Report Deliverable*: including quality control reports, marked data files or hardcopy maps if they fail QC inspection, and data files if they pass QC inspection, as well as source materials supplied by the client; and

(4) *Data Product Deliverable*: including final data files in the format specified by the contract, delivery information, quality control reports, hardcopy maps (plotted in the form specified by the contract), and all source materials supplied by the client.

The documentary resource objects involved in these processes are briefly described as follows (see Appendix A for some example documents collected):

(1) *Contract*: including legal agreements, technical specifications, production procedures and QC procedures, controlled by the client and available in hardcopy forms.

(2) *QC Report*: including both final and failure reports generated by the QC inspector based on the QC results from various inspection steps and quality control reports required from the production contractor. In the latter case, report templates are sometimes used.

(3) *Project Report*: periodic reports on project status (e.g., weekly progress of data production) consisting of data such as percentage of completion, number of data files passed or failed QC inspection, etc. Theses kinds of reports are required by the project manager and the client, and are provided by either production contractors or the QC inspector. It was sometimes created using *spreadsheet* software based on information collected manually from various reports and forms.

(4) *Official Correspondence*: including letters and memos that were exchanged on specific issues.

(5) *Hardcopy Map*: plotted at various stages over the project span for the purpose of visual examination, marking errors and inconsistencies, providing source data and displaying final products.

(6) *Source Data Specifications*: technical specifications for the source data, supplied by source data providers.

(7) *Status Map*: using various color schemes and symbols on index maps typically to express project progress and general status information such as QC status, production completions and data file distributions (see Appendix A for samples). Status maps were normally plotted on hardcopy in poster size using any GIS software and used by

participants at various levels. For example, the project manager used them to display overall project progress (e.g., blocks completed) and, as illustrated in Figure 3-4, the QC inspector used them for QC passes and failures.

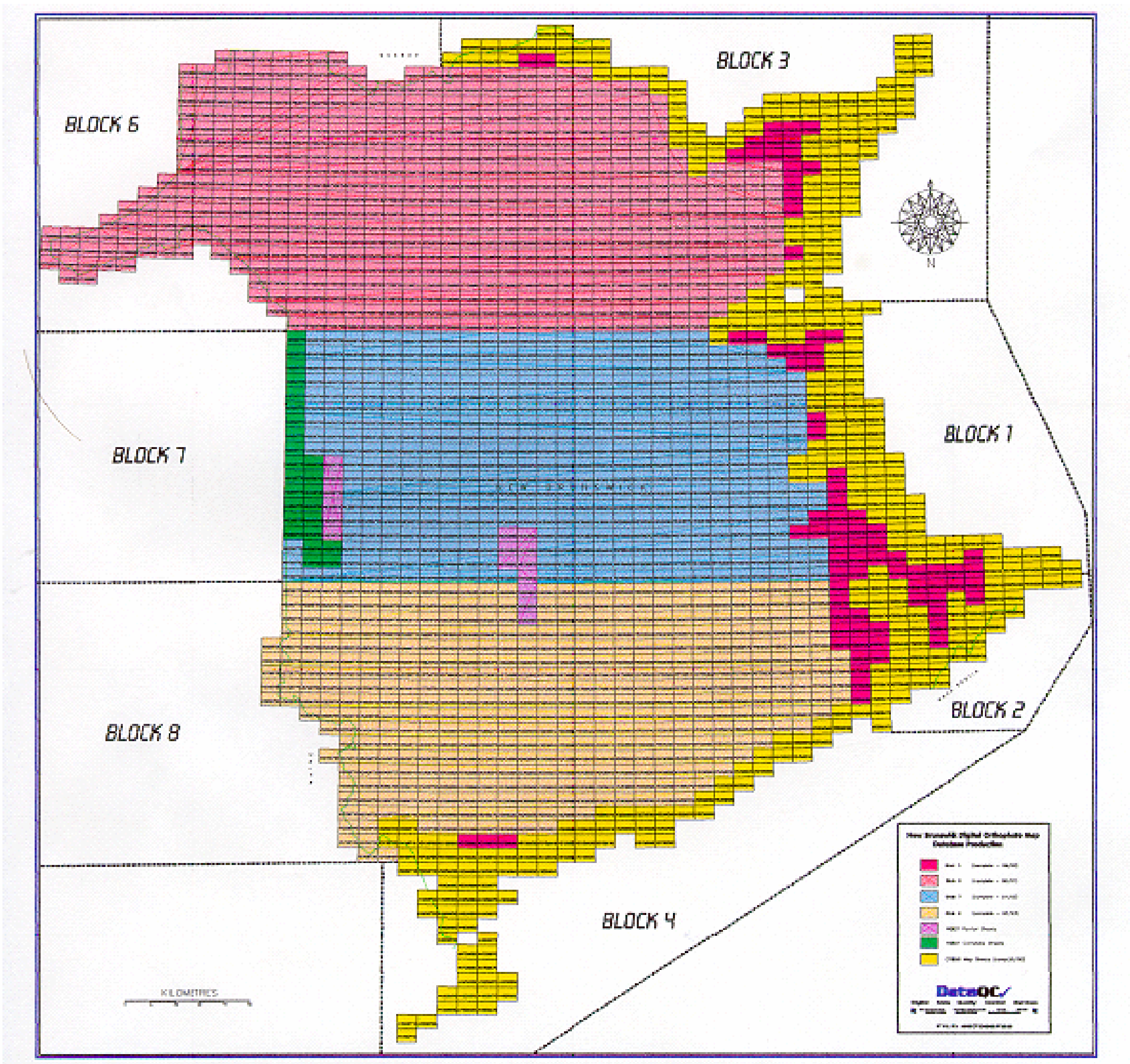


Figure 3-4 Sample of hardcopy status map from DataQC, Inc

All identified processes and resources objects not only provide basic building blocks for modeling the production process details but also will be used as the basis to model their structural behaviors and interactions among them.

Process Description and Structural Behaviors

The details of any business process modeled usually depend on the purpose of the modeling and the complexity of the underlying business process. In this thesis, the existing production project processes were modeled at three levels to ensure a clear presentation of relationships between process and functional roles, as well as to group contract-specific procedures into sub-processes so that they could be replaced easily. By "contract-specific procedures", it means those processes that are specifically defined for individual production contracts such as data processing and quality control procedures. This will be further discussed shortly.

The first level presents an overall (i.e., project-wide) process consisting of five processes identified in the previous section. The second level describes details of each of these five processes, which contains activities and, sometimes, sub-processes consisting of contract-specific procedures performed by respective functional roles. The third level models the details of sub-processes specified at the second level that consists of only contract-specific procedures. UML activity diagrams were used to capture and specify existing processes at these three levels.

Activity diagrams describe business processes by means of activities that can occur sequentially or in parallel and for which branching and synchronization can be defined. With "swim lanes" acting as responsibility lanes to assign both functional roles and internal business roles depending on the modeling level, processes could be well described with respect to the underlying organizational structures. Since each activity can be linked to another activity diagram that describes a sub-process, it makes navigation between different levels very easy.

1. Project-wide Process (Top Level)

At the highest level, the five processes identified above comprise the overall production process illustrated in Figure 3-5. The process starts with source materials preparation by the client. Once these materials are delivered to the project manager, they are checked and dispatched to various production contractors based on contracts. Production contractors then start to process the source data and submit processed data to the inspector for Quality Control. At this stage, if subcontractors are involved, production contractors reassign part of the source data to those subcontractors for processing. Subcontractors submit their processed data to production contractors. The QC inspector inspects the data submitted and, depending on the QC results, the inspected data is either returned to production contractors for further corrections or submitted to the project manager for final delivery. In either case, proper QC reports are delivered. Finally, after ensuring data is complete and contract specifications are met, the project manager delivers final data products to the client and the process is finished.

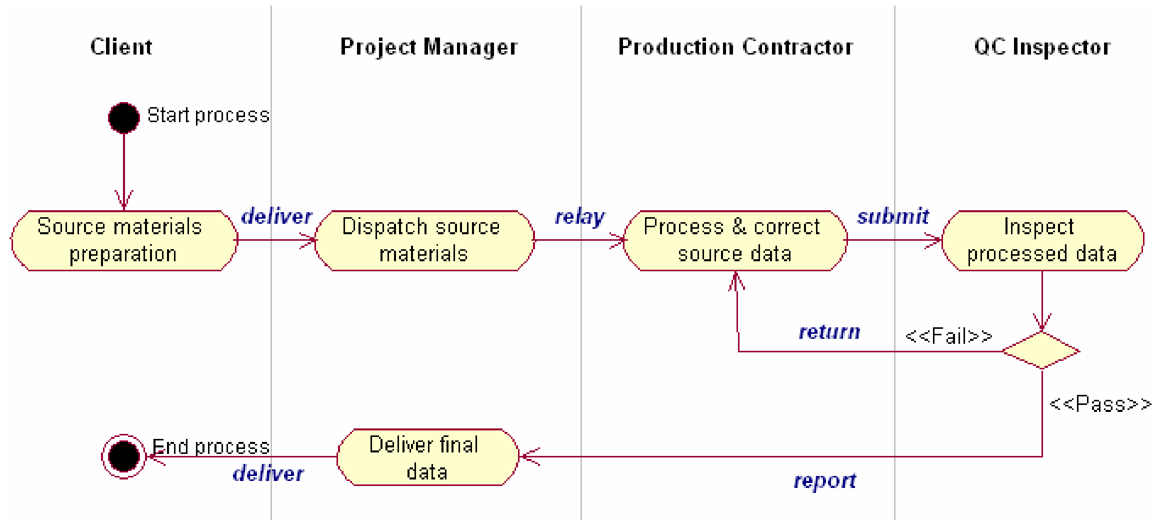


Figure 3-5 Top-level production process

The project-wide process is dominated by the flow of data materials - data files, meaning that the arrivals and deliveries of data materials mostly mark the beginning and ending of its sub-processes. For example, the delivery of processed data ends the *Data Production & Correction* process at the production contractor and the arrival of processed data started the *Data Inspection* process at the QC inspector. Because these data materials are usually shipped for delivery, submission, returning, or re-submission in shipments, the flow of each shipment depends on the completion of all data files within that shipment, resulting in a "*sequential*" process.

The data file is the basic work unit in the process. In practice, various types of information are associated with data files over the project span, including file description attributes, project related attributes and information for the financial purpose. A nonexhaustive list of identified attributes and their short descriptions are included in Table 3-3.

Table 3-3 Attributes associated with data files

Category	Name of Attribute	Description of Attribute
Data File Attributes	File Identification	A unique number or string to identify a data file
	File Name	A name of a file based on project naming rules
	Data Format	The format of the data stored in the data file
	Block Number	A number or string to identify a block
	Version	An identifier for a specific version of a data file
	Status of Edge-match	A code indicating if edge-matching is done
Project Attributes	Name of Processor	A name of the person working on the data file
	Name of Contractor	A name of the contractor having the data file
	Shipment Number	An identifier for a specific shipment
	Status of Processing	A code indicating the stage of processing
	Submission Date	The date the file is submitted
	Returning Date	The date the file is returned for further work
Financial Attributes	Expected Delivery date	The date the file is expected to be delivered
	Actual Delivery Date	The date the file is actually delivered
	Expected Accept Date	The date the file is expected to be accepted
	Final Accepted Date	The date the file is finally accepted
	Calculated Penalty ($)	Penalty calculated based on actual delay
	Penalty Occurred ($)	Penalty charged

Problems arise from time to time during the production process. Project participants have to communicate with each other to resolve such problems as different interpretations of contract specifications [Finley, 1997] and QC failures. Telephone calls, Fax, paper-based correspondences and emails are among commonly used communications methods; alternatively, face-to-face meetings are arranged. Communications are usually multi-directional, although the following patterns exist in some projects:

(1) Contract-related issues are discussed between production contractors and the client.

(2) Issues regarding project schedules, deadlines and deliverables are discussed between the project manager and other project participants.

(3) Issues regarding comprehension of QC results, especially data failures, are discussed mostly between the QC inspector and production contractors.

2. Participating Processes (Second Level)

Participating processes are those processes that are internal to the participating organizations or companies in any data production projects. In isolation, these processes are determined by (a) organizational structures, (b) business objectives or long-term goals, (c) adoption and implementation of ISO 9000, and (d) data production techniques used. When participating in data production contracts, these processes may also be affected by the contract specifications; for instance, procedures bounded with specific data formats. While process steps in organizations or companies playing different functional roles are obviously different from each other, even for companies playing the same functional role (such as production contractor), their process steps vary.

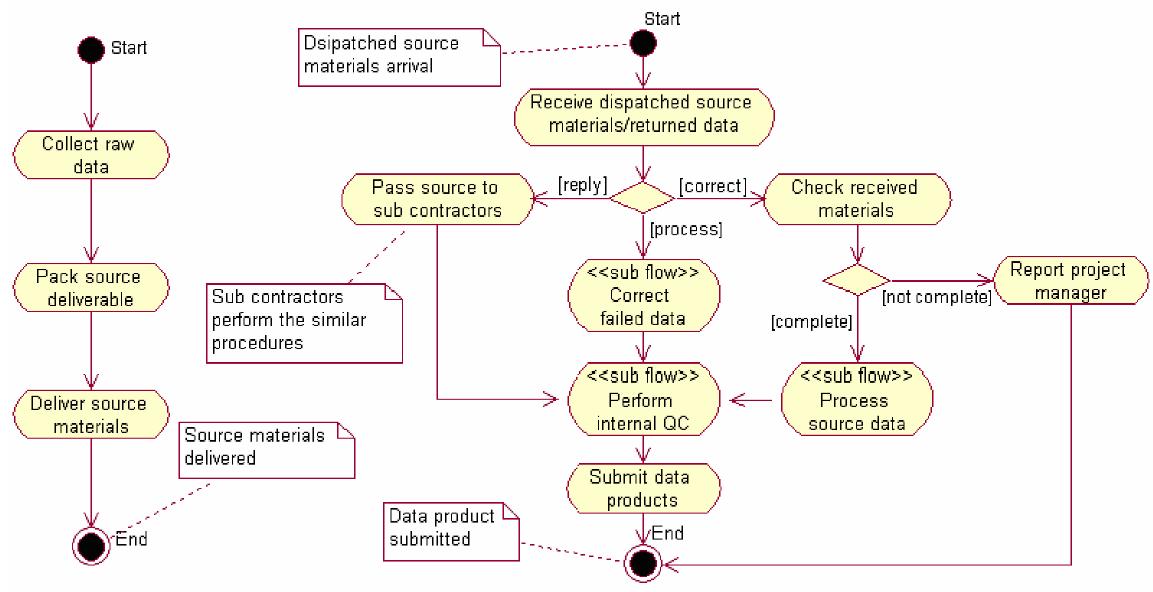
Given these complexities, it was hard to model a unified process for each of the five identified processes specified at the project-wide process to capture all variations. Instead, a "*place holder*", called "sub flow" as discussed shortly, was used to encapsulate part of the process that was affected most by the factors discussed above. Characterized from the existing processes, the rest of the process was abstracted into activities that had indirect connections with the overall project process. Results are described as follows:

(1) *Source Materials Preparation*: This process is to prepare source materials needed for data production operations through the following activities: collecting necessary raw data, packing source materials and delivering source materials to the project manager, as specified in Figure 3-6 (a). The process is performed and managed by the client.

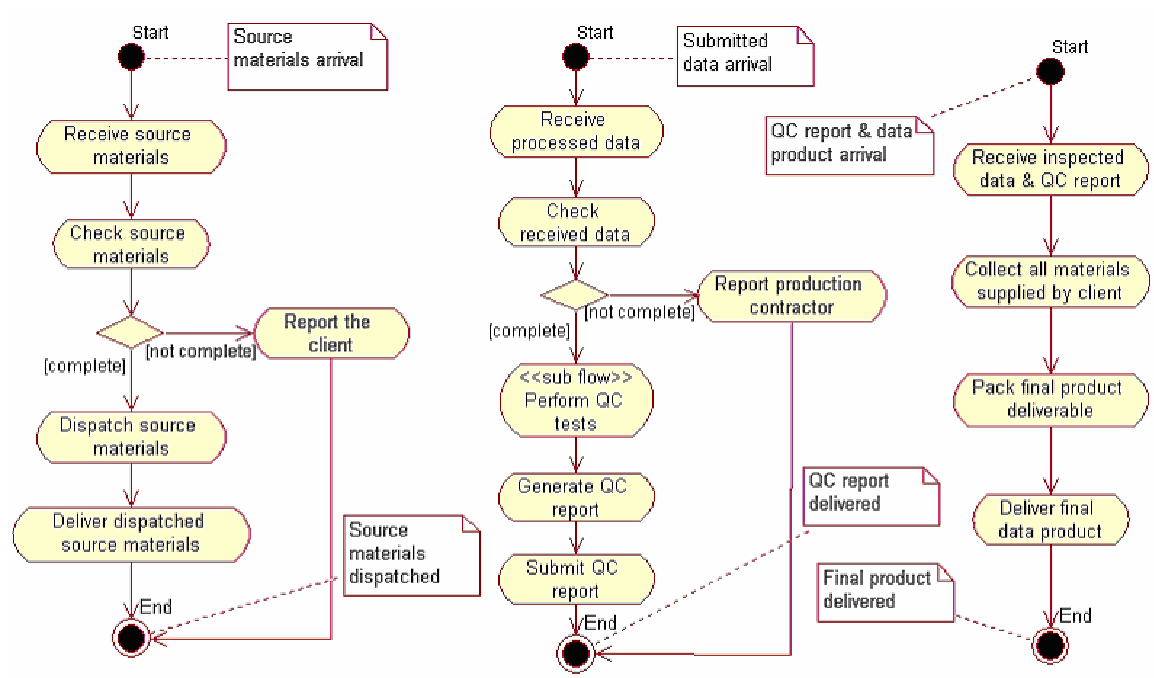
(2) *Source Materials Dispatching*: Upon receiving source data materials from the client, this process allows the project manager to check the completeness of delivered source materials and controls when to dispatch what materials to production contractors. The process goes through such activities as receiving source materials, checking materials, dispatching source materials and delivering source materials, as shown in Figure 3-6 (b).

(3) *Data Production & Correction*: This process entails production of data products or rework on failed data based on contract specifications. The abstracted activities involved include receiving source materials (from the project manager), processing source data, performing internal QC, and submitting processed data, as well as correcting returned data if it failed the *Data Inspection* process, as specified in Figure 3-6 (c). Three activities stereotyped as "s*ub flow*" are considered as contract-specific processes and they were modeled at the third level.

(4) *Data Inspection*: The objective of this process is to determine and report to the project manager if the processed data complies with quality control specifications through performing a series of QC tests. The involved activities include receiving processed data, performing QC tests, generating QC reports and submitting QC reports, as specified in Figure 3-6 (d). Again, the stereotyped activities are considered as contract-specific processes and modeled at the third level.



(a) (c)



(b) (d) (e)

Figure 3-6 Abstraction of participating processes at the second level

(5) *Data Product Delivery*: This process entails ensuring all data products satisfy contract specifications and are subsequently delivered to the client. The activities involved include receiving inspected data and QC report, collecting all materials supplied by the client, packing final deliverables and delivering, as shown in Figure 3-6 (e).

3. Contract-specific Processes (Third Level)

At the second level as shown in Figure 3-6, four separate activities stereotyped as "sub flow" were identified as contract-specific processes, including processing source materials or correcting failed data and internal QC in (c) and performing QC tests in (d). The actual data processing and correction processes involve heavy human intervention in performing contract-specified procedures. In contrast, the QC inspection process usually consists of many computerized procedures, meaning that running a set of QC testing programs or batch processing accomplished the large part of the quality control process. Human interaction is required mainly to analyze and record results from each test, as well as to manually invoke subsequent testing programs. Some example process diagrams collected in this research (included in Appendix A) demonstrate these observations.

Most quality control (inspection) processes complied with, to some extent, such criteria as completeness, logical consistency, physical consistency, and referential integrity. However, procedures and QC testing tools used to test these criteria varied from one company (specialized in data quality control) to another and from one project to another. The most determinative factors are the nature of the data to be inspected (e.g., data model and/or format) and the business logic of the QC company that is usually confidential. Discussions with industry inspectors indicate that part of this process has the potential for automation [Doucette, 2001].

Discussion

For the GIS data production projects examined, the business objectives were usually defined on an *ad hoc* basis, within the scope of a specific project or program. While some participating companies may have already implemented ISO 9000 standard [Hunter, 1996], its adoption in such a "*collective*" data production life cycle (integration of heterogeneous production processes) was considered more complicated. While workflow technologies can be valuable for achieving ISO 9000 compliance by providing mechanisms to enforce ISO 9000 certified procedures and control and update ISO 9000 documentation through the integration of document management systems, the discussion on this issue requires more dedicated research efforts that are beyond the scope of this thesis research. The existing production process was only examined and modeled within the organizational structures with considerations of various production techniques defined in procedures.

While coordinating the project processes in a controlled way was considered essential for improving process throughput, three things were found to potentially slow down the overall production process or affect the performance of involved processes. These included:

(1) *Lack of efficient mechanisms to transfer process resources*: especially for data files involved (as discussed in Section 2.1.3). Since the overall performance of project progress was largely determined based on deadlines for project deliverables, inefficient transfer of process resources very often delayed the project completion or forced the project manager to consider a longer project span at the planning stage.

(2) *Absence of concurrent processes*: Data materials were usually shipped based on a predefined contract shipment, with each shipment containing many data files (e.g., around 50 to 180 files per shipment in SNB projects). The process for any specific shipment depended on the processing of all data files within this shipment. For example, if one or more data files failed the internal QC inspection at the production stage, the whole shipment might be held rather than passing it on to the QC inspector for the *Data Inspection* process.

(3) *Change of data content during processing*: Sometimes the data provider reported changes of data content or had new data content required to be included after data was delivered to the production contractor and the data processing had started.

As described previously, there were many other types of project-related documents involved in data production projects in addition to GIS data files. Through discussions with intended users (as specified in Section 3.2.2), there was a strong consensus on having a central repository to store, control and manage all these documents, as well as to provide suitable tools for project participants to be able to find and access them easily and efficiently. The importance of having tools to automatically acquire information from the process for creating labor-intensive (e.g., one weekly QC status report may take half a week for a secretary to assemble) documents such as project reports was also perceived.

Given the fact that collaborative GIS data production projects (especially at the large-scale level) involving geographically dispersed participants in different time zones, there was also consensus on having proper tools to improve communications, to record communication results and to reduce cost for face-to-face meetings.

In summary, with all these insights obtained by examining actual projects and production processes, the collaborative workspace should provide functional support to facilitate at least the following four types of operation and management tasks:

(1) *Data File Movement and Storage* - Managing electronic submission, re-submission and delivery of data materials (files) in a controlled way and providing a central repository to store these data materials, as well as project-related information.

(2) *Coordination of Production Processes* - Automating and scheduling process activities, where appropriate, and tracking data status and process progress.

(3) *Management of Delivery Schedules* - Ensuring delivery schedules at various stages of the overall production process by means of reporting and monitoring progress, as well as notifying incoming deadlines and overdue.

(4) *Communications of Issues and Problems* - Discussing both technical and managerial issues and problems that may slow down the production process or cause confusion among project participants.

Built on the insights provided in this section, the next section will focus on defining system requirements specification of the collaborative workspace with an approach - Use Case Analysis with unified modeling language (UML), which formed the basis for the development of the architectural view of the collaboration model.

3.3 Collaborative Workspace Requirements Specification

The requirements specification was built based on an analysis of results from the previous section by using a use case analysis approach and then expressing the complete specification as a UML use case model. The goal of having this requirements specification was to provide inputs for higher-level architectural design of the collaborative workspace, to provide blueprints and testing cases for prototype development and testing as discussed in Chapter 4.

The requirements specification described in this section captures the most significant functional and nonfunctional requirement that enables the collaborative workspace to:

(1) provide data organization and workflow support for submission, returns and re-submissions of digital map files involved in data production processes;

(2) manage and automate the workflow processes dynamically in a controlled way for shortening project flow and increasing process throughput;

(3) facilitate feedback among production contractors, inspector and client on specific problems by using electronic communication and collaboration capabilities;

(4) allow tracking of project progress including tracking the status of individual digital map files and other project-assigned tasks; and

(5) be able to build a “project history archive” and manage relevant documentation repository.

The functional requirements specified in use cases were prioritized into three levels of importance: (a) absolutely needed, (b) highly desirable but not necessary, and (c) possible but could be eliminated [Pfleeger, 1998]. While the first-level requirements relate to actions or events of basic use case scenario, the second level ones deal with an alternative flow of events. Since the lowest-level requirements relate to those functions that make the use of the collaborative workspace more "comfortable", little effort was put on capturing requirements at this level.

Nonfunctional requirements specify system properties such as application standards, system quality (e.g., usability and performance), and system environments (e.g., platform dependencies and extensibility). Nonfunctional requirements related to any use case are described within the right context of that use case. However, for those nonfunctional requirements that are more generic and cannot be connected to individual use cases, they are documented separately as *Supplementary Requirements.*

3.3.1 Capturing Requirements with Use Cases

A "use case" describes interactions between users and the application system in design by identifying a set of scenarios tied together by a common user goal [Cockburn, 2000]. Although there are many variations of what entities should be included in use-case descriptions, actor and scenarios are the most-commonly used entities in use-case descriptions. Therefore, the research adopted four major entities in describing each use case - including use case goal, actor, scenarios and constraints. In some use case descriptions, an "Extension Points" element is also included to indicate necessary connections to other use cases. The "scenarios" element describes the flow of events and alternatives for each function to realize a user goal. The "constraints" element captures non-functional requirements tied or related to this use case.

Table 3-4 shows a use case description of the identified function - "Query Data File Status". While the primary scenario describes the basic flow of text-based query function, two alternative scenarios extend this basic function to allow users to either print a status report in a specific format (e.g., Microsoft Excel format) or query status based on an index map showing status information. Since this research did not intend to design an actual collaborative workspace system, the descriptions of all identified use cases are not given in detail. For example, if the use case - "Query Data File Status" were to be described for the purpose of developing an actual web-based system, the primary scenario would have been described as follows (description segment):

Table 3-4 Use case description for Query Data File Status

Name	***Query Data File Status***
Goal	To get current status information of data files in process
Actor	Client and project manager
Scenarios	***Primary Scenario*** 1. System presents the query form with default values to form elements leading to query status of all data files. 2. Actor enters query criteria (e.g., data files contracted by contractor XYZ or list of data files that have passed QC inspection). 3. Actor goes to query. 4. System performs query and presents results. ***Alternative***: Print Status Report At step 3, actor may want to print a status report. Actor selects report format options and goes to print. Then use case finishes. ***Alternative***: Query Status Map At step 2, an actor may query graphically. The system displays a status map showing the query results. The actor can pan, zoom, identify, hyperlink, and plot the displayed status map.
Constraints	1. Status information should be updated regularly, normally on daily or weekly basis. However, the time interval depends on individual project implementation.

(1) The actor selects the Query Data File Status link from the home page of the collaborative workspace web site. The system returns with a query form and instructions on how the query process works.

(2) The actor reviews the instructions and enters in query criteria and submits the query to the system.

(3) The system finds status information for all data files that match the criteria specified by the actor and presents the listing results to the actor. Each listing line includes the name of the data file, status, updating date, and current data file holder.

(4) The actor can at any time return to the query page, or the homepage, and revise the query criteria to perform a new query.

Use cases identified in this research were organized into five packages including Data File Management, Project/Document Management, Project Communications, Workflow Management, and Security Administration. All identified major use cases with packages (except for Security Administration which is common to most applications) are shown in Figure 3-7 and a complete set of descriptions of these use cases is included in Appendix B. Functional requirements captured using the use case analysis approach are briefly described as follows without distinguishing between web-based client and window-based client (called PC Client thereafter) in an Internet client/server environment. The use case model captures requirements the proposed system should satisfy from the user's point of view.

1. Data File Management

   (a) *Data file checkout*: The data file can be downloaded from the central data file repository to a user's local machine by browsing and selecting desired data files from a data file directory associated with the repository. Optionally, the user may prefer a graphical selection interface that allows user to pinpoint a specific data file or draw an area on a map graphics to select all data files covering the area. Sometimes, the user may also want to preview the content of the data file to be downloaded especially when the file size is large. Once downloaded, the system updates proper status information of the data file and locks the data file to exclude updating by other users. In case the checkout is for the client to receive the final data product, the system terminates the overall process instance for that data file.

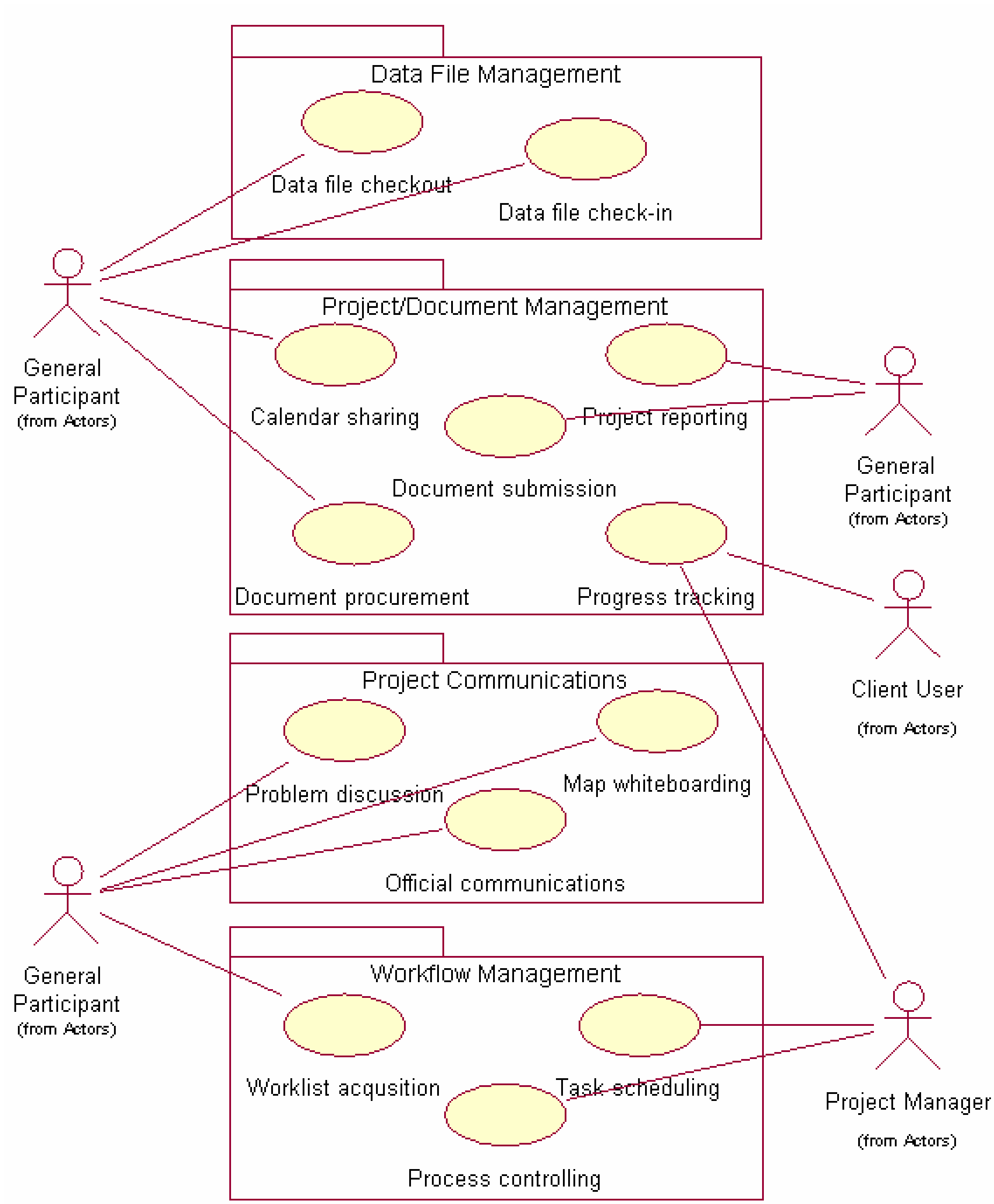


Figure 3-7 UML use case diagram for requirements specification

(b) *Data file check-in*: Data files can be uploaded to the central repository from the user's local machine by selecting a given data file from the local directory. Once uploaded, the system updates proper status information of the file and may release the lock on the file depending on check-in type. If the uploaded data file is for data re-submission purposes, the system also generates the proper version number for this copy of the file. In case the check-in is for the client to deliver source data, the system starts the process instance for that data file.

2. Project/Document Management

(a) *Document submission*: The user can publish project related documents (as described in Section 3.2.2) to the central repository by uploading document files with certain attributes such as document type, version and list of people or parties required to view it. Upon uploading, the system categorizes it and automatically sends a notification to the attention of related people and parties.

(b) *Document procurement*: The user can browse and search documents cataloged in the central repository and download them to the user's local machine. The system logs document access information regarding who accesses what documents and when it happens.

(c) *Progress tracking*: The project manager and the client can query data status in processing and the overall progress of the project such as how many data files has passed QC inspection and the completion percentage of the project. Based on the query results, they should have option to request certain forms of summary reports. Optionally, they can specify conditions for getting a digital status map.

(d) *Project reporting*: The user can send any project reports by selecting report templates and entering report entry information, such as current processing status of data files (reports templates normally various from one project to another). The system allows the user to specify to whom the report should be attended and sends a notification to attention parties automatically.

(e) *Calendar sharing*: The project calendar shows project milestones, deadlines, date of deliveries, date of major updates, daily scheduled activities, etc. The calendar can provide views of all this information on daily, weekly, monthly, or project span basis. All users can view the project calendar of which the entry information is stored in a centralized project metadata database. Project manager can update entries of the project calendar and others can send request for updating. Once the updates are confirmed, the system automatically sends a notification to all parties about new changes. Optionally, users can have a view of traditional project plan in a Gantt-Chart style.

3. Project Communications

(a) *Collaborative map whiteboarding*: Two or more different users in different locations can simultaneously access the same data file to (1) view and manipulate the data file; and (2) electronically mark up entities of the data file that requires further attention during "shared" discussion sessions (i.e., analogous to attaching "post-it" notes or comments to a hardcopy map sheet). One user can initiate a session by calling other users involved and selecting the map sheet to work on. Each party can see right away what changes or "mark-ups" others make.

Alternatively, the user can work offline during individual sessions and send copies of the "marked-up" file to specified parties for further work. In both cases, the user can select features and change their geometry and attributes with different parties properly identifying themselves by, for example, using different colors. The system creates required map sheets in raster image format (e.g., GIF or JPEG) and keeps a history of changes. Upon the termination of the session, the final copy of the worked map sheet is stored in the central repository. Ideally, the system can automatically register changes made on images to the map file in its original format. Optionally, for the synchronous collaborative session, the user can invoke a video or audio conference.

(b) *Official communications*: The user can send official correspondence to other parties. The system can have a copy of the correspondence and store it in the central repository. Alternatively, the user can browse and search correspondence stored in the repository, read the selected correspondence, and reply to that one.

(c) *Problems/issues discussing*: General issues, problems, and lessons learned can be posted as messages to the project discussion forum. The user can directly post a message, browse and search specific messages, and reply to a selected message. The system keeps threads of related messages.

4. Workflow Management

(a) *Worklist acquisition*: Users can get their worklist with all work items assigned to them either automatically sent to them or through request. They can browse through the worklist, check out individual work items, re-assign work items, and signal the system of the completion of the work items checked out. The system should allow users specifies filtering criteria to determine what workflow information related to work items are to be attached to the work items.

(b) *Data production process controlling*: The project manager can view current status of any interested data process instances and perform some controlling operations such as suspending, stopping, and aborting process instances. The project manager can also identify bottlenecks of the selected workflow process and go to reschedule work items of any process activities that may cause the bottlenecks.

(c) *Task scheduling*: This function allows the project manager to do general task and resources scheduling, usually tied with data files in processing, associated with the overall data production workflow by specifying tasks, duration, resources, task dependencies, and so on. Then the system should be able to track, alert, notify, and route the scheduled tasks based on the saved scheduling attributes.

(5) Security Administration

(a) *User Authentication*: Allows the user to log on to the collaborative workspace using a pair of username and password and the system to validate the user.

(b) *User Account Creation*: Allows the new user to create an account with proper profile information.

(c) *User Profile Updating*: Allows the user to update his/her profile information.

(d) *User Access Level Control*: Enables the control of access to project information and tasks based on the user's authority.

3.3.2 Supplementary Requirements

Many nonfunctional requirements are associated with some "hard" numbers; for example, "the system should be available 24 hours a day and 7 days a week". Since these numbers are highly related to individual systems in design, those nonfunctional requirements that do involve numbers were described either using example numbers or brief explanations in use case descriptions.

Apart from those described in the use case descriptions, there are also some nonfunctional requirements generic to either many use cases or the whole collaborative workspace. Rather than capturing all nonfunctional requirements at the system-level, from which many are similar to other Internet-based systems, this research only captures those that characterize the proposed collaborative workspace. They are briefly discussed as follows.

Usability Requirements: The biggest concern here is the amount of training time required by project participants to be able to use the workspace productively in their routines. As discussed in Chapter 2, collaborative GIS data production projects are normally based on production contracts, implying that users may only employ the workspace from time to time. It is not feasible to assume a week or two for them to learn how to use the workspace. Therefore, the system must be straightforward, easy to learn, and require no more than a few hours for users to become familiar with it. An example is to use Web-based interfaces that are already familiar to most people.

Reliability Requirements: The proposed collaborative workspace may have to be available on a basis of twenty-four hours a day and seven days a week since (1) project participants may schedule later data file transfer; (2) global-wide projects may involve different time zones; and especially (3) there may be more than one project underway and managed by the same workspace. Compared with other commercial systems, the workspace has significantly reduced access load. It is always possible to schedule downtime during weekdays provided users are given prior notice.

One of the potentials of using this Internet-based workspace is to reduce wastage of project resources, for example not to store data files at multiple locations. The central storage must be secure and, at least, it should be replicated or backed up on a daily basis.

Performance Requirements: The collaborative workspace should be able to handle and store large size data files, ranging from several MB up to 250 MB. While the number of users accessing the system will be small compared with other types of Internet-based systems, it should at least allow concurrent access to the same resources stored on the system.

Standards Requirements: Standards should be referenced including OpenGIS standards (e.g., Architectural Specifications and Metadata Cataloging Specifications), FGDC standards (e.g., data quality), and standards from the Workflow Management Coalition (e.g., Workflow Reference Model). While OGC metadata cataloging has a focus on data discovery, the metadata related to quality control in FGDC standards describes five quality aspects of GIS data.

Both functional and nonfunctional requirements are discussed in this section in terms of "*what*" users can do with the proposed workspace. In order to be used as direct inputs for later higher-level architectural design, these use cases need to be further analyzed to find out "*how*" they can be realized in the collaborative workspace in terms of the system's architecture. This aspect will be discussed in the next chapter.

3.4 Comparative Study of Current Groupware Functions

Rather than reviewing a large number of groupware packages, which was out of the scope of this research, several major groupware systems were selected to identify typical collaboration functions available in the market, including Lotus Domino ™, Microsoft Exchange ™, GroupWise ™, and Netscape SuiteSpot ™. While these systems had been reviewed up to 1999 based on product manuals, white papers, and published evaluation reports, only Netscape SuiteSpot 3.5 was installed[5] on a workstation to allow:

(1) Close examination of available functions;

(2) Test of initial requirements identified for collaborative data production; and

(3) Easy communication with intended users during requirement analysis.

[5] In later 1998 and early 1999, Netscape SuiteSpot was the only system that allowed downloading of a fully functioning version through the Internet for evaluation purposes. Details regarding the installation, configuration and prototyping evaluation of this system may be found in Boettcher [1999].

By comparison with initial collaboration requirements, the identified available function list was used to determine (a) if the existing groupware capability satisfies the collaborative GIS data production requirements and (b) if there were any missed requirements that could be crucial for collaborative data production. While the details of groupware software review were reported in detail in [Li et al., 1998; Boettcher, 1999; Coleman and Li, 1999], Table 3-5 presents an overview of the review of these selected groupware packages.

Table 3-5 Comparative results of groupware functional review

Contents	Function List	SuiteSpot 3.5	Domino 4.6	Altavista 98	GroupWise 5.5
Document Management	File Transfer	V	–	–	–
	Publishing	V	V	V	V
	Searching	V	V	V	V
	Catalogue	V	V	V	V
	Version control	V	–	V	V
	Multiple formats	V	V	V	V
Communication	Open email	V	V	V	V
	Correspondence	–	–	–	–
	Discussion forum	V	–	V	–
Group Decision Making	E-Whiteboarding	V	–	–	–
	Conferencing	V	–	–	–
	Chat (IRC)	V	–	V	–
Group Calendar and Scheduling	Calendaring	V	V	V	V
	Scheduling	V	V	V	V
	Conflict Checking	V	–	–	V
	Email Notification	V	–	–	–
	Share Calendars	V	V	V	V
Workflow Management	Workflow Design	X	V	V	V
	Progress Tracking	X	V	V	V
	Documents Routing	X	V	V	V
Security Implementation	Authentication	V	V	V	V
	Encryption	V	V	V	V
	Access Control List	V	V	–	–

Legend: "V" Supported "X" Not Supported "–" Undefined

Compared with the initial requirements, it was found that most collaboration requirements were supported by functions located sporadically in various groupware packages. However, special consideration has to be taken in order to satisfy collaborative data production requirements. For example:

(1) *Needs for Integration*: The integration level of current groupware tools was still very low at the time of comparison in the sense that groupware components from any system are usually bundled tightly with its system core, making them hard to interact with the components from other packages. In addition, neither GIS software packages nor GIS tool kits provide support for interacting with those identified collaboration functions.

(2) *Limited Capability*: Given the special nature of GIS data, the current design and capability of groupware functions may not be able to fully support collaborative GIS data production needs. For instance, most existing functions for file transfer do not provide sufficient provisions for dealing with transferring very large data files such as image files. Especially, as discussed later in Chapter 6, transferring very large files over the Internet is still not reliable. These kinds of functions need to be extended to include capabilities of compression/decompression, transfer scheduling, and resuming of interrupted transfer, to satisfy GIS data production requirements.

Another example is “whiteboarding” or "collaborative editing/viewing" functions. When dealing with graphics, there is no support for any other graphics format than GIF, JPEG, BMP or TIFF. For some simple applications such as pinpointing problematic features, this may be sufficient. However, in more complicated cases such as group collaborative editing of the same map, the capability to access native GIS data formats, keeping the annotations or mark-ups on a separate layer, or even having these annotation/mark-up layer registered to the actual map may be desired.

The discussions in this chapter provide necessary insights to business surroundings and requirements of the collaborative workspace. It was based on these insights that the proposed collaboration model was designed. The next chapter will focus on discussing the design of the model.

Chapter 4 COLLABORATIVE WORKSPACE MODEL: METHODOLOGY AND DESIGN

While the collaborative workspace may be designed and developed based on a specific GIS data project environment resulting in a "best-fit" system for supporting that project, obtaining the more generic characteristics is more desirable since individual projects vary from one to another and their life cycle is usually short (controlled by data production contracts). The collaborative workspace (CW) model discussed in this chapter serves the purpose of modeling these characteristics and providing suitable guidelines for developing the collaborative workspace.

This chapter first describes the general methodology used in this research. Based on the established understandings of current practices (as discussed in Chapter 3), the development of the CW model is then presented in Section 4.2, and 4.4. Finally, Section 4.5 presents a brief summary on potential strengths and limitations of the designed model.

4.1 General Methodology Considerations

Given the methodology diversity of designing and developing CSCW-based system as discussed in Section 2.3.1.1, the overall approach adopted in this research was a participatory, prototype-oriented paradigm. By comparison with typical software life-cycle models (e.g., the waterfall model), this approach emphasizes two important aspects that are critical for CSCW system development: end user participation and define-design-build-evaluate iteration (see Figure 4-1 for the framework of this approach).

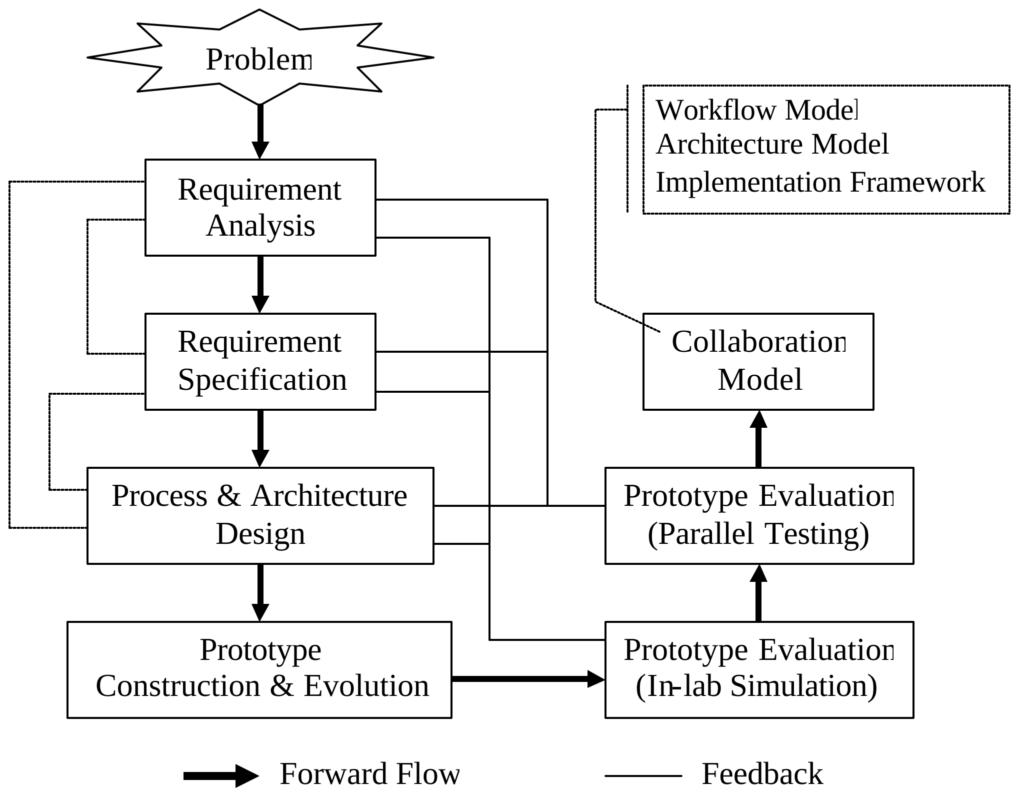


Figure 4-1 A participatory, prototype-oriented development framework

The involvement of "potential users" (defined in Section 3.1) in the development of CSCW systems is vital if the systems are to be accepted both functionally and by the user community [Hughes et al., 1994]. Most activities involving users' participation occurred in different forms in almost all phases illustrated in Figure 4-1, except the phases of design and prototype construction where results from user's participation served as feedback for further improvement.

Compared with a traditional sequential software development model (e.g., waterfall model), the iterative process allows feedback between phases to refine the artifacts from each phase based on the work of the phase immediately after that phase or all the phases followed, especially the two evaluation phases. In such a way, an incremental design could be realized with the first few rounds staying on core requirements. The feedback from user's participation at later phases could also be included in the design through iterations.

Throughout the whole development process, Unified Modeling Language (UML) was used to specify artifacts resulting from each phase (see Section 2.2.3.3 for UML details) where appropriate. This has been accomplished with the Rational Rose CASE tool that supports current UML standards. The methodology details and tools employed at each phase illustrated in Figure 4-1 will be further elaborated in later sections where appropriate when actual design and development work are discussed. Special considerations regarding the selected methodology will also be discussed.

While the first two phases – requirement analysis and requirement specifications – were discussed in the previous chapter, the output will be used in the subsequent phases to develop the proposed CW model, which is the ultimate goal of this thesis research. Before describing the development of the CW model, the definition of the CW model and its compositional components are presented here.

4.1.1 Definition and Components of the Collaborative Model

The motivation of developing a CW model was to provide both process and structural views of the collaborative workspace supporting distributed GIS data production operations and project management, as well as implementation guidelines. As illustrated in Figure 4-2, the collaborative GIS data production environments encompass several main structural components including the production processes involved, the people and organizational structures participating in these processes, as well as management processes, current management and contract policies, relevant technical standards, and domain-level systems and applications used for actual production activities such as data QC testing. To support cooperation among distributed project participants, supporting technologies have also to be integrated into a wide area network environment (the Internet in this study).

The actual implementation of a collaborative workspace supporting this type of GIS data production environment may be different depending on many factors such as contracts at hand and production procedures utilized. However, two issues in common need to be addressed: (1) how to design the group work processes that can improve the efficiency and productivity of the collaborative production environments, and (2) how to design the computer systems to support them. Once these two issues have been considered and results obtained, a development process with a set of guidelines is needed to guide the construction work of the supporting system – *the collaborative workspace*. The CW model was developed to address these issues, while both the model and the collaborative workspace built based on the model are used to verify the hypothesis of this research.

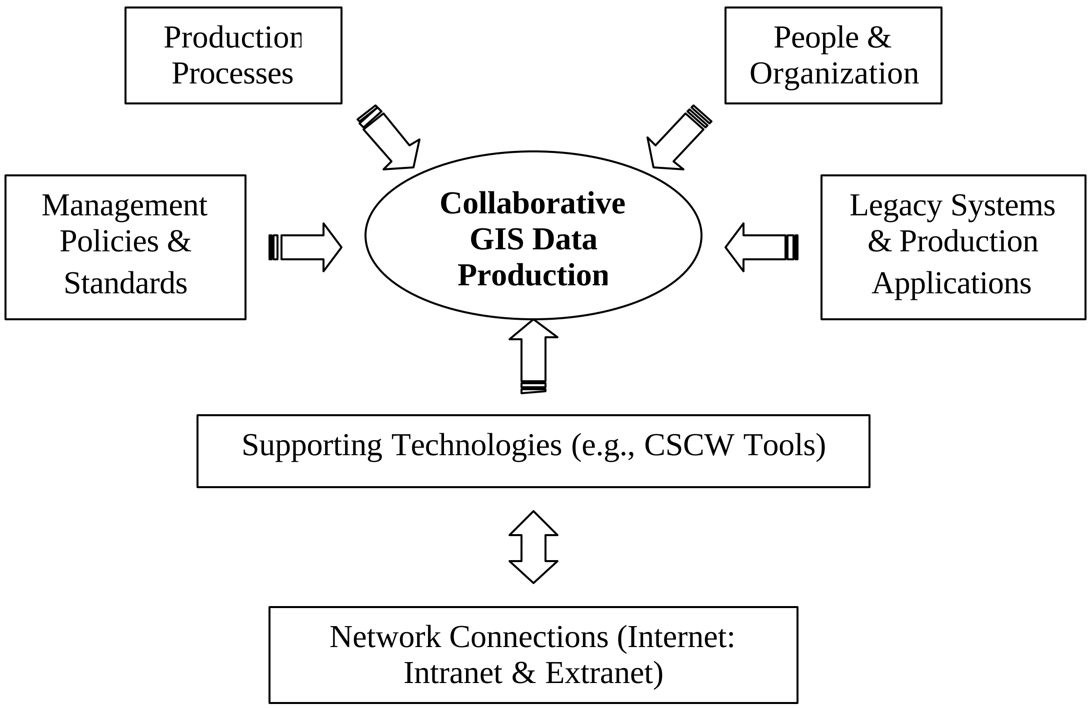


Figure 4-2 Structural view of the collaborative GIS data production environments

The CW model presents three views of the computerized collaborative GIS data production environments. While the workflow model is the main focus of this research, the architecture model and the implementation framework are necessary components for realizing the workflow model in actual workspace implementation. Three components of the CW model are defined as follows:

1) *Process Perspective – Workflow model*: a schema for automation and coordination of a specific set of geomatics production processes to complete a given project. The model elements include activities (procedures), roles, rules, relationships among them, policies, task scheduling components, and triggers for initiating external process (applications).
2) *Structural Perspective – Architecture model*: the structure of involved software and hardware components in the workspace including workflow components, groupware tools, scripting programs, GIS, DBMS, data warehouse and web servers such as HTTP server and Internet map server. Focus was on the interaction and information flow among them.
3) *Deployment Perspective – Implementation framework*: a set of specifications that can be used in a real production environment to help the project management design and implement the collaborative workspace.

While the issue regarding how well the model can be used to improve distributed GIS data production projects will be analyzed and discussed in Chapter 6 based on the designed model and the research prototype developed, the following sections focus on the design and development of the proposed collaborative workspace model.

4.2 Workflow Model

As described in Section 2.2.3, workflow deals with the automation of existing or redesigned business processes and manages their associated tasks, information and domain applications in a controlled way. The designed workflow model presents a structured view of these automated processes with their relevant workflow resources in the context of GIS data production operation and management. The purpose of developing a workflow model was to:

(1) establish workflow process templates that can be "understood" by workflow management systems to assign performers and tasks, to execute work assignments by computer, and to track the status of performed tasks;

(2) combine information and work materials (e.g., GIS data files and project documents) handled by production process into workflow instances. In case the information and work materials cannot be computerized (e.g., hardcopy plots), provide access linkages to the information through the workflow instances; and

(3) specify the potential external applications that can be invoked by the workflow management system or that can be "workflow-enabled".

4.2.1 Workflow Modeling Process

Workflow is not really an "application" *per se*, meaning that one cannot buy it "off the shelf" and install. It requires an understanding of the individual underlying processes and rules within the work environments. Workflow modeling largely depends on the results of a thorough examination of the existing distributed GIS data production processes and their environments as discussed in Section 3.2. These results, falling into four categories (see Figure 4-3), serve as the main input for modeling workflow. While relationships among these categories are important for characterizing data production environments (for example, processes are controlled by people and ruled by policies, standards and business rules), it is those elements within each category that essentially constitute the model.

The *existing and redesigned processes* include both management and operational procedures adopted in the current GIS data production practices (as described in Chapter 3). From these processes, insights can be obtained regarding how data production projects are run and what must be done to realize the business objectives. Domain-level applications and other IT applications that support these processes can also be identified. And most importantly, if the workflow principles have not been applied previously, the improvement that the workflow technology may bring in may be predicted.

While the *management policies, standards and business rules* help to identify business logic involved in production processes to determine the sequence of process steps, route project information and production work materials, and schedule tasks in the workflow, the current *organizational structures and business roles* provide the basis for defining or redefining functional roles and relationships within the context of the new integrated project-wide organizational structure.

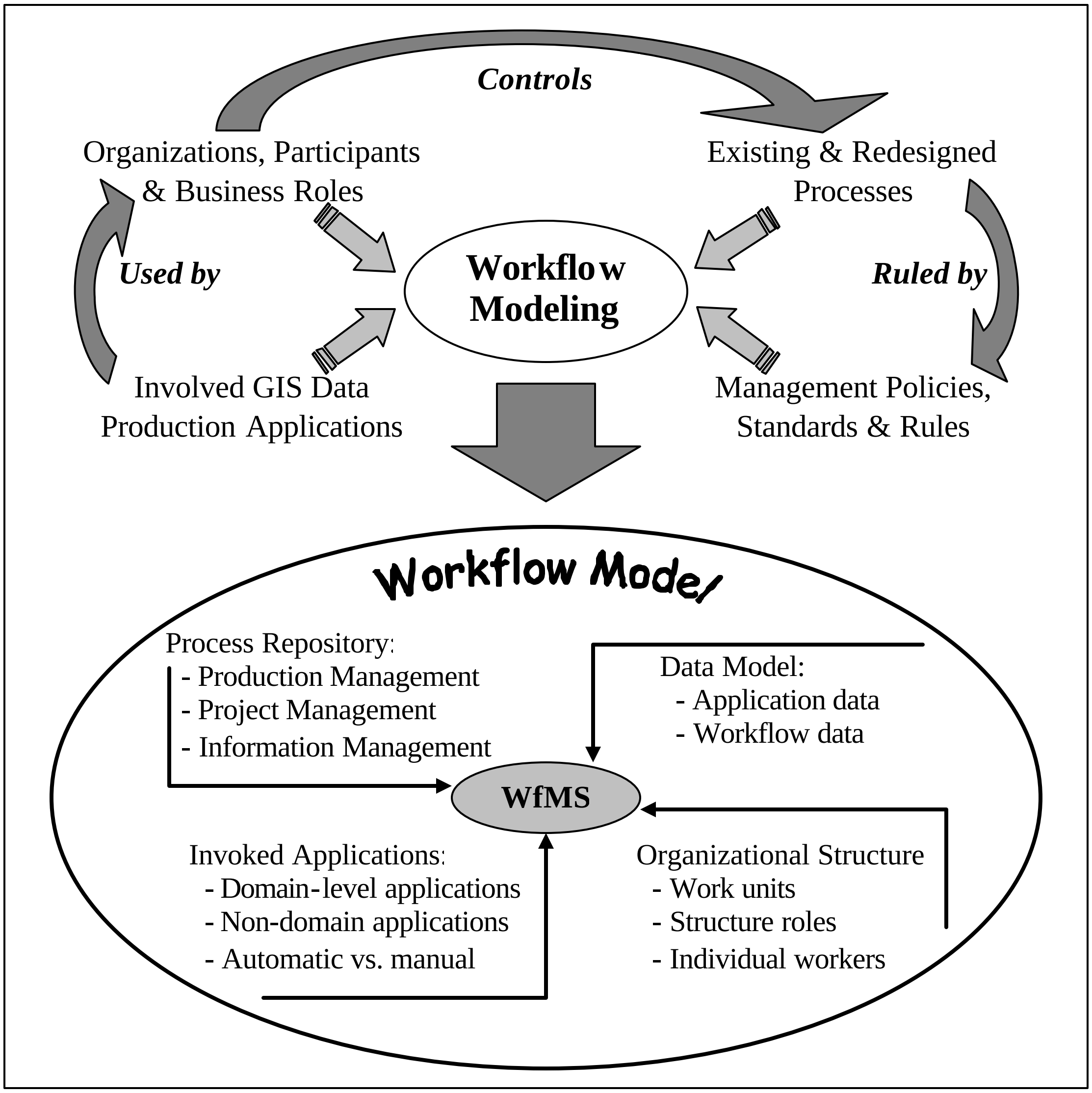


Figure 4-3 Workflow modeling process for collaborative GIS data production

Most workflow activities invoke existing applications in two ways: (1) the engine calls the application through direct procedural calling (local or remote); or (2) the engine queues the application to be invoked and it becomes the application's responsibility to check the workflow invoked application queue. One way or another, these applications must be associated with proper workflow activities in workflow process definitions. By identifying the *involved GIS data production applications* in existing production processes, the potentials for being invoked by the workflow engine can be better determined.

During the workflow modeling process, all the factors discussed above have to be thoroughly analyzed with the focus on the possible better coordination among them. The developed workflow model is structured to contain workflow process repository, invoked workflow application catalog, organizational structure, and data model (see Figure 4-3). While these structural components will be discussed in the following, an example WPDL description of the workflow model based on the discussion is included in Appendix B.

4.2.2 Workflow Process Repository

The process repository is organized in several categories (see Figure 4-4). Each category addresses a specific type of activity and contains a set of processes, while each process is described with items classified into one of three groups: business goal, process resources and involved activities. For example, the project management category may address all activities associated with processes for contract negotiation and project plan controlling. Example process categories involved in the collaborative GIS data production are described as follows:

1) *Operation Management*: including data delivery management process, data quality inspection process, and data producing process. Processes in this category are

directly related to GIS data handling procedures. Therefore, the performance of these processes to a large extent determines how fast and efficiently the final data products can be delivered to the customers (or clients).

2) *Project Management*: including contract negotiation process, project progress tracking process, and project reporting process. These processes help manage and control the execution of operational processes in the *operation management* category.

3) *Information Management*: including project document management process, contract specifications revision process, and project issues resolving process. These processes provide controlled approaches for sharing and managing project information and process resources, as well as communications of problems.

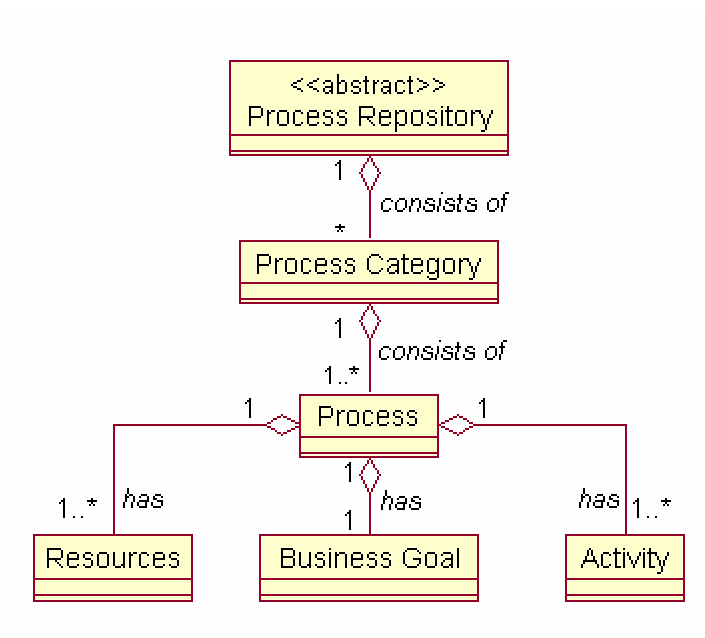


Figure 4-4 A hierarchy structure of the workflow process repository

As illustrated in Figure 4-4, each process consists of three major components: a collection of process resources, a business goal, and a set of activities. The process resources are input and output work products used or produced by the process, and additional materials (e.g., documents or other information objects) required to facilitate the process execution. In this object-oriented process structure, other process components such as pre- and post-process/activity are treated as process resources, while the information referencing things such as higher-level and low-level processes may be defined as attributes of a process class. The business goal comprises the main objective to be realized through the execution of the process, while the activities are defined as GIS data processing and management practices, which in the context of workflow model contains linkages to the applications to be invoked or linked.

Building a process repository allows inclusion of some industry best practices so that valuable lessons learned are beneficial for the repository users. The process repository also takes into account most business models commonly adopted by GIS data production projects, which makes it more adaptable to various situations. Since relevant standards such as ISO 9000 and domain specific standards (including *de facto* standards) are considered as important input for establishing the workflow process model (as illustrated in Figure 4-3), the process repository favors industrial standards for process improvement. Finally, the process repository provides a useful framework for managing, assessing and improving the processes, which is very important for providing GIS data production services in a timely fashion. The idea of having a process repository may be especially useful and beneficial for those companies whose major businesses involve dealing with contracts with processes such as GIS data quality control.

There are, however, no commercially available software packages that allow the creation and management of such a process repository. An alternative is to use existing workflow management systems or business process definition software with which the accessibility of the process repository would be limited to the process modeling language and constructs adopted by that specific software.

Workflow Process Determination

The selection and design of a suitable workflow process is widely considered as the first step for a successful workflow implementation since not every process can be a suitable workflow process for automation if it involves too many manual activities or requires constant human interaction. As discussed in Section 3.1, most first hand knowledge was obtained from existing GIS data quality control projects. Therefore, the following discussion mainly uses a data quality inspection process as the example.

GIS Data Quality Inspection processes usually involve running a set of QC test scripts or batch-processing procedures (mostly based on the DOS batch command in current practice). However, each company has its own QC procedures associated with various specific-testing scripts and deals with spatial data in various formats. It is not feasible to have a set of unified QC procedures that is accepted by all firms and a set of unified testing scripts that can deal with data in every format. However, it may be practical to have a set of higher-level activities that comply with data quality standards such as FGDC[6] standards. Based on this understanding, a data quality inspection process was designed as illustrated in Figure 4-5.

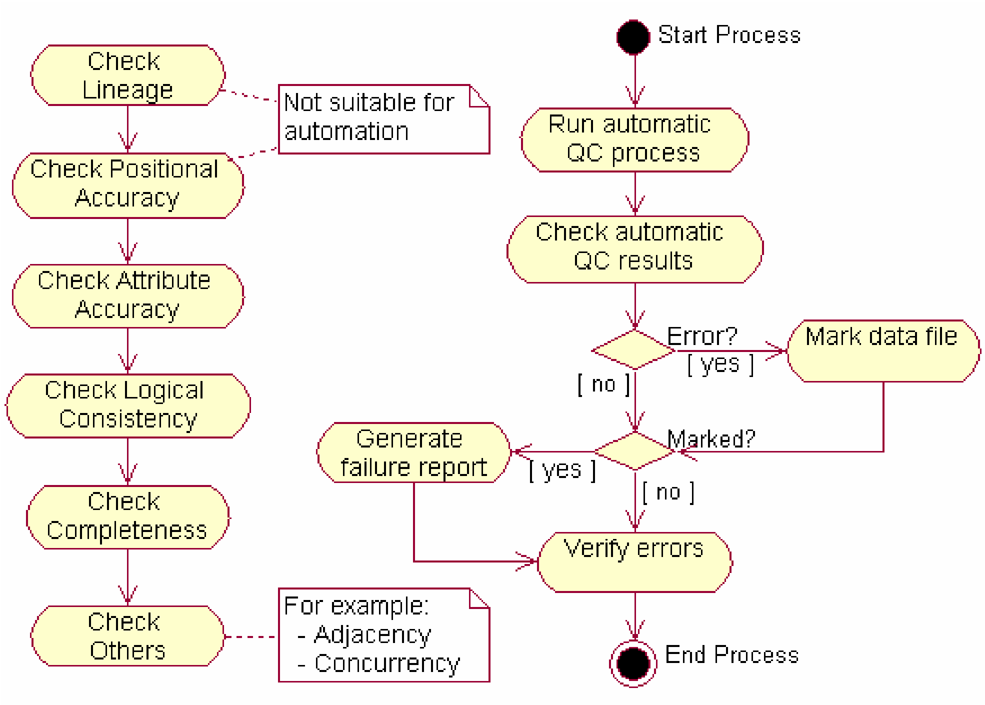


Figure 4-5 Example process for GIS data quality control

The overall process for data quality control includes six abstract activities based on the FGDC standard. The last activity encompasses all other data quality requirements specified by individual contracts (see the diagram on the left-hand side of Figure 4-5).

[6] FGDC is a non-profit organization promoting spatial data standards for geospatial information sharing (see its web site at http://www.fgdc.gov/ for more information).

Each activity has a number of QC testing actions that are essentially QC testing programs that may be invoked by workflow management systems. Each activity can be realized through a sub-process that has similar process structure as illustrated by the diagram on the right-hand side of the Figure 4-5, which has an automatically-executed part and a manually-handled part. As such, once a data file is submitted for inspection, the overall QC process is executed in such a way that automatic QC batch processing procedures at all levels are run first, followed by a manual check of results. Upon approval by a QC inspector, the workflow invokes the application for data markup to mark the errors found and generate the failure report, followed by final manual checks. The final QC report is submitted to the project manager for the reporting purpose.

GIS data quality control procedures require such predefined parameters as input, test statement, test criteria, and test results [WaterMark, 1995], as shown in the example in Figure 4-6 which validates theme numbers of all features in a GIS data file. Once QC procedures are controlled and executed by a workflow system, these parameters should be included in the workflow models as either workflow data or workflow application data.

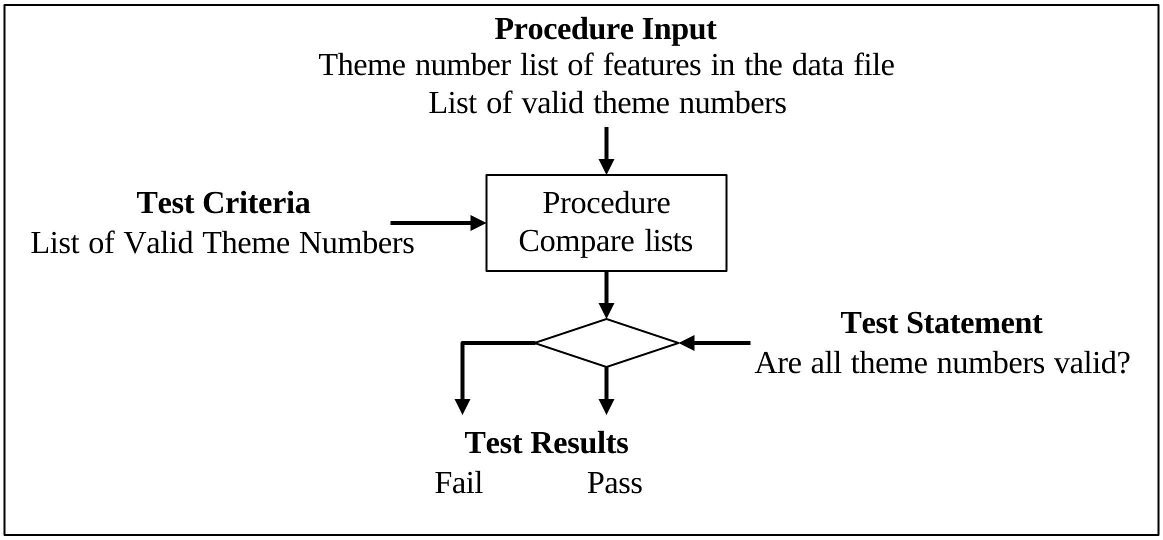


Figure 4-6 Example parameters needed for QC testing procedures

While workflow processes in each process category may be managed and executed separately, it would be much more efficient if they can be integrated under a high-level project-wide process and implemented through process integration mechanisms. The integration can be within a specific process category or across different categories. For example, for the contract-based production projects, the project-wide workflow process should avoid direct inclusions of any processes internal to participating companies. However, proper linkages should be designed to allow process integration if participating companies also run their own workflow management systems with internal workflow processes. Apart from the above basic considerations, the following criteria have also to be taken into account when designing the project-wide workflow process:

(1) The process should allow a continuous flow of project-wide activities over the project life span.

(2) The process should include those activities internal to project-participating companies that provide important information for project management, such as internal work status, or affect the flow of the overall process.

(3) The process must be spanned over distributed functional roles (organizations or companies) and time is the important factor of process execution.

(4) In workflow process design, each process activity should be atomic in the sense that there should be no branching or joining between actions or steps within the activity.

An example project-wide workflow process is illustrated in Figure 4-7 that integrates workflow processes (stereotyped as "sub flow") from the *Operation Management* category. Compared with the current project process where process resources (mainly information and work materials) were physically transferred back and forth among project participants, the new workflow process focused on the check-in and checkout of these resources with respect to the project central repository. The workflow process starts with the arrival of source materials from the client and ends when the client receives the final data products in the form specified by project contracts. Every time a unit of source materials (zipped source materials package containing a number of source files) arrives, a process instance is created and instances of all relevant workflow activities are created as well.

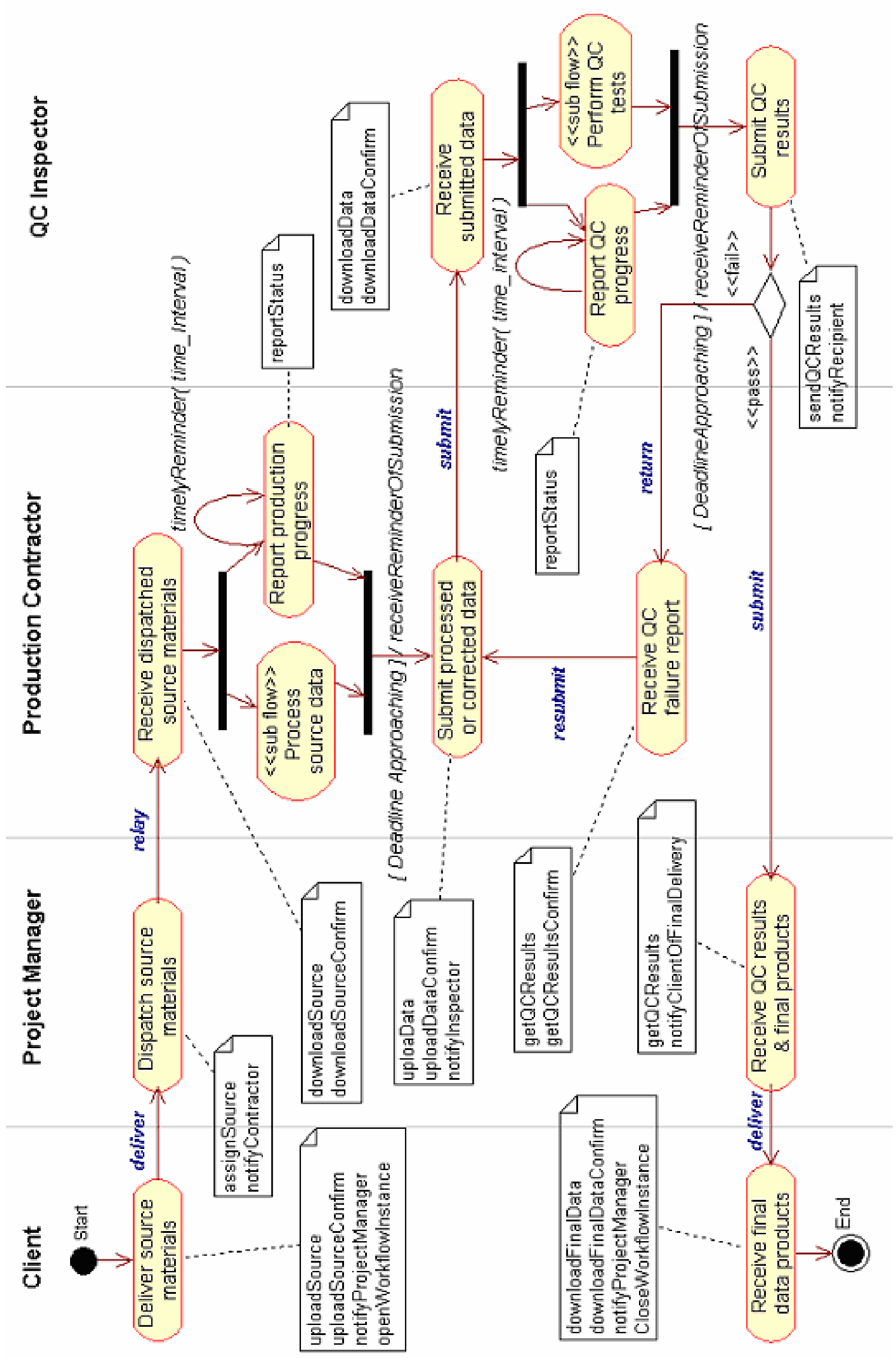


Figure 4-7 Example project-wide workflow process

The workflow process includes two sub-processes stereotyped as "sub flow". These two components were designed here as pure "placeholders" for the purpose of process integration. If proper internal workflow systems are running at the participating company's sites, these two placeholders will initiate internal workflow when the project workflow executes to these points. Otherwise, the project workflow just ignores these two activities of this process type. While the ***Process Source Data*** process may consist of both manual and automatic activities, as discussed in Section 3.2, more human interaction is required than the ***Data Quality Inspection*** process discussed above.

Another example is the project document circulation process that serves the purposes of distribution, updating, revision and management of project related documents such as contracts, technical specifications, production procedures, and project reports (see Figure 4-8). In this process, the project manager plays the role of the mediator to facilitate activities involved. Among all activities assigned, the project manager can assign the "Review Document" activity to other workflow participants if needed.

While this section discusses issues regarding the process repository elements of the workflow model such as the organization of the repository, process selection and benefits of having a workflow process repository, the next section will focus on discussing other components of the workflow model, again in the context of collaborative GIS data production.

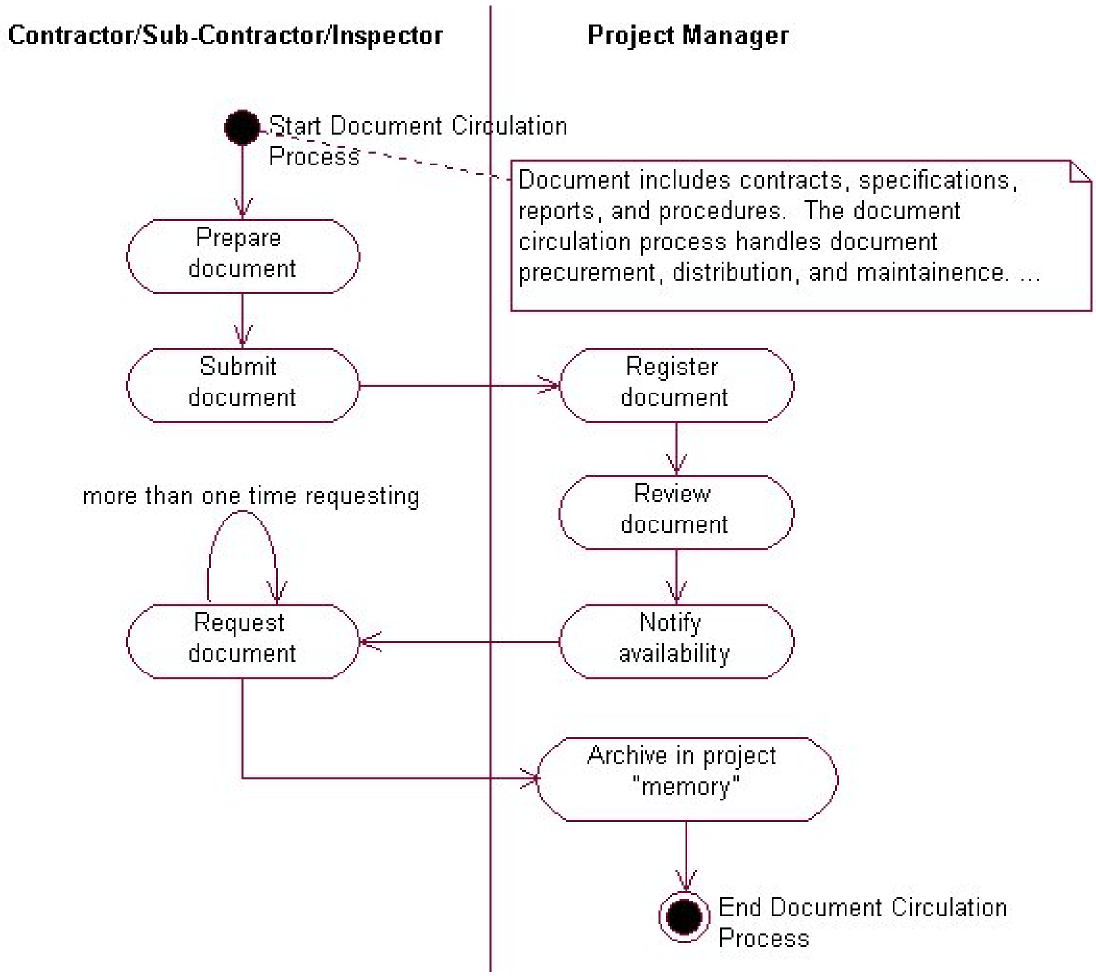


Figure 4-8 Document circulation workflow process

4.2.3 Discussion of Other Workflow Model Components

The process repository discussed in the previous section is the core component of the workflow model. It either uses or references the elements from other workflow model components, which provide necessary information required by activities specified in processes such as participants, associated applications and data. While these references or linkages may be expressed in actual workflow model specifications based on WPDL, the individual components are briefly described here.

Organizational Structure and Task Scheduling

Workflow participants were defined based on the organizational structure shown in Figure 3-1 (A) as discussed in Section 3.2.1. While all workflow activities were assigned to different functional roles in the model, assignments of activity instances to actual human performers were designed to be done at run time. To eliminate redundant process flow, subcontractors are treated as production contractors when participating in the redesigned workflow process.

There are several ways to schedule work tasks to workflow performers based on defined roles, for example *round robin*, *workload distribution*, or *manual assignment* [Chang et al., 2000]. With the first two methods, the task assignments are completely controlled by the workflow management system. For example, if there are three people acting in the QC inspector role, the system will either select one with minimum workload or assign tasks in turns. The last method allows explicit assignment to a specific human worker specified by a super QC inspector who also has a role as, for example, QC inspection manager. In this case, under each functional role, at least two roles may exist: management role and operation role as depicted in Figure 4-9.

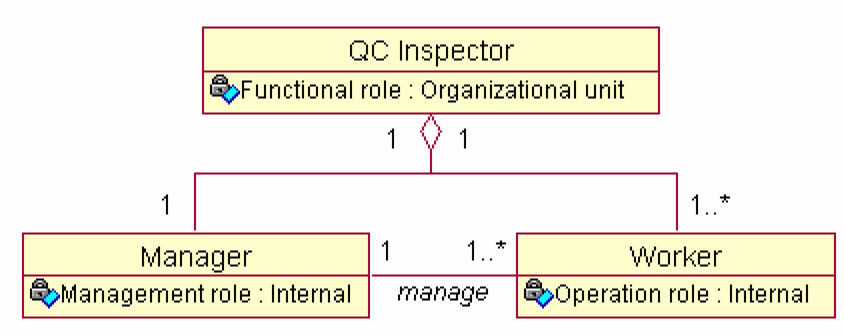


Figure 4-9 Example of split roles under each functional role

GIS data production workflow is highly restricted by delivery deadlines. Deadline tracking must therefore be reflected in the workflow model: (a) to warn production contractors and QC inspectors of appropriate deadlines, and (b) to notify the project manager of any overdue work. One example is to send timely reminders for submitting periodic progress reports (as illustrated in Figure 4-7). These kinds of time scheduling and tracking requirements are accomplished in the designed workflow model at both workflow process and activity levels through the inclusion of such attributes (workflow relevant data) as “expected duration” and “actual working time”, and handled by the workflow engine that is able to define deadlines for each process and activities.

Invoked Applications

With respect to workflow applications, there are two major types of applications that were included in the redesigned workflow process associated with actions within process activities. One type includes such applications as programs used for transferring data materials and project documents, and for generating project reports at various stages. In this case, applications are invoked by workflow performers manually and must include mechanisms to allow signaling the workflow system about the completion of work so the workflow execution can continue.

Another type of workflow application includes those that are invoked by the workflow engine automatically. For the ***GIS Data Quality Control*** example, the programs associated with the activities in "automatic QC testing" sub-processes discussed in the previous section are of this type. The workflow process may also include pre- and post-uploading and downloading applications that perform automatic tasks of data material packing (zipping), compression and decompression, organizing file directories, updating progress status, and hardcopy map plotting, if deemed necessary by individual projects.

QC testing programs (as workflow applications) can be categorized into three groups that perform "pass/fail" tests, "quantifiable" tests, or "partially-quantifiable" tests, which may or may not need human interaction during the testing procedures. As illustrated in Table 4-1, testing programs that have "pass/fail" or "quantifiable" results not requiring human interaction tend to be the candidate workflow applications that can be automatically executed and pass on workflow controls to the next activities. Other QC testing programs can be invoked by the workflow engine but need human interaction to either finish testing procedures or pass on workflow controls.

Table 4-1 Example of QC testing programs to be invoked

	Pass/Fail Tests	Quantifiable Tests	Partially Quantifiable Tests
No Human Interaction	***Fully-Automated QC Test Program***	***Fully-Automated QC Test Program***	Partially-Automated QC Test Program
Need Human Interaction	Partially-Automated QC Test Program	Partially-Automated QC Test Program	Partially-Automated QC Test Program

In the long run, QC programs needed by activities of this automatic QC process may be developed gradually and categorized based on different GIS data formats to form a kind of data quality testing program repository. This would need a software specification framework (e.g., following OGC[7] Abstract Architectural Specifications) to deal with platform and operating system compatibility issues and interface, especially an open interface to various workflow systems considering WfMC's standard - Interface 2 (see Section 2.2.3.2).

It is also desirable to investigate the corresponding QC detectable error model for building the QC program repository. Given the time constraints and the amount of work needed, this thesis research did not include further discussions on these issues.

[7] OGC stands for Open GIS Consortium; a non-profit organization participated by many GIS related organizations and companies. Information may be found from http://www.opengis.org/.

Workflow Data Model

The workflow data model presents two types of data: workflow application data and workflow relevant data, as well as relationships between these data. Different from workflow application data that is accessed or managed by workflow applications, workflow relevant data is defined at the workflow process level to be passed between activities or sub-processes to maintain their values. While some of the data values may not be changed in the scope of a workflow model such as data file identifications, some data values are changed by the execution of individual activities or sub-processes.

Based on the identified attributes as described in Chapter 3 and the workflow processes discussed above, Figure 4-10 presents the most significant data entries and relationships between them for the management of the workflow processes involved in GIS data production. Most workflow relevant data are among the circled data entries. Other data entries are mainly used by the related workflow applications. This designed data and relationship diagram does not include all data entries that could be involved in the designed workflow-enabled collaborative workspace since the actual implementation of this type of collaborative workspaces may bring more relevant data into the system. This will be further demonstrated in the next chapter when the research prototype is described.

Among the workflow-relevant data shown in Figure 4-10, three types of data entries play an important role in the execution of workflow processes discussed above. The data file identifier (DataID) is of the first type. Since the production workflow instances are created and executed for individual data files, this data identifier can be the most important identification to differentiate the workflow instance underway from others. The second type includes those entries (e.g., Actual_Submission_Date and Accepted_Date) that present important project milestones during the production life cycle of each data file. They are mostly used to specify “timeout” attributes of workflow and workflow activities. The last type includes the status attributes (e.g., StatusID and Status_Code). These data entries can be attached to the workflow process as activity attributes. Their values normally change, depending on the nature of the activities undertaken, during the workflow execution. The domain of the status data entry may vary from one project to another. However, the value domains described in Table 4-2 were identified from the existing projects.

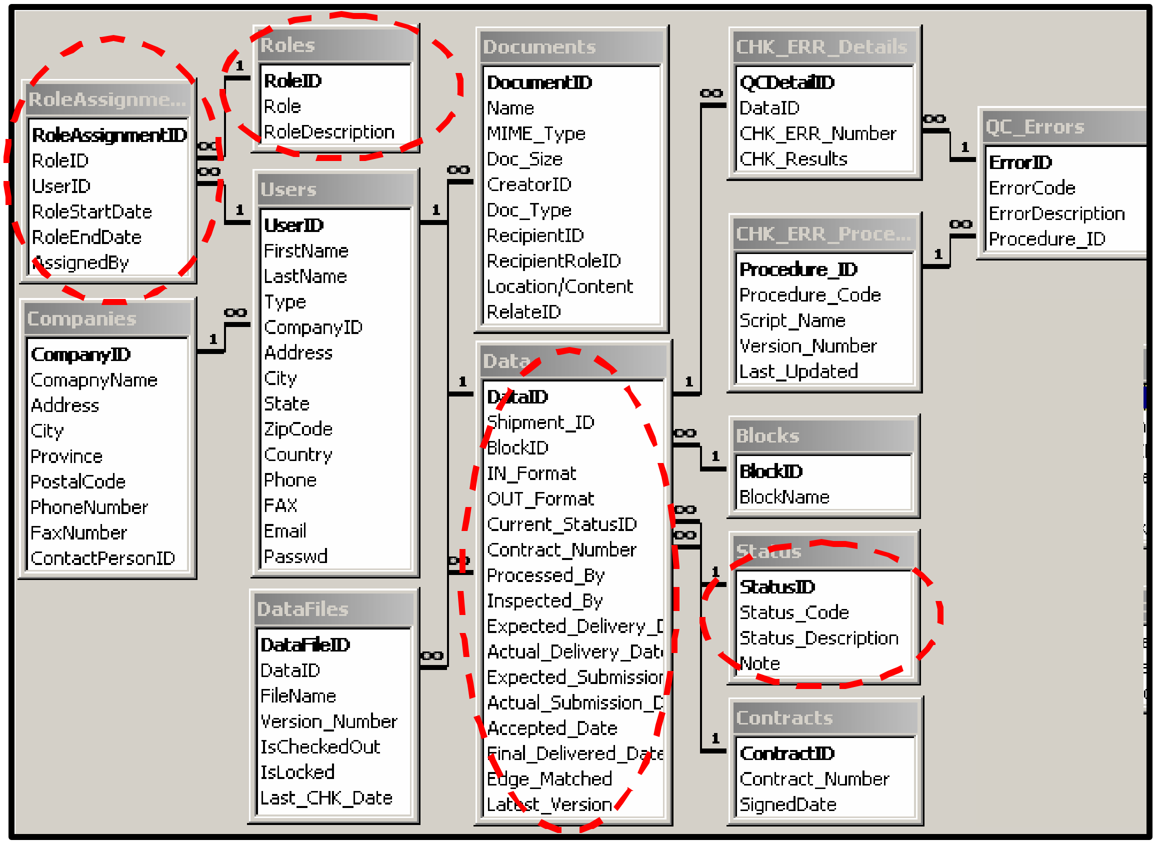


Figure 4-10 Data entities and relationships in the workflow data model

While UML diagrams (especially Activity and Class diagrams) are used in this section to describe some workflow model elements, the overall discussion is referenced to the MfMC's meta-models discussed in Chapter 2. The WPDL workflow modeling language is used to describe some of these discussed elements for implementation purposes. Some sample WPDL descriptions can be found in Appendix B.

Table 4-2 Value domains used for status tracking

Status Name	Code	Brief Description
Raw Data Waiting	RDW	Raw data has not been delivered.
Raw Data Delivered	RDD	Raw data has been delivered by the client.
Raw Data Validated	RDV	Raw data has passed data validation process.
Raw Data Production	RDP	Raw data has been delivered for processing.
Data Submitted	FDS	Produced data has been first submitted for QC.
Data Returned	DR	Submitted data has been returned for reworking.
Data Resubmitted	RRS	Reworked data has been submitted for inspection.
QC Passed – Level 1	QC1	Data has passed Level 1 QC procedures.
QC Passed – Level 2	QC2	Data has passed Level 2 QC procedures.
QC Passed – Level 3	QC3	Data has passed Level 3 QC procedures.
QC Passed – Level 4	QC4	Data has passed Level 4 QC procedures.
QC Passed – Level 5	QC5	Data has passed Level 5 QC procedures.
Data Accepted	DA	Data has been finally accepted.
Final Data Delivered	FDD	Final data has been delivered to the client.

Note: "Code" values are the abbreviations of the corresponding "Status Name" using the first letter of each word.

4.3 Architectural Design

This section discusses the architectural structure of the proposed collaborative workspace based on requirements specifications defined in the Chapter 3. The description of the system architecture starts with a brief illustration of the architecture's key concepts. This is followed by an overview of the architecture. Then the components of the approach are described in more detail with possible network connection and hardware mappings.

Basic Concepts for IICW Architecture Design

There are several key concepts that are important for understanding the discussions in this section. These concepts are presented here before describing the actual architecture's structure of the collaborative workspace.

Internet-based integrated collaborative workspace (IICW): A collaborative workspace, as defined in Section 2.1.4.2, has an integrated GUI interface or a window such as a web browser as the front-end and a set of collaboration functions (groupware components), a shared database, a collection of files and file directories and a workflow management system as the back-end. The workspace should be able to invoke external geoprocessing applications or components through the workflow systems based on workflow process definitions. All the production project participants access the IICW through its front-end user interface (see Figure 4-11).

The information comes with various explicit or implicit inner structures from the system components; for example, native presentation from individual groupware tools or table forms from databases. As the WWW is currently the most successful structured information space and requires little effort to understand, most information, if not all, is synthesized into HTML-encoded hypermedia format for unique presentation within IICW. However, the architecture's structure is not restricted solely to the Web-based client/server model, as discussed shortly.

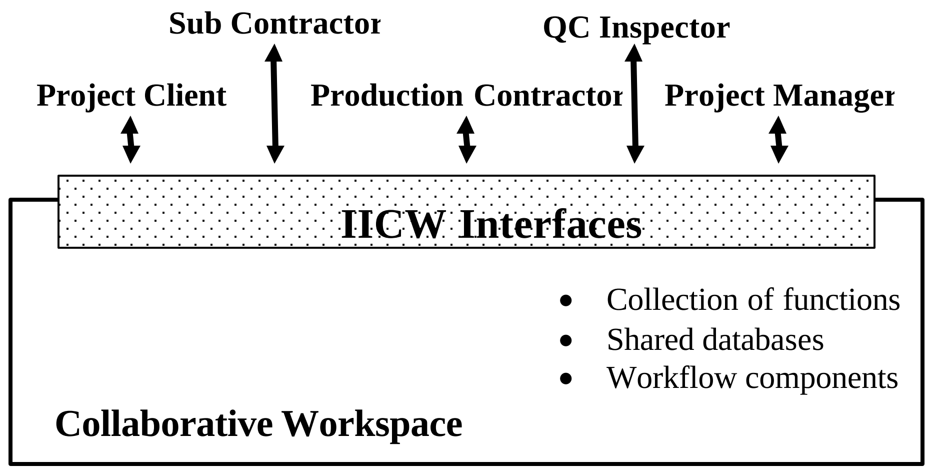


Figure 4-11 Access to the collaborative workspace

Integration Mechanisms for existing components: The architectural design of the IICW focuses on describing the structure of involved components - software and hardware, the relationships between them, and the way in which they interact with each other dynamically. There are a variety of ways to integrate tools from different sources. Integrating existing software components may be more efficient, reliable and effective simply because there is no need to reinvent the wheel yet the software quality is ensured. However, not all required software components are commercially available although most required functions are available from stand-alone software packages (see Section 3.4). Therefore, adopting the concepts of the "adapter" or the "object mapper" is helpful for bridging the gap between software components and stand-alone tools when a hybrid system is desired [Shih et al., 1996].

Domain software integration: Existing domain-level tools in general do not have proper interfaces or an API that is able to interact with CSCW or information management tools.

This fact needs to be taken into consideration. By “domain-level”, it means those components that are directly used to handle GIS data during data production processes (e.g., GIS packages) and those that are integrated in IICW for sharing data views and communicating data problems (e.g., services provided for displaying a status map and electronic map whiteboard).

Given the variety of CSCW tools available, it is not feasible to implement a set of unique interfaces or APIs that can interact with different CSCW tools. There are two ways to make this kind of integration an easier task by using some *de facto* industrial software interoperability standards such as OpenGIS’s Architectural Specifications (http://www.opengis.org) and the Workflow Reference Model from the Workflow Management Coalition (http://www.wfmc.org). For example, APIs may be implemented in GIS tools that can talk to the interchange interfaces specified in the Workflow Reference Model. In turn, the workflow management systems can communicate with the same interchange interfaces as illustrated in Figure 4-12.

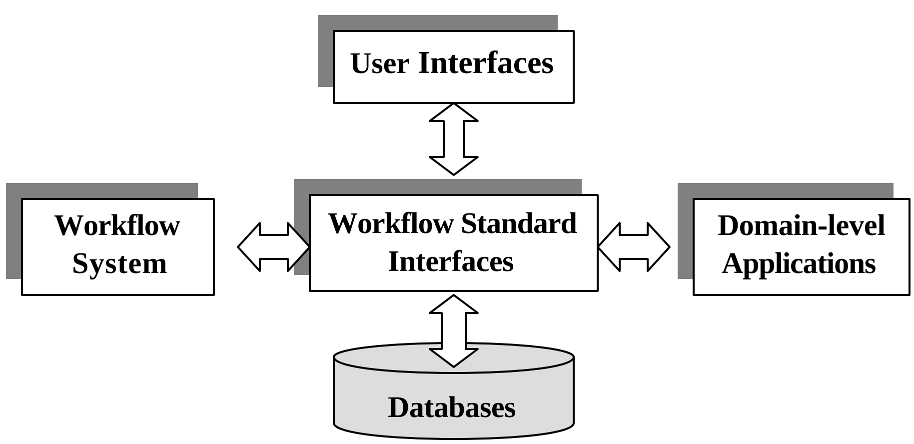


Figure 4-12 Coordination of workflow system and domain (GIS) applications

Abstraction level: The architectural model was designed as a conceptual representation of the system to be built instead of the system being built or the system that was built. Although hardware and network nodes are more closely related to implementation, the inclusion of these components helps describe computer platforms and their physical connections, as well as service levels. As such, the model was designed to serve the purposes of (1) allowing easy selection of integrating software components or packages; and (2) providing a basis for selecting physical computer systems and a network model, in which the integrated collaborative workspace will execute.

With the above premises, the following sections introduce the main artifacts of the IICW architecture structure.

4.3.1 Logical View of IICW Architecture

The overall architecture follows the Internet-based client/server structure that contains three main component groups: client components, servers and connections. While the network connections between IICW client and server components use standard Internet protocols (e.g., HTTP, FTP and NNTP), internal connections among server components are service calls realized through component interfaces. Although the discussions in this section focus on the Web-based client, access to the IICW is not only restricted to this type of client interface. The overall structure can also be applied to PC-based client or API-based client that may be used by domain-level tools.

A component-based approach with a layering style is adopted to describe the IICW architecture, of which all components can be divided into three major layers, namely presentation layer, business logic layer, and data access layer. The presentation layer is built on top of the business logic layer that depends on support from the data access layer (see Figure 4-13). While each component involved is treated as a "*black box*" with interfaces and services specified for the external access, its implementation is realized usually with more than one component or object as discussed in the next section.

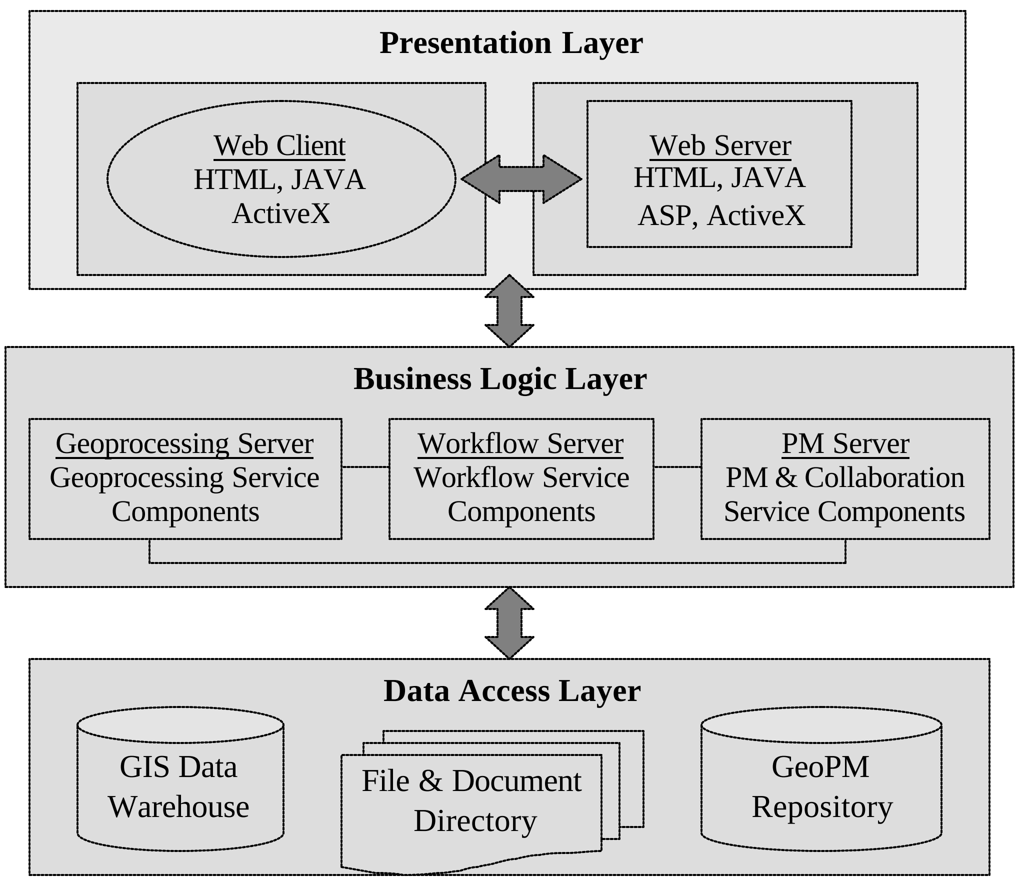


Figure 4-13 Architectural model of the collaborative workspace

The *presentation layer* holds two groups of components: the Web client component and the Web server component. The Web server component uses technology such as HTML, JAVA, ActiveX and ASP to provide web-based services to IICW users and the client component uses the same technologies or HTML-coded outputs from the Web server component to interface with users.

The *business logic layer* implements two main types of IICW components, servers (or engines) and services, which are grouped into three subsystems: geoprocessing, workflow, and project management (PM). Geoprocessing subsystem encompasses those components that provide GIS related services to be executed on the server-side of IICW. While the workflow subsystem possesses components that deal with all kinds of workflow services provided through the workflow engine (the Workflow Management Coalition's website provides a full list of recommended services), the PM subsystem provides services for both project management and project cooperation activities.

The *data access layer* contains mainly database and file/document directory components. The project repository has both operational and analytical databases that hold data structures and project structural information. The databases also provide data access services upon request calls from the components on the business logic layer. The file/document directory components provide both storage and access services for physical GIS data files and project-related documents. If spatial and project metadata data warehouses are implemented, the file/document directory component communicates to the data-warehousing component. In the case of a database system supporting storage of large computer files (e.g., Oracle Database system), the file/document directory component may be eliminated.

With this architectural structure, the inter-connections between the server/services components on the business logic layer are made possible through service calling and project information transfer. The PM services can call the geoprocessing services through the geoprocessing server to initiate a GIS-enabled collaboration session – for example – an electronic map whiteboard session to discuss project-related issues. The workflow server component can also call geoprocessing services that are used for certain automatic server-side processing. For example, some data quality inspection testing procedures may be implemented as geoprocessing services and invoked upon the submission of the processed GIS data files. In any of these cases, if the involved components do not have interchangeable interfaces, the component adapters have to be included to transfer required service and information requests between the participating software components.

The above system architecture was designed based on a component model, which provides guidelines for integrating off-the-shelf software components from various vendors to develop the collaborative workspace. In many cases, it may be necessary to use several stand-alone software packages in a collaborative work environment simply because the required off-the-shelf software components are not available from the market. For example, a CARIS TM, Microsoft Project Manager TM, and Oracle Workflow TM system, plus web browser and the Internet infrastructure may be integrated. In this case, the needed system architecture is different – a separate layer consisting of object mapper and internal controls would have to be included [Shih et al., 1996]. In this case, an alternative architecture may be described as in Figure 4-14, where the object mapper provides access to information stored in the databases – both spatial and non-spatial – within the project repository.

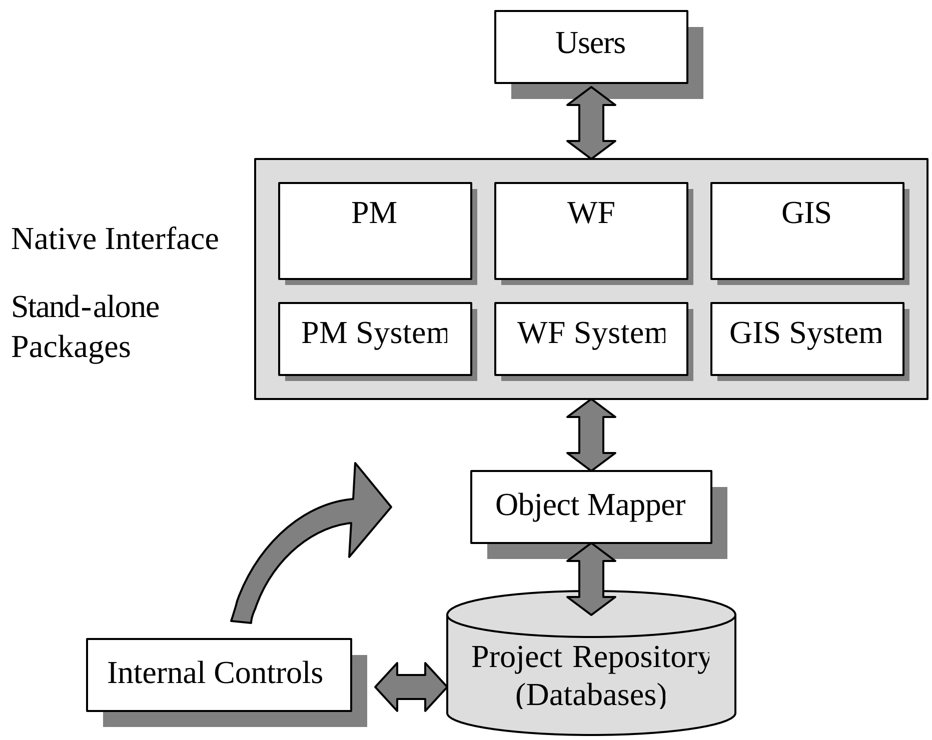


Figure 4-14 Alternative structure for integrating stand-alone software packages (Adopted from Shih et al. [1996])

Other situations may involve a hybrid integration of both off-the-shelf software components and stand-alone software packages. Since the focus of this research is on providing a component-based architecture structure, these variations are not further discussed in the context of the architecture model of IICW. This issue will be addressed in the descriptions of the research prototype presented in Chapter 5.

4.3.2 Architectural Components

In this section, the major functionally significant components of the IICW architecture are described at higher level without getting into much implementation detail, meaning that each component may be further decomposed into several smaller components if needed. The possible mapping between these components and hardware units required is briefly discussed. The descriptions comply with the requirements specification discussed in Section 3.3. Figure 4-15 illustrates an overview of these software components organized into several subsystems, where "dependency" relations are denoted as dashed lines with an arrowhead (UML notation). All components are categorized either as "server" component or as "service" component.

The *IICW Web Server* subsystems contain three "server" components and one "service" component. The Session Manager component plays a unique role in the architecture. Services provided by this component are primarily logon/logoff procedures for user authentication, session timeout and, if cookies[8] are used, user's access information. The access to all services provided by the IICW has to go through this session manager. Once

[8] A cookie is a mechanism that allows the server to store its own information about a user on the user's own computer.

the client information is verified, the client request is passed on to the proper server component.

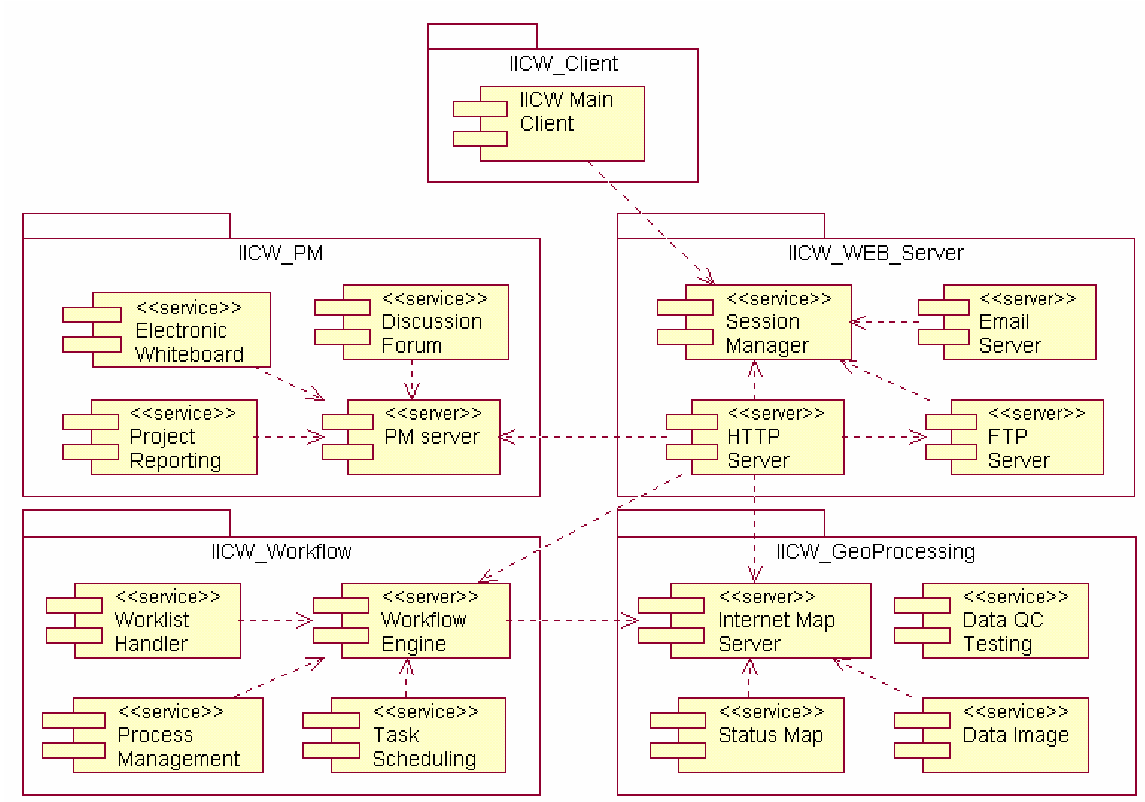


Figure 4-15 Overview of IICW functionally significant components

The HTTP server is a hypermedia server that stores and distributes information written in HTML. The basic service provided by the HTTP server is to relay a client's request to the IICW server components and respond to the client when a response to that request is received. It is also responsible for responding to those client requests that are not handled by the IICW server components such as providing information HTML pages.

The Email server manages the asynchronous flow of messages between individual project participants. Since services such as official correspondence provided by the project reporting component and open discussion provided by the discussion forum are based on Internet mail protocols (POP and SMTP), the Email server really acts as the project messaging gateway.

The FTP server provides file transfer services for the synchronous exchange of project data files and documents. This service can be enhanced with a Web-based FTP client to allow users to attach extra information with the transferred files, and to schedule a later time for transferring. In this case, the service from the session manager is required to ensure proper session connections.

The *IICW Geoprocessing* subsystem has two “service” components that are dependent on the Internet map server component. These two components provide GIS data services used for information sharing and project management purposes. For example, the data image component provides data images for previewing or for map whiteboarding. Another “service” component is responsible for providing automatic server-side data quality testing at different levels. Although these services are mostly called by the workflow engine component, there are cases where the project manager or QC inspector wants to run individual services to verify certain QC results of GIS datasets.

The *IICW Workflow* subsystem contains three “service” components and one “server” component. The workflow engine provides run-time execution services for any workflow instances based on the workflow process definitions as discussed in Section 4.2

(a full list of the services a workflow engine should provide may be found from the Workflow Management Coalition's workflow reference model).

The worklist – the queue of work items assigned to a particular performer by the workflow engine – is accessible to the workflow engine for assigning work items and to the worklist handler for retrieving work items. The worklist handler should at least provide services for fetching the worklist and presenting it to the IICW web client, and invoke services from other IICW components to retrieve work items. The process management component mainly provides services to input, edit, and monitor the workflow process definitions. Through its services, the client is able to suspend and stop any process instance, and resume any suspended process instances.

The task scheduler component actually belongs to both IICW PM and IICW Workflow subsystems. For the IICW Workflow subsystem, it provides services to allow access to performer and performance (e.g., work time and waiting time) information of the workflow process. Rescheduling workflow activities to different performers based on the process execution can also be done through the services of this component. For the IICW PM subsystem, the component provides services that enable other project-related task scheduling and ensures proper project calendar entries are referenced by calling the PM server services.

The *IICW PM* subsystem encompasses components providing both project management and collaboration services. The PM server manages audio/video conferencing sessions, provides agent service to check and update project progress (e.g., Gantt chart) on a real-time fashion by calling proper services from the workflow engine, and maintains project discussion forums. The PM server also manages a project calendar so that the IICW client can access it and initiate new calendar entries.

The electronic whiteboard component provides services for both synchronous and asynchronous whiteboarding sessions. The clients can call these services to initiate discussion sessions with other project participants sharing the same graphical view of a particular problematic data file. It is this component's responsibility to ensure that proper service from the data image component is called and images are obtained. In case a collaborative map data markup session is desired, the component provides services to update multiple views and record markups and annotations.

All these architectural components may be installed on different network nodes and connected with the Internet infrastructure by adopting the Extranet model. However, given the lower number of project participants, lower volume of simultaneous access to the workspace, and the temporal consortium nature of the involved projects, it would be more feasible and efficient to have all server-side components installed at the same location, if not on the same network server. Figure 4-16 illustrates the simplest deployment scenario that allocates involved software, hardware and network connections in a web-based collaborative workspace.

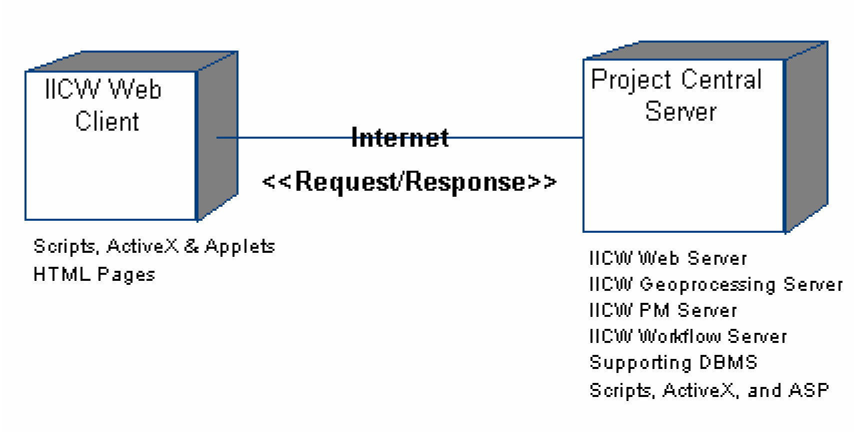


Figure 4-16 A simple deployment scenario for a collaborative workspace

4.4 Implementation Framework

The successful implementation of workflow-enabled collaborative workspaces to a large extent depends on how implementation processes are planned, organized, and conducted. Based on lessons learned from existing CSCW application implementation practices and problems encountered during design and development stages of the proposed collaborating systems, this section discusses a framework for collaborative workspace implementation processes, which can be used by project managers and developers to deploy collaborative workspaces through structures of implementation processes.

The implementation process presented here focuses on the deployment of workflow and architectural design discussed in the previous sections, starting with detailed process and requirements modeling which deals with individual business requirements of the data production projects at hands. Since most GIS data production projects will go through, formally or informally, the steps of contracts negotiation, technical (data) specifications, specification validation, and contract execution, the implementation process should be done after technical specifications are defined and before the contracts are executed – if possible, parallel to the specification validation process. Rather than being detailed, the framework captures some major steps including process modeling, software/component selection, workflow/tools deployment, and pilot testing (see Figure 4-17). These are further described in the following with actions involved and problems to be considered during the implementation process.

It should be noted that the implementation process for any CSCW-related application is driven by many contextual forces, such as organization, technology, users, and work as defined by Sanderson [1992]. This section presents an implementation process that is mainly driven by technology required and work to be done for distributed GIS data production projects.

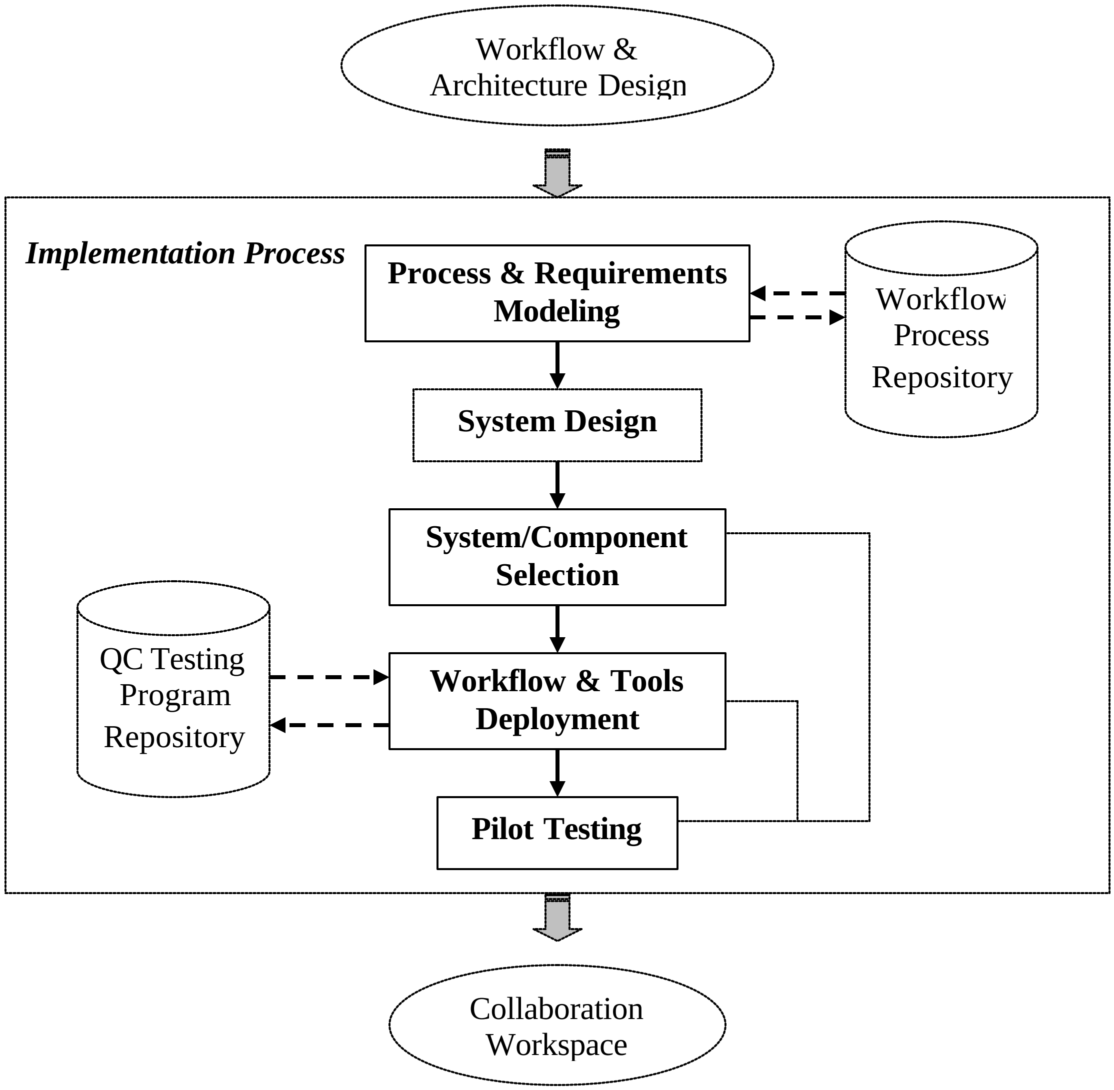


Figure 4-17 An implementation framework for a collaborative workspace

The ***first phase*** of collaborative workspace implementation processes deal with gathering detailed information relevant to individual collaborative GIS data production projects. Empirical studies based on interview, workshop, and interactive presentation techniques, as well as analyses of collected documentation (such as forms, reports, memos, etc.) are used. The information collected is used as the basis for selecting proper process templates from the process repository (if there is a pre-populated one) and modifying selected processes, where appropriate, to accommodate special requirements of the project to be accomplished.

In the case where no appropriate process templates can be found in the repository, process modeling involves specifying extra data production processes based on the information gathered and analyzed. These new processes can then be added to the process repository, which will increasingly enhance the usability of the process repository for accomplishing later projects. The process constructs of the repository have to be observed by all added processes to ensure easy-to-understand representations using formal modeling notations (e.g., UML or WPDL) so that information system experts and domain experts are able to validate and redesign production process models to achieve added value later through utilizing collaborative workspaces.

Aside from the process modeling, other collaboration requirements have to be identified at this stage. Although the general requirements are discussed in Chapter 3, there are still some project-specific requirements to be dealt with on an individual basis, such as those non-functional requirements, for example, requirements on platform (using existing platform or implementing new one?), network model (Intranet or Extranet?) and performance (critical or not critical?). This information is definitely crucial for the following phases.

The ***next phase*** is to design the project collaborative workspace, where the architectural model discussed in the previous section can be used as a reference. Since the collaboration model was developed based on the understanding of generic characteristics obtained from the existing data production projects, its workflow model and architectural structure can be modified based on the identified requirements from phase one. Workflow processes are designed at this stage by applying project rules, associating involved application programs, and assigning priorities of process automation.

During the ***third phase***, the software systems and/or software components have to be selected, evaluated, and licenses purchased based on the results from the system design stage. Given the diversity of CSCW software and software components in the market, the major objectives are to acquire appropriate software for the implementation of the collaborative workspace which will be used for best supporting your project. In order to do so, additional criteria have also to be defined to guide the selection process and evaluate the selected software, such as software integration, platform compatibility, run-time and network performances, and deployment flexibility.

Among these criteria, integration may be the most important considerations due to the difficulties discussed in Chapter 3. How do you want the involved software systems/components integrated, loosely coupled or seamlessly integrated? How will other components interact with the workflow engine and how will application data and relevant materials flow in the project by executing workflow processes, especially when the workflow system is considered as a "glue" component to integrate other components of the collaborative workspace? Should, for example, QC testing programs be invoked automatically by the workflow engine based on the workflow definitions or manually upon the confirmation from an operational role to a workflow notification? All these kinds of questions should be answered before the selection process starts. Criteria regarding software integration should specify both software and software-related data integration aspects, especially in dealing with domain-specific components that handles GIS data.

GIS installations for data production projects are almost always determined by the nature of the data at hand and it is not likely that they can be changed for the sake of implementing a supporting collaborative workspace. Therefore, it is very important for the selected software to be appropriate to the existing GIS installations. For example, when selecting the file management system/component, the fact that GIS data files are usually very large has to be considered carefully.

The procedure of software selection may include such activities as selecting software, evaluating and testing the selected software against system design and pre-defined criteria, and making selection decisions. In many cases, if the software systems/components available in the market cannot satisfy pre-defined criteria, they may be modified and used for testing the selected software again.

Given the fact that sometimes the results from the following phases may reverse the system design and software selection, it would be wise to use a demonstration version of selected software systems or components. In this way, if any selection has to be reversed or proves to be in appropriate for the project by later phases it will not increase the cost occurred for acquiring software.

The ***fourth phase*** involves actual deployment of workflow processes with the corresponding rules and roles (as defined in the workflow model) into the selected workflow management system, and software integration of selected tools for supporting project management and collaboration. Depending on the software selected, activities in this phase often involve a considerable amount of coding and software testing. Most effort is in coding proper interfacing/interacting components for selected software components and domain-specific processing components that are not available, such as QC testing programs.

Finally, the ***last phase*** is to test the implemented collaborative workspace, whose overall goal is to obtain information about the usability, the technical stability and the organizational suitability of the collaborative workspace. In order to achieve the goal, certain testing scenarios have to be defined based on the project requirements and contract technical specifications. This is also why it was suggested the implementation process to be done parallel with the validation process of the technical specifications at the beginning of this section.

It is better to test the implementation in two steps: lab simulation and field test. The lab simulation testing focuses on identified requirements and designed workflow process models and defining testing scenarios while the field testing is to verify if the collaborative workspace is able to handle real world situations by using a sub-set of GIS data to be produced with selected major production processes. During field testing, the testing environment should be real and involve actual system users and all critical project information. In reality, the field testing phase is also a suitable time for training potential users of the collaborative workspace although the use of this type of system should not be difficult to learn.

4.5 Summary

This chapter has contributed to the design and development of the proposed collaborative workspace model, focusing on workflow and architecture design. To present the designed artifacts, UML and WPDL modeling languages are used. During the model development, emphasis was given to higher-level presentations of logic relationships among the designed components. The chapter demonstrates that existing practices of distributed GIS data production projects should be considered as the most essential sources for modeling the required workflow. Although the reality check of existing projects is limited in this thesis research, the information obtained is sufficient for model development. However, the lack of specific target projects for which these models will serve imposes certain restrictions on developing model details, especially those related to implementation of the model. In this sense, the developed model is considered as a higher-level one rather than one that can be directly applied into any specific projects.

Several new concepts are discussed in this chapter. One concept presented is to use a Process Repository to organize involved workflow processes in the workflow model. A framework of this Process Repository is also described and related issues are discussed. Many advantages have been identified for using the concept of a process repository, among them, the inclusion of the best industry practices, process standardization, and establishment of project memory are most significant. Another concept introduced is the QC testing program repository, which is suggested to be the foundation for managing external applications that could be attached to the QC workflow.

The hybrid architecture was briefly introduced, and is based on a component structure. The implementation work to be described in the next chapter demonstrates that the architecture of hybrid systems is largely affected by the selected software systems, especially the workflow management system. In this respect, the component-based architectural design is not only more generic in nature, but also provides guidelines for selecting software systems if the hybrid approach is adopted.

Chapter 5 PROTOTYPE SYSTEM: IMPLEMENTATION AND TESTING

This chapter is devoted to the implementation and testing of the concepts discussed in the previous chapters, especially the collaboration model developed in Chapter 4 The prototype (called *GeoPM*), built as an Extranet-based client/server system over Internet infrastructure, was primarily developed for the purposes of proof-of-concept, verification of the model's usability, and facilitating the test of research the hypothesis within the context of a GIS data quality control process.

This chapter is organized into four parts. It first presents a description of the prototype development, followed by discussions on prototype implementation - three case studies presenting software realization of project information management, communications, collaboration and work coordination within a collaborative workspace. It then presents some simulation testing of the prototype performance. The chapter concludes with a final discussion on the results obtained and problems identified.

5.1 Development of Prototype System

An incomplete prototyping approach was selected in the prototype development, implementing some of the selected key functional requirements that were defined with top importance level as discussed in Section 3.3. These functions focus on demonstrating the following capabilities of a collaborative workspace:

(1) *Workflow capacities:* status tracking and querying of project activities on a real-time basis, task and document routing, task assignment based on roles and rules, automatic triggering of batch processing programs, e.g. feature code checking, based on pre-scheduling, and report and notification routing.

(2) *Collaboration capacities:* email correspondence, discussion forum, whiteboarding, audio-conferencing, and personnel and resources scheduling.

(3) *Other capacities:* centralized project repository for storing data files, documents, and project-related information, allowing data file check-in and checkout, directory browsing, and project attribute query, statistics, and summary.

The prototyping efforts in this research emphasized the data quality control process and its related requirements. The selected data quality control workflow captures all activities from initial data (un-inspected data product) submission through inspection to the final data delivery to the client. Whenever a production contractor submits a data file, a separate workflow instance will be created based on the workflow process template and started by the workflow engine. The selected workflow process was designed based on the contract-based production model where client, production contractor, QC inspector, and project manager are the primary roles involved. To simplify the prototype implementation, process integration was not considered, i.e., there were no interactions between different workflow management systems. All workflow processes were executed by the solo workflow management system installed on the centralized server.

5.1.1 Prototype Components

The components of the prototype system architecture (as discussed in Chapter 4) include a web server, a database management system (DBMS), an Internet map server (IMS), a workflow management system (WMS), server-side scripting or Visual Basic (VB) programs, a web browser with downloadable ActiveX Control[9] components on the client side, as well as corresponding hardware and networking components. The overall configuration of these prototype components is illustrated in Figure 5-1, where server-side components are allocated on two separate computers: TYR and GEOPLUS. On the client side, only a web-based IICW client interface component was implemented in the prototype.

As discussed in Chapter 4, the architecture design of the collaborative workspace employs a component-based approach. The initial investigation of existing software market revealed difficulties in obtaining required components due to the commercial strategies followed by the software vendors and varieties of interface standards used (see Chapter 3 for details). Therefore, the actual software implementation of the prototype adopted a hybrid approach by integrating available software components, especially programmed components and existing software systems. The software selection took into consideration several factors such as fast realizations of required functions, cost of prototyping, and availability of required software components and systems.

---

[9] ActiveX controls, formerly known as OLE controls, are the Common Object Model (COM) based objects that can be used to develop window applications.

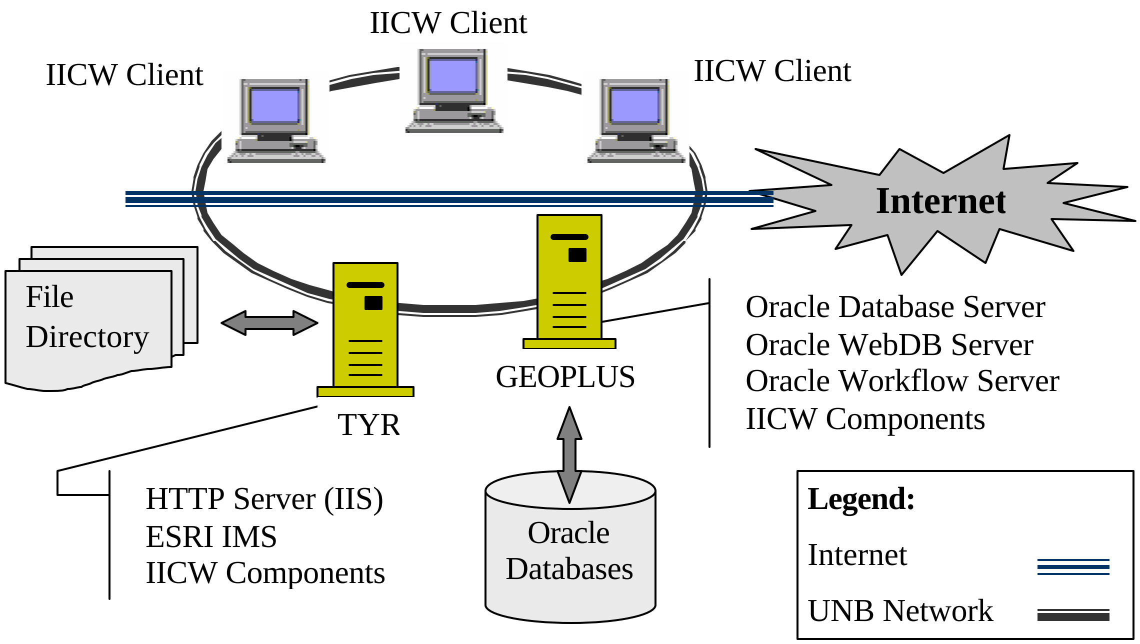


Figure 5-1 GeoPM system configuration

Software Components

Microsoft Internet Information Server (IIS) bundled with Windows 2000 Professional operating systems was the Web server selected for the prototype system (see Windows 2000 Help Documents at http://localhost/iisHelp/). The 5.0 edition of IIS is a built-in component which allows only one web server instance and one FTP server instance. IIS 5.0 can also be installed under the Windows NT Workstation operating system by installing Service Pack 6.a. The initial testing of the web server was undertaken both on Windows NT Workstation 3.5 and Windows 2000 Professionals platforms. The results show that the IIS 5.0 version is sufficient to support the research prototype. In addition to the limited number of server instances IIS can create, it is also not supported by UNIX operating systems.

A use of the Netscape Enterprise Server software supported by both Windows and UNIX systems was explored in the preliminary research stage [Li, et al., 1998; Boettcher, 1999]. Although the Netscape Enterprise Server was identified as an alternative web server for supporting *GeoPM*, it was not selected for the final development because of the lack of licenses available at UNB and the lack of funds to purchase additional ones.

ESRI MapObjects IMS (version 2.0) is a tool for authoring, customizing, and administering Internet mapping applications, and serving users' mapping requests through web browsers [ESRI, 1998]. It is comprised of programs, utilities, and files that can be installed on distributed computers or used in a single system environment. IMS works with two of the major commercial web servers discussed above through a key dynamic link library (DLL) component – "esrimap.dll" for Microsoft IIS or "esrimapn.dll" for Netscape Enterprise Server. Services provided are realized through the server-side applications developed by Microsoft Visual Basic or C^{++}, within which ESRI MapObjects (to be discussed shortly in Section 5.1.2) is set up as a separate project component. As such, all services to be provided through IMS have to be programmed as stand-alone applications beforehand and set up in the IMS administration environment. Whenever a mapping request is received by the web server from the IICW web client, the web server will pass on the request to the IMS server. The IMS server will then instruct appropriate IMS services to generate the required response and send it back to the web server. Eventually, the web server will send the response back to the IICW client in HTML formats.

In developing IMS service applications, a fundamental ActiveX control provided by MapObjects IMS (called WebLink control) was used and added to VB programs in the VB integrated development environment (IDE). The WebLink control is responsible for establishing and managing mapping service sessions, and receiving and parsing mapping service requests from the web server.

Oracle Database Server (version 8.*i*.16) is an object-relational client/server based DBMS system that was selected for storing all project related information as well as supporting the selected Oracle Workflow System (as described shortly) [Leverenz et al., 1999]. Since the datasets used in prototyping are mostly in CARIS data formats, the Oracle databases are not used to store GIS data files directly. Only information (i.e. metadata) describing GIS data files are stored and managed in the Oracle databases.

Oracle Workflow Server (version 2.5) was selected as the supporting WMS system to provide workflow related functions [Chang et al., 2000], although a large amount of work was done in reviewing and evaluating several workflow systems. This Internet WMS system relies on an embedded workflow engine in the Oracle Database Server system and interfaces with the workflow participants through a notification mechanism. Participants can have their own worklist that contains all notifications associated with the tasks to be attended, where rules for sending notifications are pre-defined in process definitions through the Oracle Workflow Builder – a native process definition tool. The system allows monitoring of the status of individual process and activity instances both graphically and in textual summary form on a real-time basis. With the Oracle Workflow, one can use the full power of PL/SQL™, which is the language of the Oracle Database Server, to implement sophisticated business logic.

In order to coordinate the server-side applications of *GeoPM* on both TYR and GEOPLUS computers (see Figure 5-1), an Oracle Workflow Client and an Oracle Database Client are also installed on TYR to enable server applications to interact with the Oracle DBMS and the workflow server on GEOPLUS.

Microsoft Internet Explorer (version 5.5) is the web browser selected for the prototype since some client-side downloadable components require Microsoft ActiveX controls, which are not currently supported by other web browsers with their native capabilities such as Netscape Communicator. Although a special plug-in may be pre-installed for Netscape browsers to support ActiveX controls, this was not tested in the prototype development.

Server Scripts and Applications include those scripts or compiled VB programs that are developed in this particular project and deployed on the centralized project servers as either separate components or add-ons to the selected software systems discussed above. These programs will be described elsewhere in this chapter where appropriate.

CARIS for Windows ™, a GIS software package, was selected as the existing GIS installation since all datasets available to the project are in CARIS formats. This package is mainly used to perform necessary GIS data preparation and processing for the purpose of prototyping and testing and its macro modular programs are used to compile some batch processing procedures, such as those for QC testing purposes.

Hardware and Networking Components

Hardware configuration includes a dedicated high speed (at the time of the prototype development) IBM PC with 17" monitor (TYR) and a shared Windows NT server (GEOPLUS). TYR was used to perform all prototype development tasks, as well as to host the ESRI *MapObjects* ™ Internet Map Server and the related server applications at the deployment stage. Started from October 2000, GEOPLUS – the department IBM server – was used to host Oracle Database Server, Oracle Workflow Server and Oracle WebDB Server, where the Oracle WebDB is used to develop and manage *GeoPM* web site and its associated web pages. At the time of the prototype deployment and testing, the server was also used to host other teaching related software systems such as PCI and *CARIS* Spatial Fusion ™. However, the access load was not very significant due to a lower number of simultaneous access requests to the server. Several other desktop computers in the Geographical Engineering Group (GEG) laboratory at UNB were used to access the prototyped collaborative workspace. Table 5-1 summarizes major specifications of these computers.

Table 5-1 Computer Workstations and Server used in the prototype

Specs	TYR	GEOPLUS	workstations
Model	IBM	IBM Netfinity 5000	IBM or DELL
OS System	Win 2000 P	Win NT Server 4.0	Win NT 3.5
CPU Speed	667 MHZ	450 MHZ	200 – 450 MHZ
Memory	262 MB	320 MB	64 – 128 MB
Storage	15 GB	45 GB	4.3 – 15 GB
Network Card	3 COM Ethernet XL 10/100 PCI	AMD PCNET Family Ethernet Adapter 10/100	3 COM Ethernet XL 10/100 PCI

All computers involved in the prototype development and testing are connected through the departmental LAN network system and to the Internet through UNB campus networks. The accessibility and bandwidth is treated on an as-it-is basis, i.e., depending on the existing Internet connection (NBNet - T3 with 45 Mbs speed). Because the processes involve transfers of huge amounts of digital data, sometimes on a real-time basis, the bandwidth factor is especially important. This will be further discussed later in Section 5.3.

Discussion

In summary, selection of all aforementioned prototype components, especially software components, were based on the architectural design discussed in the previous chapter. This, however, involved a lot of work on reviewing product documents, installing, evaluating, and testing selected software. In many cases, software vendors had to be contacted through a number of email correspondences and phone calls. Among the work, the most difficult task was to determine an appropriate workflow component due to the high cost of commercial workflow systems, unavailability of evaluation versions, and technical and implementation complexity of most workflow systems.

Several reviewed commercial workflow systems indicate a price range from $20,000 to $40,000 (based on the quotations obtained), subject to the number of seats for servers and users. In addition, most workflow system vendors do not provide an evaluation copy of their systems. Therefore, the evaluation work was conducted mainly based on the product literature and results from personal communications to vendor's technical support staff, which was very limited. In fact, even if evaluation copies could be obtained, it would have been very difficult to actually install and try them due to the technical and implementation complexity.

5.1.2 Programming Work

The prototype development involved using a number of development tools and programming languages. In general, several considerations were taken in selecting tools and languages such as availability, compatibility with deployment systems, and level of learning curve needed in order to use them. The selected development tools and programming languages include Microsoft Visual Basic Professional (v 6.0), ESRI MapObjects (v 2.0), Oracle PL/SQL language, and scripting languages such as VBScript and JavaScript. While examples of program code are included in Appendix C, these development tools and programming languages and corresponding development work are briefly described in the following.

Microsoft Visual Basic Professional is an integrated development environment (IDE) based on Visual Basic computer programming language, which is very easy to learn. MS VB has a rich collection of pre-defined foundation classes which provide basic building blocks for developing various types of VB programs, such as stand-alone executables, ActiveX controls, ActiveX DLL, and ActiveX documents. In addition, there are many free VB programs and ActiveX components with their source code available for downloading over the Internet. However, the reliability of this freeware code has to be verified before integrating them into the applications under development. In this prototype development, MS VB was used to particularly develop several ActiveX controls that can be deployed as Internet applications and accessed by IICW clients.

The developed ActiveX controls build on not only basic building blocks provided by the MS VB IDE but also ESRI MapObjects components (to be discussed shortly) associated with VB IDE. The deployment of these controls to IIS and IMS allows *GeoPM* to realize two groups of services: mapping related services and data file check-in and checkout services. Mapping service controls were built on top of ESRI MapObjects Mapping components and other VB IDE built-in components to provide such functions as previewing data contents, generating hardcopy map plots, displaying and querying status map, and providing geographical (e.g., index map) referencing to other projects information such as reports and documents relevant to specific data files. Built on top of

Microsoft Internet Transfer Control and other components, data file check-in/checkout services provide interfaces for an IICW client to check in/out data files with necessary attributes and for *GeoPM* to manage checking in/out processes on the server side.

ESRI MapObjects ™, a widely used GIS toolkit[10], contains a set of mapping software components that allows adding of maps to applications which can run on Windows 95, 98 and Windows NT 4 or higher. MapObjects consists of a major ActiveX control called Map control with a set of over forty-five ActiveX automation objects [Hartman, 1997]. These objects allow applications to provide most basic mapping functions such as zooming, panning, overlaying, labeling, and identifying/querying/updating attributes associated with selected features. When installed together with VB IDE on the same computer, the MapObjects components can be included in VB programs in the same way as other VB IDE built-in components are added. Alternatively, MapObjects can be associated with Microsoft Visual C^{++} and used in the same way as in VB IDE.

MapObjects was mainly used to develop mapping related components as mentioned above for *GeoPM*. It was used to program two types of components: ActiveX controls with a Map control that can be deployed as Internet downloadable applications (registered on the IICW client's computer when downloaded) and stand-alone applications with both Map and WebLink controls that are deployed on the project server computer and registered as IMS services. In developing the downloadable ActiveX

[10] At the time of thesis writing, ESRI has already released a new GIS toolkit based on new data model and built into its ArcGIS software.

controls with the Map controls, it was found that these controls do not work on computers other than the one used for the development – the controls cannot be rendered in web browsers. Investigations indicated the downloadable ActiveX controls, when downloaded to and registered on the client computers, lose one critical DLL file called "shape20.dll", although it has been included in the deployment packages. It was further confirmed through discussions with ESRI technical support staff that MapObjects was not originally designed for the purpose of developing Internet-based applications.

PL/SQL TM, Oracle's procedural extension to structured query language (SQL), is an advanced fourth-generation programming language (4GL) which offers features such as data encapsulation, overloading, collection types, exception handling, and information hiding [Portfolio, 1999]. Since the Oracle Database Server is used in the prototyped collaborative workspace, the use of PL/SQL allows seamless SQL access (tight integration) with the Oracle databases and server tools, portability, and security.

The development of PL/SQL procedures implemented in *GeoPM* was done within the Oracle WebDB environment, which allows management of all Oracle database objects including databases, tables, procedures and interfaces components. There are two types of PL/SQL procedures developed in the prototype: procedures providing form-based database application components (such as *GeoPM* discussion forum), and procedures providing web interfaces through which users can access project information stored in databases.

VBScript, JavaScript and Pearl scripting languages are used to develop necessary scripting programs and ASP pages that can be executed on both IICW servers and the client's computer to facilitate easy and efficient access to the collaborative workspace.

All project-related attribute information is stored in an Oracle database and managed by the Oracle server. The design of the project database focused on tables holding not only all collaboration attributes such as map file processing status, but also those attributes necessary for the workspace administration. The Oracle Database Server is expected to manage spatial data of the collaborative project in the future as well.

5.2 Prototype Implementation

The prototype collaborative workspace was developed to demonstrate three constructive aspects that helps improve the performance of GIS data production projects. These three aspects include: (1) work and process coordination; (2) project data file and document management; and (3) project issue-resolving through collaboration and communication. Emphasis has been placed on presenting and discussing implemented supporting functions, as well as problems and constraints associated with them.

The GUI of the IICW web clients presented is organized in three HTML frames, with some embedded ActiveX control interfaces displayed in the main frame. Figure 5-2 illustrates the main GUI of the IICW client interface, where the top frame displays hyperlinks to the major functional groups, the left frame displays an index map that allows linkage to all project artifacts related to each individual data file, and the right frame – the main frame – acts as the container to present activities performed within the workspace or HTML pages containing hyperlinks to further options (e.g., hyperlinks to GeoPM tools). The index map in the left frame provides two extra options other than panning and zooming in/out: *Identify* and *Hyperlink*. While the *Identify* option allows quick links to all predefined attributes of data files, such as delivery dates, the *Hyperlink* option provides linkages to other project artifacts related to the specified data file (click on the index grid associated data file) such as documents, reports, and workflow.

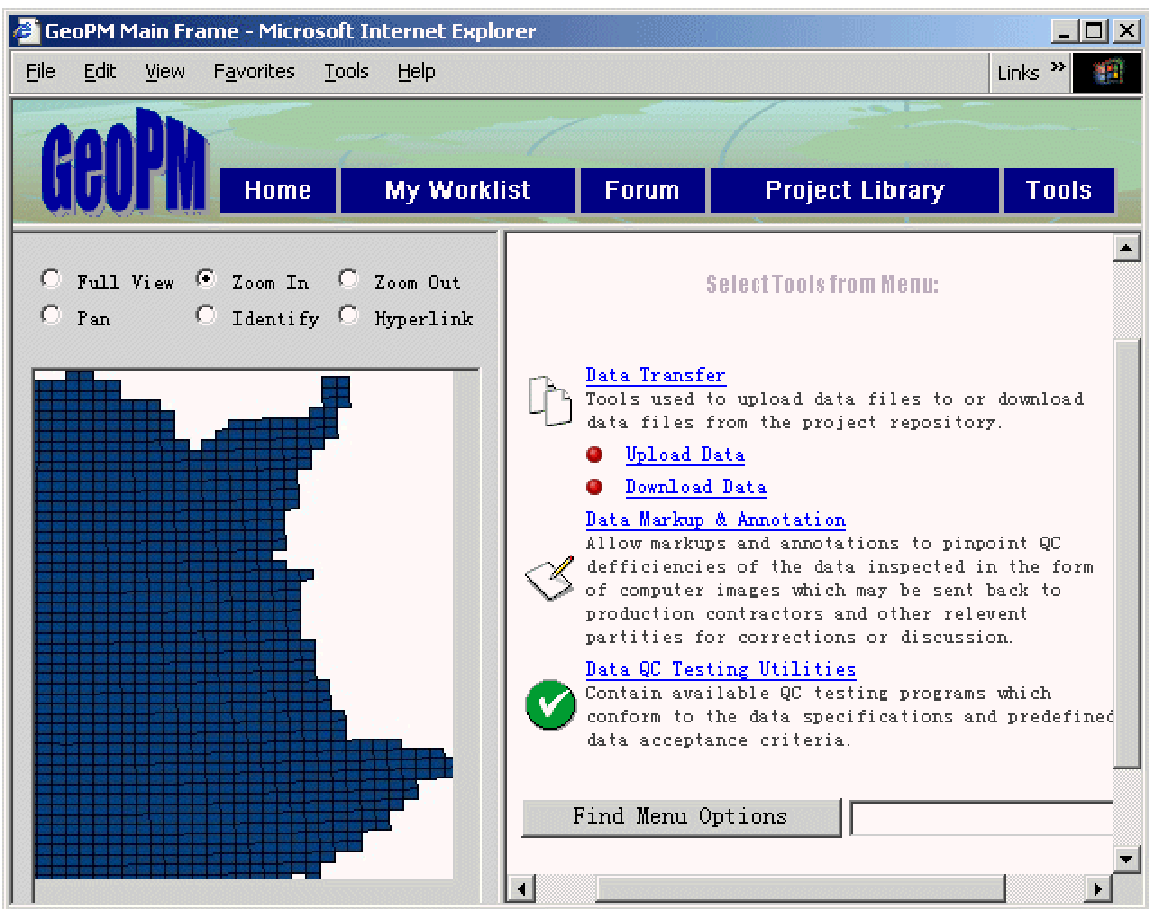


Figure 5-2 The main GUI of the prototyped collaborative workspace

Generally speaking, a web browser can visualize and display every file format (e.g., text, raster image, sound, movie, etc.) that has a MIME type, in some cases, with the help of the required plug-ins or helper applications. This prototype only involves simple formats including text, raster images, and compound text such as tables and forms. With ActiveX controls, only Microsoft Internet Explorer is used as discussed previously. While these presented GUI interfaces are briefly described in the following sections, the actual implemented functions that support these interfaces will be discussed in detail.

5.2.1 Progress Tracking and Control

This section is devoted to the descriptions of software implementation for supporting project progress monitoring and control, and status querying of individual data files in processing in the project.

5.2.1.1 Data File Status Querying

The GUI that allows *GeoPM* users to query the real-time status of working GIS data files is presented with two formats: textual- and graphics-based interfaces. Chapter 3 has already discussed these two different approaches in querying the processing status of data files. Each format allows users to specify query conditions that will be translated into basic SQL querying strings at the server side to query the project information databases to obtain required status information. The results are presented in HTML files displayed in web browsers in graphics format or tabular format, depending on the query GUI interface selected.

The “Status” view may vary depending on the query conditions specified in both GUIs, such as “display the status of all files inspected by certain contractor or contractors” or “display all files whose inspection status are of ‘Data Accepted’ and completion date is before April 1997”. In addition to viewing these results, *GeoPM* users also have options to output them in other different formats for later referencing purpose or sharing with other project participants. Currently, results can be stored in text-based formats (see Function 1) and graphics format (see Function 2).

Function 1: Textual Query of Data Processing Status Based on HTML Form

The form-based version of the interfaces for status querying was developed with Oracle WebDB tools. It provides several HTML form components in its interface, including buttons and pull-down menus that allow users to compose query conditions through specifying values of certain attributes associated with data files, e.g., data file identifier (a unique identification for each data file) and data status (see Figure 5-3). The query result is dynamically created on the server side in a tabular form and displayed in a separate HTML file.

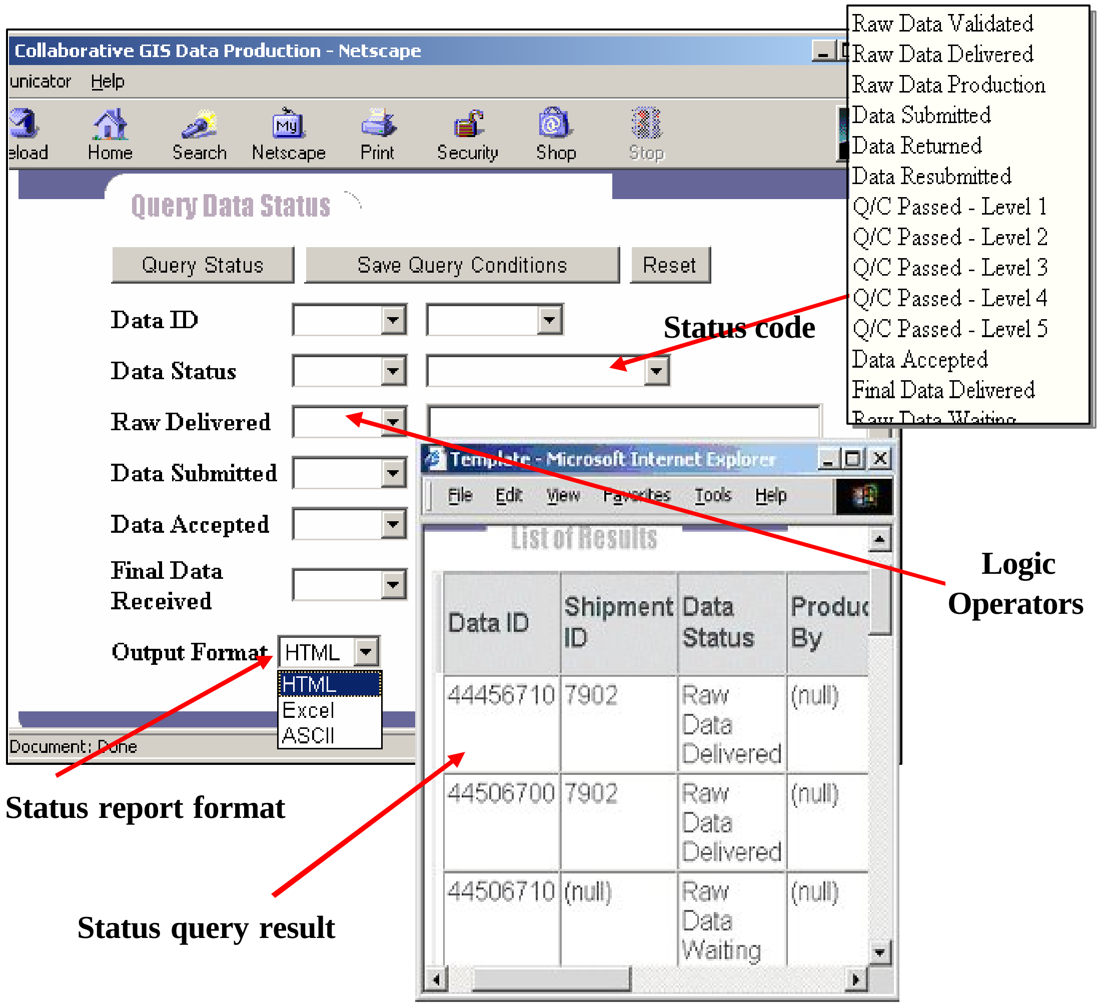


Figure 5-3 Text-based data status query

In this HTML form, not all attributes are included using pull-down menus for specifying query conditions. The current query interface contains six attributes of data files – *Data ID*, *Data Status*, and four attributes allowing queries based on some significant milestones in the overall project life cycle for each data file. The milestones include dates when raw data materials are delivered, when produced data is submitted, when submitted data is accepted, and when final data product is delivered. Users are allowed to specify values of any of these attributes and select logical operators for the query (the pull-down menus in the middle of the form allows most commonly used operators such as “+”, “-“, “=”, “>”, “>=”, “<”, “<=”, “like”, “not like”). If none of these attributes is specified (default query), the query goes for the status information for all data files in processing in the project. An example query that could be performed by this form is to find “all data files that have been submitted by production contractors but not yet accepted by the project manager”. In this case, data files may be in the process of data inspection or returning for further corrections.

If users decide to store the query results for later referencing, they can select to store the results in HTML, Microsoft spreadsheet (Excel), or pure ASCII formats. For HTML and ASCII format, the results will be displayed in a separate HTML page directly in the web browser. For Excel format, users will be prompted with a pop-up window to select if they want to save or open the result file in the same way the web browsers deal with file formats other than HTML.

Function 2: Graphical Query of Data Processing Status Based on Index Map

The graphics version of the interfaces, called *Status Map*, will use a grid map (e.g., index map) covering the project area to display status by using different colors. This is a kind of simulation of the actual paper-based status management approach found in real data production projects (see Appendix A for hardcopy status map examples). Each grid cell (or index grid) relates to one data file at a certain scale, as discussed in Chapter 3. Therefore, symbols or colours filled in each cell will indicate a certain status of the data file associated with that cell. When users first access the function, a coloured status map with a legend showing most updated status information will be rendered (see Figure 5-4).

The GUI, in addition to rendering the status map, presents several other capabilities in the displayed HTML page. Users can pan, zoom in/out, and draw the full extent of the map. While zooming in or out, users can directly click on zooming "bulbs", the panning function needs users to click on the status map to specify a new display center point. Further, users can request new displays of the status map by restricting "block", "status", or "shipment number" using pull-down menus and text input box. For example, status maps can be obtained only displaying status information of data files falling within certain blocks, or with the same shipment number. Users can also request a status map highlighting all data files having the same status.

For the purpose of future references to status maps at any specific date or time, users are allowed to save or plot hardcopies of status map as images. However, the plot capability is not implemented in this prototype. Other unimplemented capabilities, such as "*Identify*" and "*Hyperlink*", are designed to establish linkages to all information and documents relevant to specific data files.

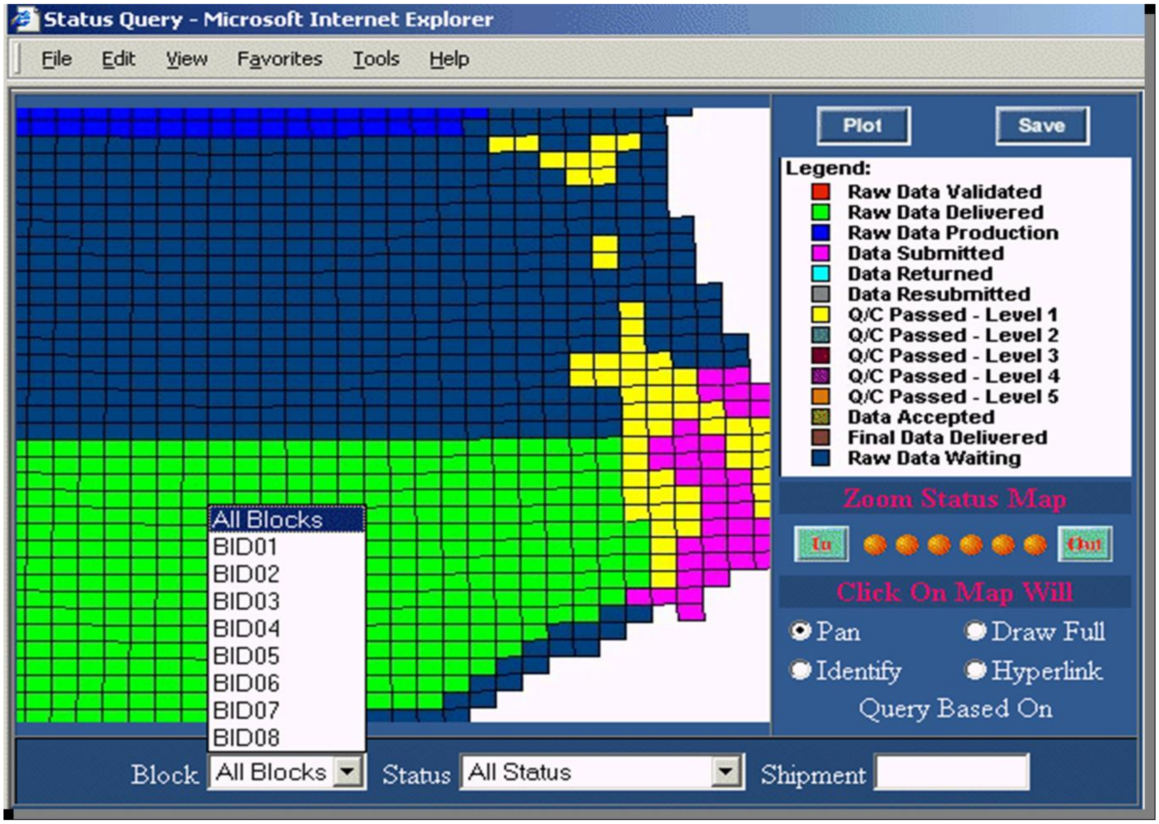


Figure 5-4 Graphics-based data status query

This graphics-based GUI for status querying is supported by an ESRI IMS service application at the server side. The service application is not only responsible for creating status map images based on user specified query conditions, but also for obtaining current status information from the project metadata database which is kept updated by project data production workflow execution as discussed in the next section. When a query is received by IMS, the service application connects to the project metadata database to acquire status data and then, based on the status data, a new status map is created by the Map control as the image embedded in the returning HTML page.

Function 3: Status Summary of All Data Files in Processing

In many cases, project management needs the overall picture of the current status of all data files in processing, representing as either tabular forms or Gantt charts. The example illustrated in Figure 5-5 presents one way to extract status information from the project metadata database and renders the status summary in Gantt charts: horizontal and vertical, based on the chart setup. Both web-based GUI and chart outputs are produced by the server-side procedures created from the Oracle WebDB basic objects – procedure building blocks.

The two Gantt charts illustrated in Figure 5-5 are created based on the "status" attribute of data files, where the base axis represents status values and the other axis represents actual counts of these values. Note that the orientation of base axis determines if the chart is horizontal or vertical. Both charts summarize the current total of data files that is with each status and present the maximum and minimum numbers among these totals (Note: the "zero" counts are not presented on the charts). For example, currently there are 564 data files in data production stage and only 70 data files submitted. Gantt charts can also be created based on other data file related attributes such as "Data Submitted Date" and "Final Delivery Date". In this case, the base axis will be date values with intervals in month or week. Since the Oracle procedures for GUIs of setting up charts and creating charts based on attributes other than "status" have to be predefined, this particular prototype only defines procedures for charts based on "status" attribute.

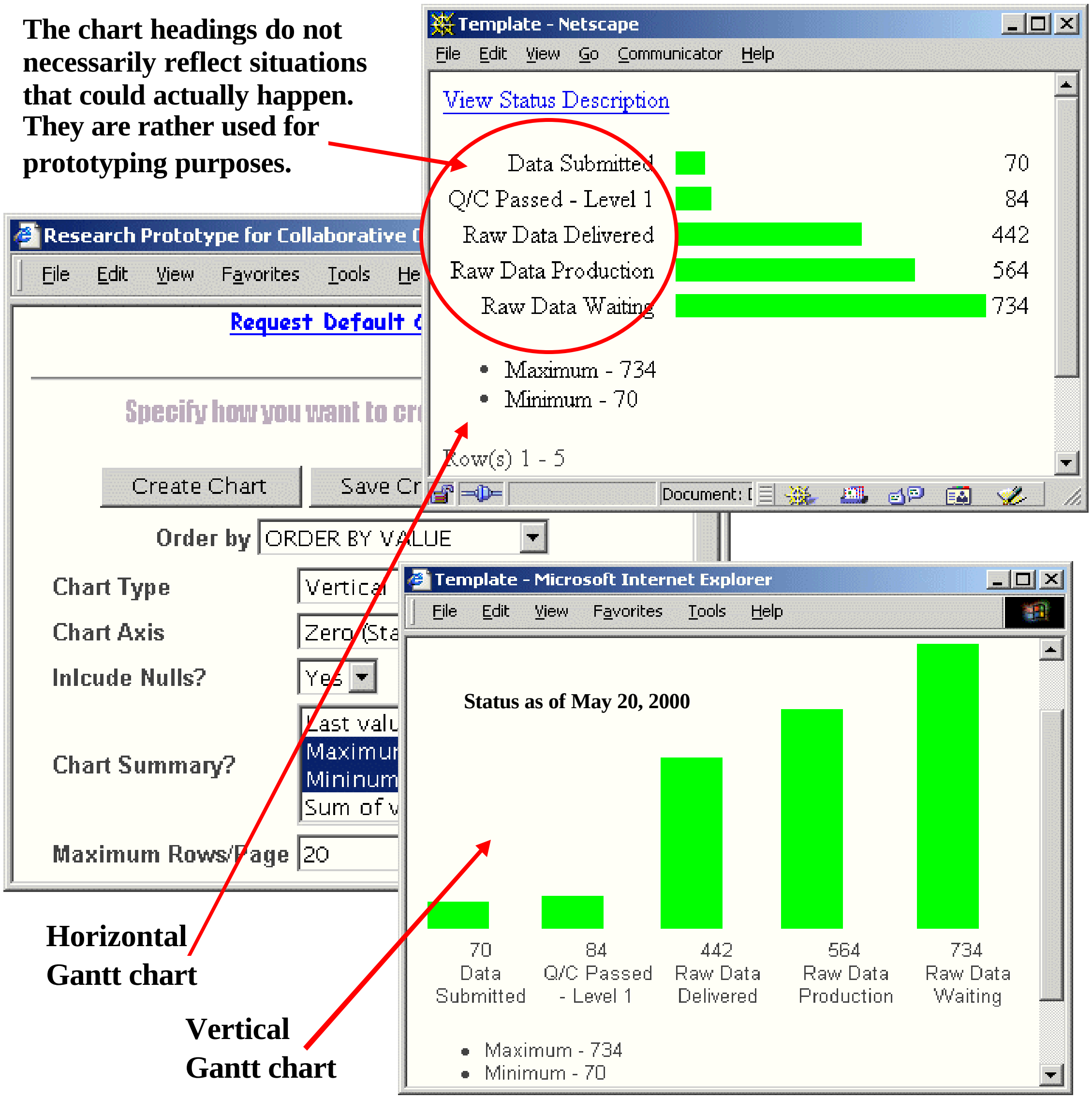
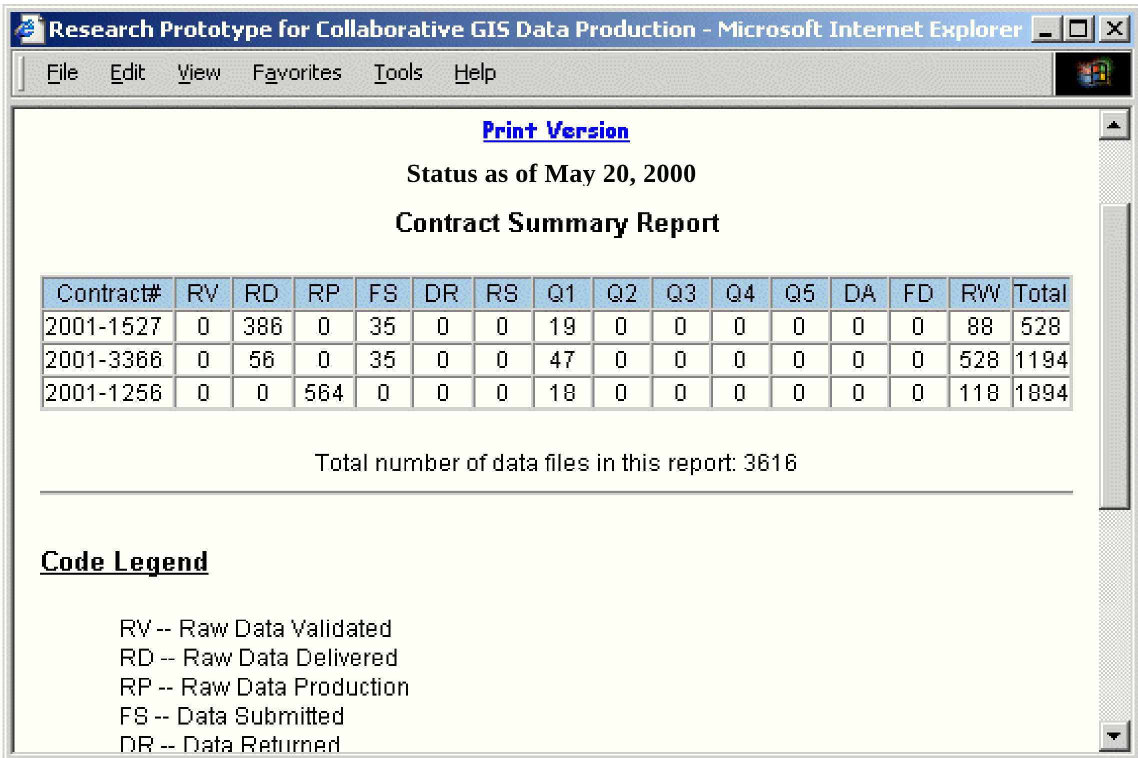


Figure 5-5 Status summary in Gantt charts

The prototype also provides a status summary in tabular form that is rendered in pure HTML pages. The summary can be presented based on contract numbers, shipment numbers, or blocks. Figure 5-6 illustrates a summary based on different Contract Numbers involved in the project (Note: Contract Numbers here are hypothetical). The summary table presents total numbers (counted based on different status) of data files under the same contract number on one row. To facilitate the user's understanding of summary table's column headings, a code legend is included which briefly explains the meaning of abbreviations used in headings.



Contract#	RV	RD	RP	FS	DR	RS	Q1	Q2	Q3	Q4	Q5	DA	FD	RW	Total
2001-1527	0	386	0	35	0	0	19	0	0	0	0	0	0	88	528
2001-3366	0	56	0	35	0	0	47	0	0	0	0	0	0	528	1194
2001-1256	0	0	564	0	0	0	18	0	0	0	0	0	0	118	1894

Total number of data files in this report: 3616

Code Legend

RV -- Raw Data Validated
RD -- Raw Data Delivered
RP -- Raw Data Production
FS -- Data Submitted
DR -- Data Returned

Figure 5-6 Tabular form of status summary

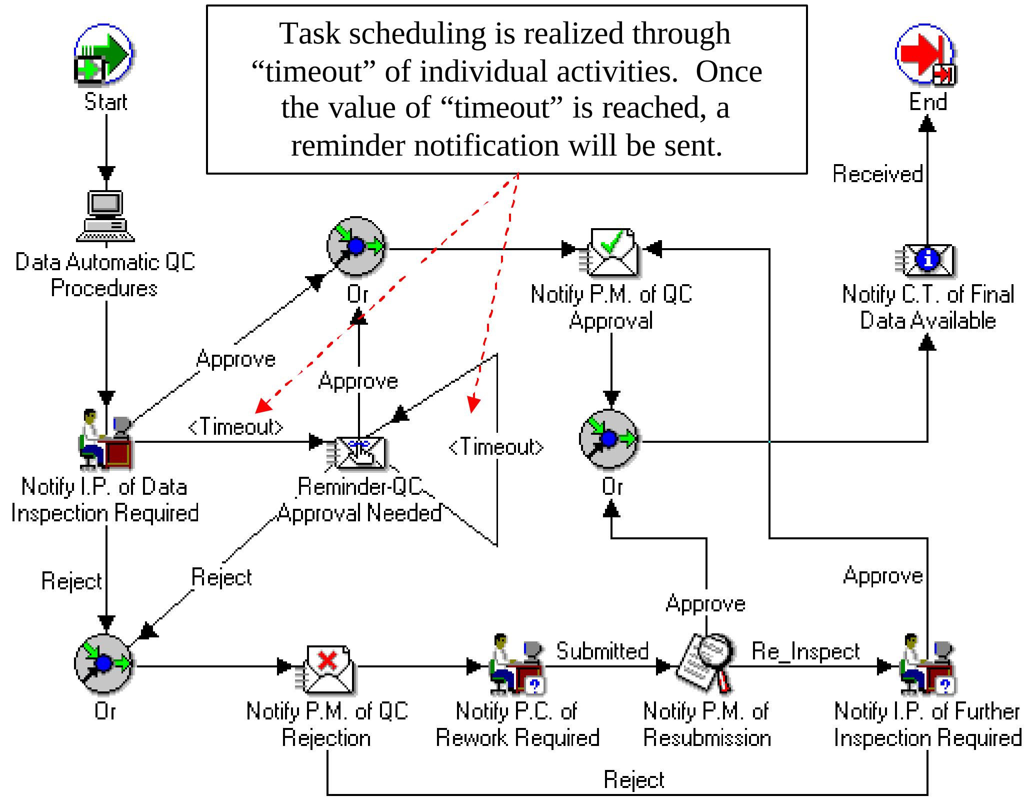
The summary results are generated on-the-fly from the project metadata database every time a request is sent from the IICW client and received by the *GeoPM* servers. Since the status information is updated primarily by the production workflow automatically and, in some cases, by the project manager manually, it ensures the summarized status information is maintained on a near real-time basis.

5.2.1.2 Project Progress Tracking and Control

"Progress tracking and control" capabilities in the prototype cover the tracking of data files along the production workflow and monitoring/controlling of GIS data processing work. Since the Oracle Workflow management system is adopted in the prototype development, the workflow processes discussed in Chapter 3 are described with the Oracle Workflow Builder[11] TM, entered into the Oracle Workflow system, and executed by the workflow engine. Figure 5-7 illustrates the data QC workflow described and used in the Oracle Workflow system, which contains major milestones involved in the data QC process. Many workflow instances are created and executed on an individual data file basis. Two example functions (see Functions 4 and 5) are discussed in this section.

---

[11] The Oracle Workflow Builder is the trade mark of the Oracle Corporation. This software package allows representation of workflows in the way the Oracle Workflow engine can understand.



C. T.: Client **I. P.:** QC Inspector **P. M.:** Project Manager **P. C.:** Production Contractor

Figure 5-7 Workflow process defined in the Oracle Workflow system

The workflow process illustrated in the above figure is for data quality control, during which various data inspection tasks are performed after produced datasets are submitted. The workflow starts with the submission of produced datasets and end with notifications to the client of available final data products. Once a workflow instance is started, a "Data Auto Check" activity is performed that executes some automatic QC testing procedures managed by an external functional program. Upon completion, the external program signals the workflow to continue and the QC inspector is notified of the data file for inspection with the results from the "Data Auto Check" activity. Based on the QC results, the workflow goes to notify the project manager of either the approved data file if it passes all QC testing or the rejected data file if it fails all or some of QC testing. If this file is rejected, the production contractor is notified that the rejected data file requires further corrections. Further, upon the resubmission of the corrected data file, the project manager determines if the QC inspector needs to be notified to inspect corrected data file.

Function 4: Management of Data Production Process

While managing project progress using workflow management systems is one of the primary objectives of this research, the process management function includes several sub functions that are used to control, monitor, and view current progress of workflow processes executing by the workflow engine. These sub functions are provided by the Oracle Workflow system with some supplementary database procedures developed in this prototype to accommodate needs connecting workflow capabilities to other workspace components such as obtaining project metadata within workflow operations.

The first example of sub functions is to check status of all workflow process instances underway. Figure 5-8 shows a web interface that lists all instances of the workflow process defined in Figure 5-7. The process instances listed in the tabular table contains process identification (*type*, *key* and *name*), instance key, execution status of process instances (*Complete*, *In Error*, and *Suspended*), and the beginning date of each process instance. In this particular process list screenshot, all visible process instances are in running condition so none of the above mentioned execution status is checked.

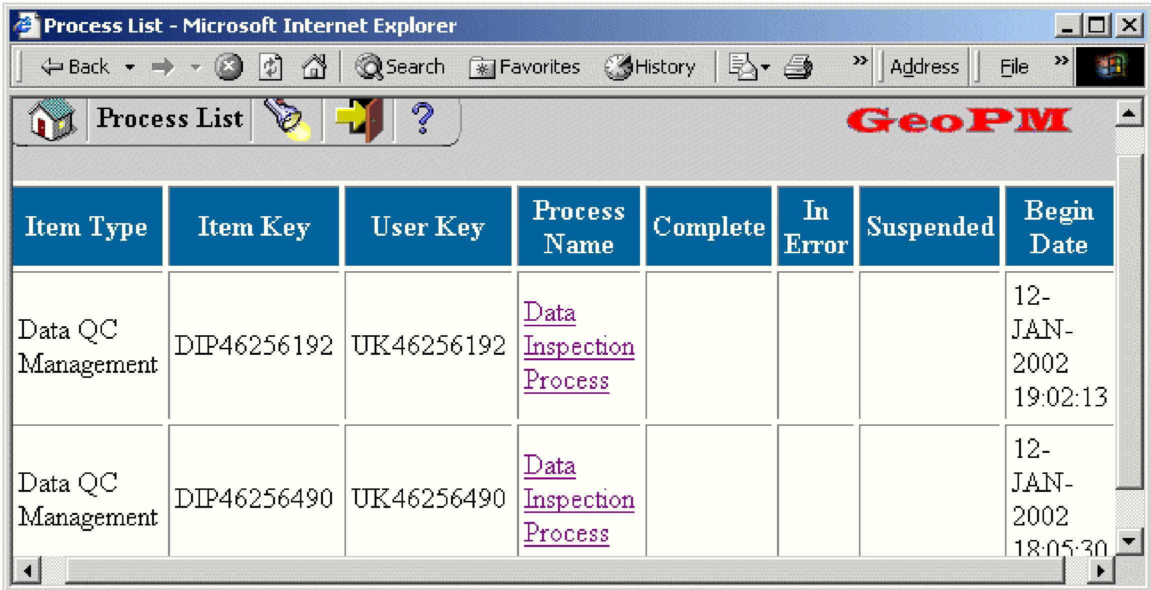


Figure 5-8 Monitoring execution of workflow process instances

Through the hyperlink under the column headed as "Process Name", detailed information about activities involved in the corresponding process instance can be obtained and rendered as a separate HTML page, which is titled as "Notification List" (see Figure 5-9). However, this notification list only presents activities that require user response or interaction, since the Oracle Workflow system is based on sending "notification" to realize interaction between human performers and workflow execution. Those activities executed automatically are presented on the process diagram (see example in Figure 5-9).

The process diagram is presented by a Java applet provided with the Oracle Workflow software. Downloaded and run within IICW client's browser, the applet dynamically tracks the progress of workflow execution on the process diagram by highlighting the activity that is currently active. In the screenshot presented in Figure 5-9, the highlighted activity is "Data Auto Check". In addition, this dynamic monitoring capability enables the authorized users to abort and suspend process as well as to expedite activities.

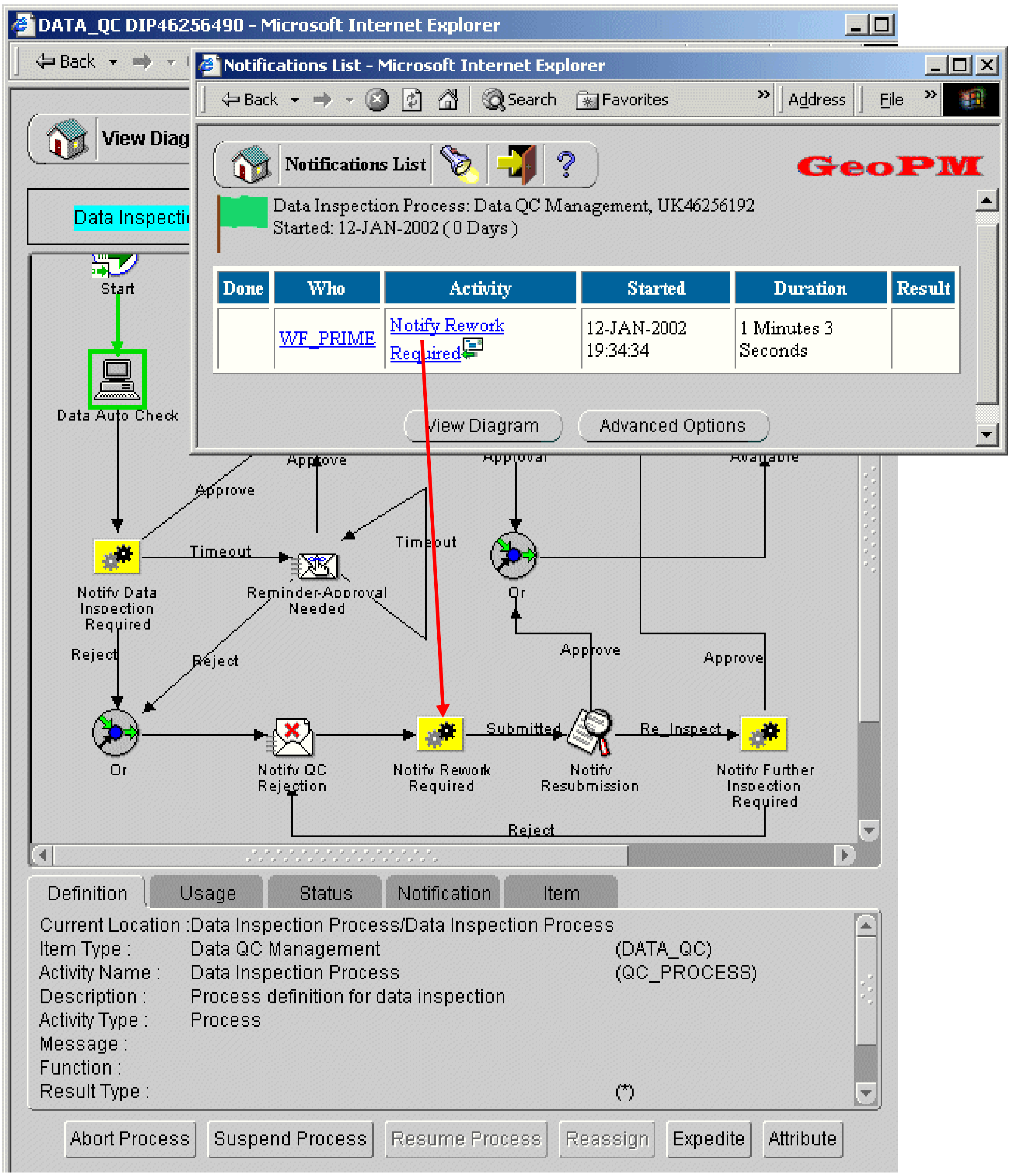


Figure 5-9 Viewing and controlling workflow execution details

All sub functions described above are mainly used for process management purposes, i.e., they are used by specially-authorized users who have been assigned the "project manager" role. The view of the notification list presented here is different from the one presented for all regular project participants, which is described in the following function.

Function 5: Processing Workflow Notification Activities

One useful feature of most workflow management systems is the worklist provided to individual workflow participants, where all tasks assigned to a specific workflow performer (specific user or workflow role) are listed. The interfaces presented in Figure 5-10 contain three functional GUIs: worklist, notification details, and task reassignment. The worklist interface presents all work a workflow performer has to do, with the due date of each task and the date the task was sent to the user. From this interface, the user can check a detailed description of each workflow task by simply clicking on the hyperlinked task subject.

The "Notification Details" interface describes the workflow task in detail, where the format of description is predefined as message templates through the Oracle workflow definition tool. In the particular case illustrated in Figure 5-10, the task is assigned to the WF_INSPECTOR – the role of QC inspector – to inspect the submitted data file. The interface has two major parts: a task description (based on message template); and a comments text area where the task performer can write any notes and/or comments on the task to be submitted or to be reassigned to other project participant.

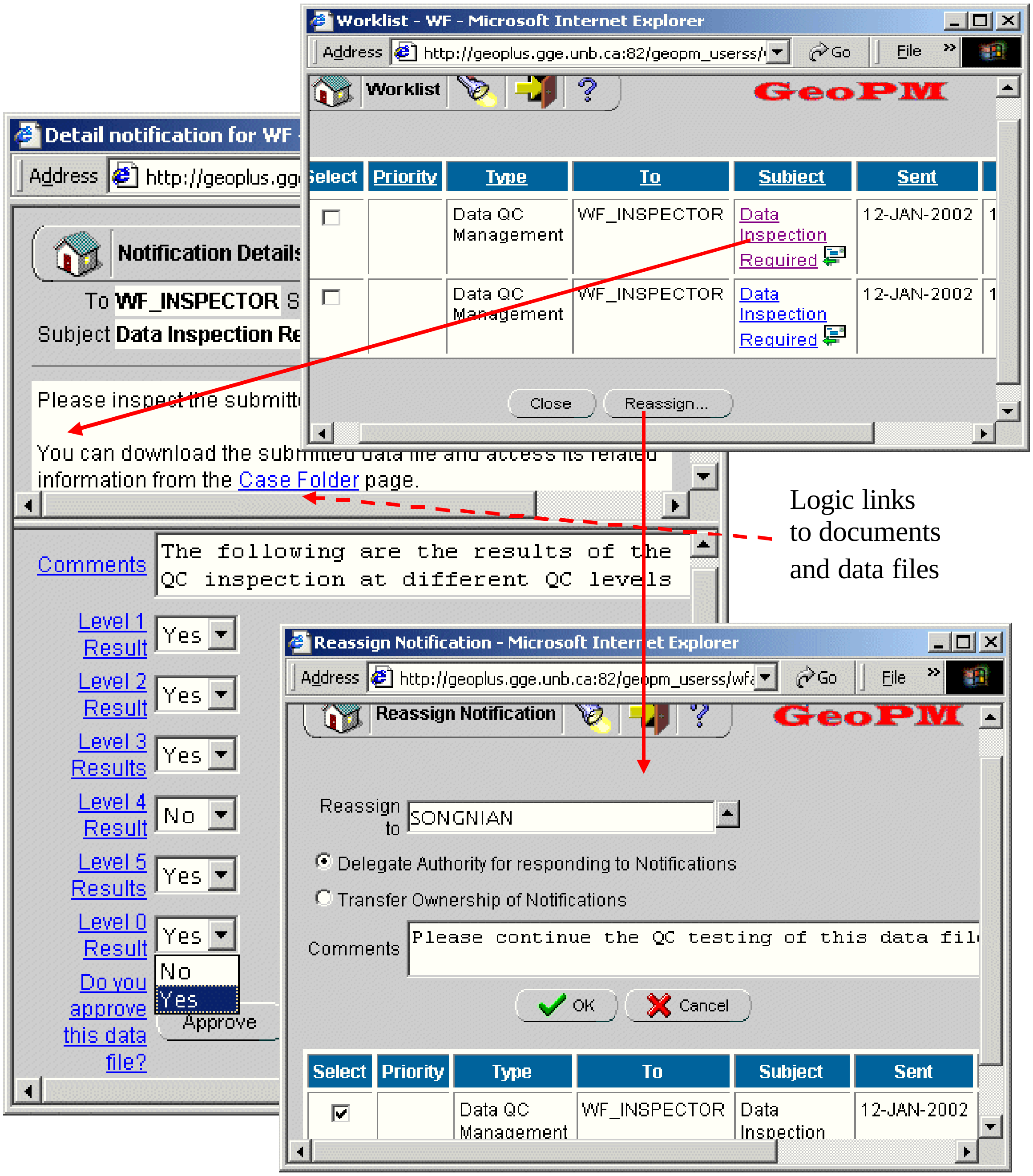


Figure 5-10 Handling notifications on the worklist

Different types of tasks have different formats of task description based on predefined message templates. The task description illustrated in Figure 5-10 notifies

WF_INSPECTOR of a newly-submitted data file to be inspected. The data file and related documents can be found and downloaded from the hyperlinked "Case Folder".

Ideally, a workflow system should support and/or integrate with a document management system so that both document and data files can be automatically attached to the notification messages. The Oracle Workflow claims to support "Open Text Livelink Intranet (8.0.2 or higher)" and "Oracle Internet Document" systems. Consulting with the Oracle technical support, there was no such "Oracle Internet Document" developed at the time of testing in Winter/Spring 2002[12]. There was also no existing license of "Open Text Livelink Intranet" at UNB at the time of this research. Therefore, a "Case Folder" concept was developed in this prototype development to address this defect, ensuring at least relevant documents and data files can be linked easily. The "Case Folder" does not physically contain any documents or data files. Rather, it is built as a logic folder that contains links to proper documents stored in the Oracle database and data files stored in a file directory. This "logic" case folder was included in all workflow notifications as a pointer to necessary workflow resources as discussed in Section 4.2.2.

There are two ways to reassign workflow tasks to other project participants through the integration of the Oracle workflow management system. One way is to use the Worklist interface where the tasks to be reassigned are checked (under "Select" column) and the

[12] The Oracle Internet Document was replaced by the Oracle Internet File System and the Oracle Workflow system did not reflect this change. Based on personal communications with people having the same problem and with technical support staff from the Oracle Corporation, it was possible to develop new interfaces for the Oracle Workflow system to interact with the Oracle Internet File system. However, there was no successful implementation reported at the time of the prototype development.

"Reassign" button pressed. Another way is to read the notification details of a specific task first and then decide if that task should be reassigned to someone else to perform. In both cases, once the reassignment is requested, a new interface is presented (the bottom window in Figure 5-10), which allows the current performer to specify a new task performer and comment on reasons or any other issues related to this reassignment.

All the interfaces described under this function are provided by the Oracle workflow with both HTML and Java Applets. However, a lot of work was done to implement these workflow functions based on the workflow model discussed in Chapter 4. In using the Oracle Workflow Builder to describe workflow processes, a number of notification and message templates were designed and populated in the Oracle Workflow system. PL/SQL procedures are developed to support extra features such as "Case Folder" implementation and post-processing of notification responses (e.g., updating status information based on the results from notification activities).

5.2.2 Project Data and Document Management

As discussed in Chapter 2, some of the most challenging aspects of the existing GIS data production projects involving in distributed project participants are related to efficient management of project documents (e.g., specifications, reports and predefined procedures) and GIS data files. The prototype implementation described in this section deals with these challenges, focusing on three aspects: (1) single-point access to all project-related documents; (2) process control of document updating, especially for technical specifications; and (3) centralized storage and check-in/checkout of data files. While data file submission, return and resubmission is controlled through the production workflow processes discussed in the previous section, the project document submission and updating is controlled through the related workflow process discussed in Chapter 4 and implemented in Oracle Workflow.

5.2.2.1 Management and Updating of Project Document

Project documents serve as containers for many types of project related information such as contracts, technical specifications, production procedures, reports, correspondence, and memos, as already discussed in Chapter 2 and Chapter 3. In this particular prototype, these documents are logically grouped into six categories: ***Contracts***, ***Specifications***, ***Production Procedures***, ***Official Correspondence***, ***Project Reports***, and ***Others***. Physically, all these documents are stored in one database table as files [13] except official correspondence which are stored in a separate table because their information can be structured. The table schemes are listed as follows:

Documents [DOC_ID, DOC_NAME, MIME_TYPE, DOC_SIZE, CREATOR, CONTENT, RECIPIENT, DOC_TYPE, VERSION]

Correspondences [ID, SUBJECT, TITLE, BODY, SENDER, RECIPIENT, SEND_DATE, RESPONSED_REQUIRED, THREAD_ID, DATA_ID]

While these documents may be submitted to the project database through different web-based interfaces (see Function 7 and 8), documents from Contacts, Specifications, and

[13] Oracle databases allow large object (LOB) data types (e.g., BLOB, CLOB, NCLOB and BFILE, see List of Abbreviations) enabling storage of large files [Portfolio, 1999].

Production Procedures categories will go through a submission/review workflow process (see Function 6). This is because most important document updating is required for these types of documents and the updated results have to be available to related project participants as soon as possible to avoid unnecessary misunderstandings and waste of time and resources as discussed in Chapter 2.

Function 6: Control of Document Submission and Updating Processes

The capability of controlling the document submission and updating process, implemented using the Oracle Workflow system (see Figure 5-11), is transparent to project participants. Depending on the type of documents and the procedure chosen by the project manager with respect to that submitted document, project participants can get three kinds of notifications in their aforementioned worklist.

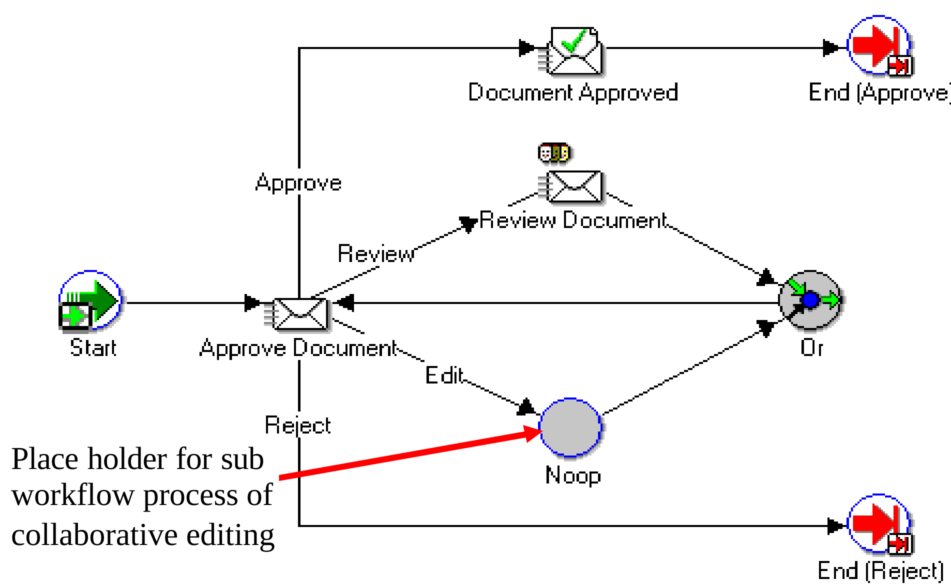


Figure 5-11 Document submission/review workflow process

The first kind is a pure "for your information" notice that notifies them of newly available or updated documents, for which participants do not have to do anything except check out the documents as needed. The second kind requires participants to take some action such as commenting on or approving documents. The last kind involves collaborative editing of draft documentation.

In the case of collaborative editing (asynchronously), a sub workflow process (see Figure 5-11) will be initiated by the project manager and proper workflow instance created, where many rounds of editing and reviewing procedures are realized. Each notified participant contributes to the editing by checking out, revising, and resubmitting documents until a final version is reached. However, this sub-workflow is not implemented due to the lack of integrated document management system in the Oracle Workflow as discussed previously. The documents are not really attached to the notification messages directly and versions (or "time maps") of the document in editing cannot be easily controlled. Therefore, they have to be accessed through hyperlinks embedded in notification messages.

Function 7: Access to Project Technical Library

This function allows project participants a single-point access (called *Project Technical Library*) to all relevant project documents through a HTML page that contains hyperlinks to the above six categories, as illustrated in Figure 5-12. From each hyperlink, users can access documents within the corresponding category in a separate HTML page that lists them with appropriate document description information. Currently, the prototype does not impose any restrictions on the formats of stored documents. The formats obtained through accessing this technical library will depend on the formats stored in the database. Based on a discussion with potential users of this kind of workspace, there may be a need for a utility tool that allows conversion of documents between some file formats such as Microsoft Word and WordPerfect since not all project participants have required “helper” applications. This capability, however, is not implemented in this prototype.

The main interface for accessing the project technical library also includes the access point for publishing documents to the project database, which initiates the “Publish Document” form as illustrated in the right window shown in Figure 5-12. To publish a document, users have to specify some required information such as a descriptive name, document category, and required/suggested project participants who should read the document. If the document category of “Contracts”, “Specifications”, or “Production Procedures” is specified, a document submission/review workflow process (see Function 6) instance will be created and started to control.

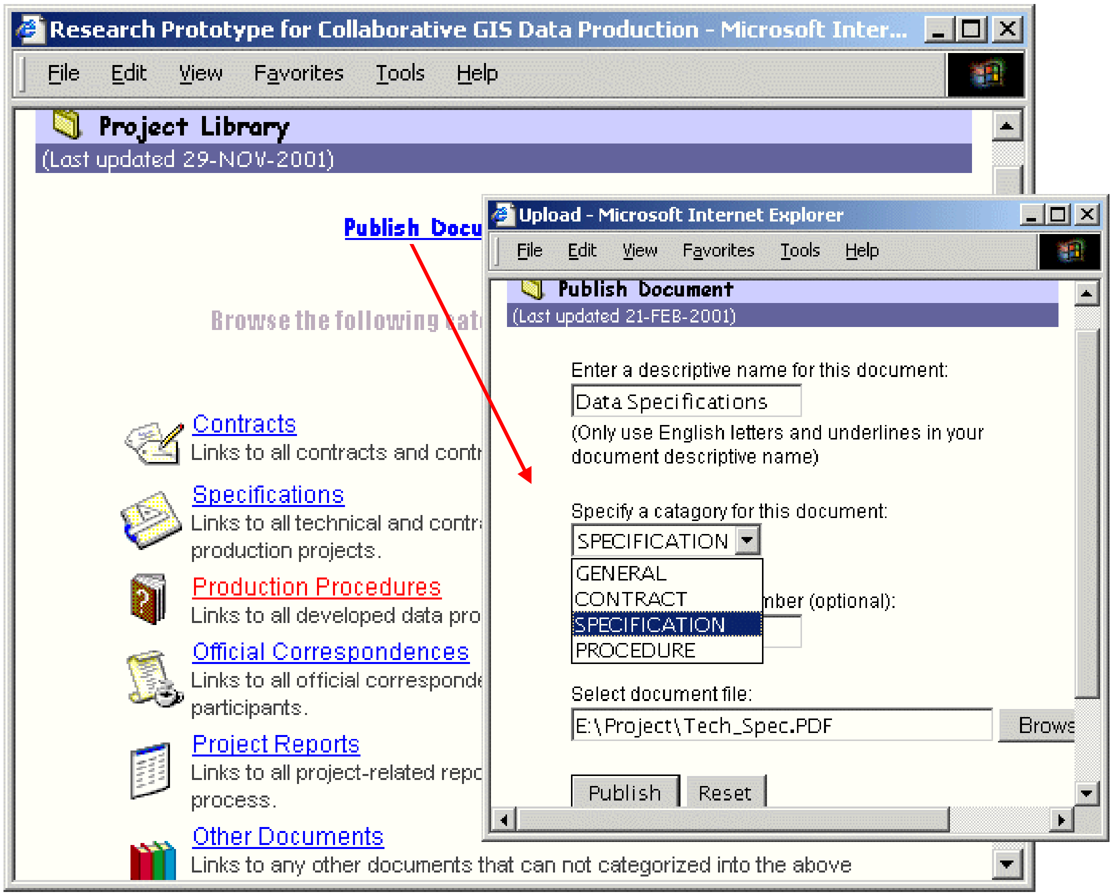


Figure 5-12 Single-point access to the project technical library

While the *Menu* object of the Oracle WebDB is used to implement the major interface for accessing the technical library, the document publishing interface and the presentation of detailed lists of documents within each category are realized through its *Form* object since documents are read from or stored into the Oracle database.

Function 8: Management of Project Official Communications

Official project communications are realized in the prototype in two forms: project reporting (see Figure 5-13) and official correspondence (see Figure 5-14) as discussed in Chapter 3. While official correspondence deals with those official memos, notes, and announcements that cannot be sent through project workflow executions as workflow notifications, the project reporting capability allows participants to file and submit formal project reports using predefined reporting formats (templates).

The interface for project reporting is presented with a HTML form that takes from reporters such report attributes as the report's descriptive name, recipients, and shipment number if needed, as well as an ActiveX control interface (see the window on the right-button corner of Figure 5-13) that allows reporters to fill in necessary report contents in a structured way. The HTML form implements two ways to submit a report. One way is to browse a report file reporters have already created using other word processing tools and saved on the local computer. Another way is to allow reporters to activate the ActiveX control interface to compose the report in a structured way "on-the-fly". Once all required contents are filled in, the ActiveX control component will create a report file that is temporarily stored on the local computer and specify the appropriate path in the HTML form so that the report file can be submitted.

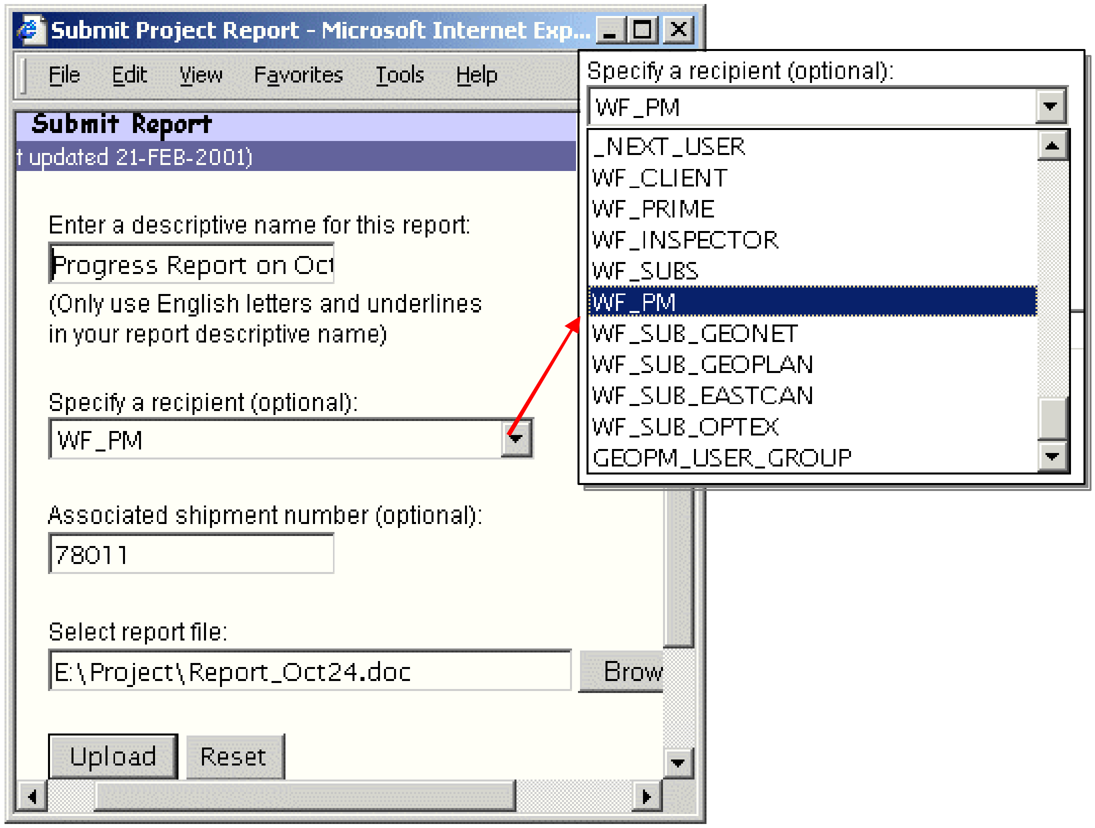


Figure 5-13 Interface for project reporting

As discussed in Function 6, after the report is submitted, a submission/review workflow instance will be created to take care of the review and distribution activities of the report. As such, it will be ensured that the report reaches the appropriate project participants, especially the project manager.

For sending official correspondence, the HTML form illustrated in Figure 5-14 is implemented. To construct an official correspondence requires from users such mandatory inputs as subject, recipients, title of correspondence, and message body. Optionally, users can specify the specific data file identification to which the correspondence relates, and indicate whether or not a reply is required. If not specified, the default data file identification is empty and the default reply request is “Yes”. The structured contents of the correspondence will be directly uploaded and saved in the database table described in the beginning of this section.

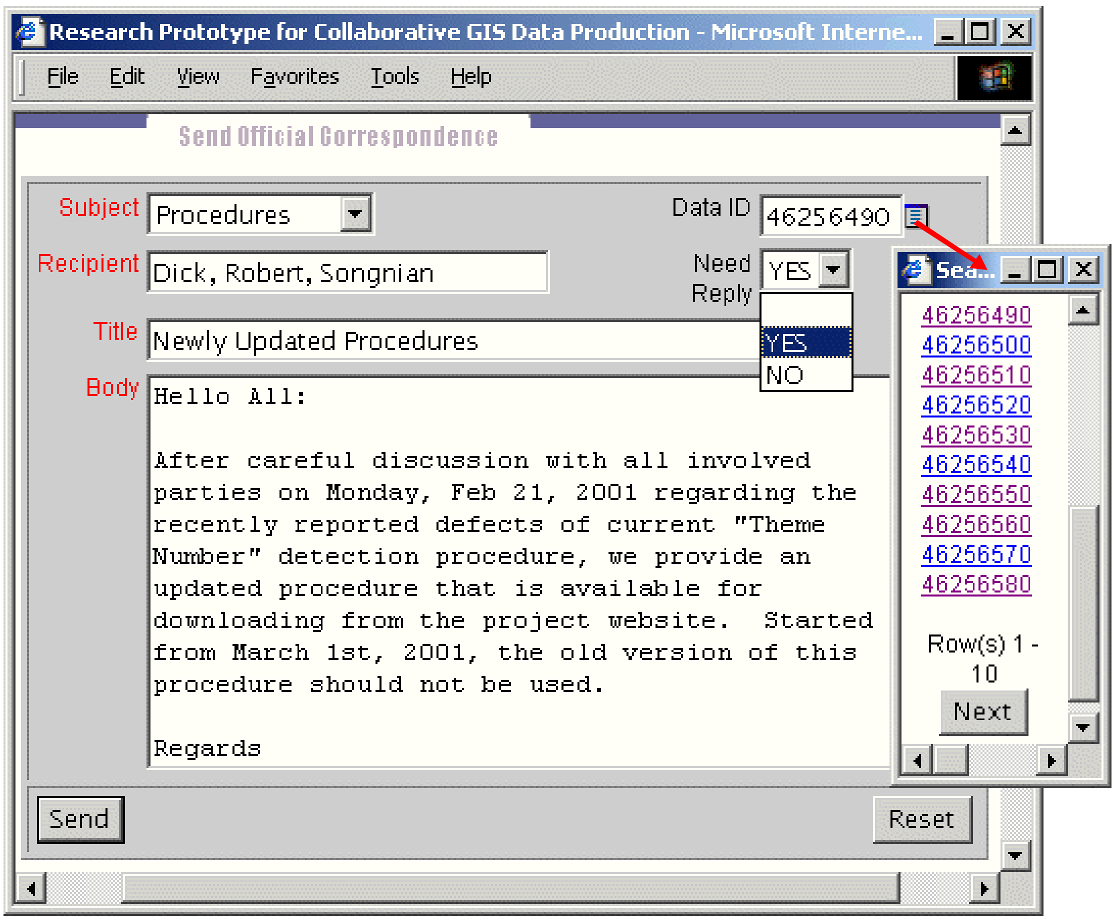


Figure 5-14 Interface for sending official correspondence

HTML forms implemented for this function are realized through the Oracle WebDB *Form* object and the ActiveX control was developed using Visual Basic 6.0. The same

ActiveX control is also implemented as a separate utility tool (as discussed later) to allow project participants to compose structured reports that will not be submitted right away.

5.2.2.2 Management of Data Files and Data Transferring

All data files – including files containing raw data materials, submitted data files and final data files which pass data QC inspection – are stored on the project server (in this case, the server named TYR). The files are organized into a structured file directory that takes into considerations the multiple submissions. Although it would be better to have data stored into a DBMS such as the Oracle Database system in terms of controlling versions at the data entity level, there are difficulties in doing so in this thesis due to the format of sample datasets used and amount of work to be done to develop programs to implement this idea.

Different projects have different ways of organizing data files. In this particular research, it is assumed that all files associated with one map sheet required by CARIS data format are packed into one ZIP file. The project data file directory organizes data files in a hierarchical way. Data files related to each dataset (map sheet) are stored under one folder which subsequently has three sub folders: raw data folder, submitted data folder and final data folder. This enables control of different versions of the same data files.

Based on this discussion, two functions described as follows examine capabilities implemented in the prototype to check-in and checkout data files from the project file directory. Through theses functions, data files can be check-in to or checkout from the project data file directory.

Function 9: Check-in of Data Files (Data Uploading)

The data file uploading interface implemented using ActiveX control as illustrated in Figure 5-15 is capable of checking in multiple data files at the same time into the project file directory. The supporting ActiveX control communicates with a server-side application program that manages uploaded data files.

The ActiveX control interface allows users to select multiple data files from local file directories and add them into an "uploading list". Users can modify the list by removing any unwanted data files. Once the final "uploading list" is determined, users can specify upload type including "*Raw Delivery*", "*Submission*", and "*Re-Submission*" and optionally enter comments or memos for the uploading. The upload type is an important parameter used by the server-side application to decide how the uploaded data files should be handled. For example, if the upload type is "Raw Data Submission", the server-side application will start data production workflow process instances for every uploaded data file.

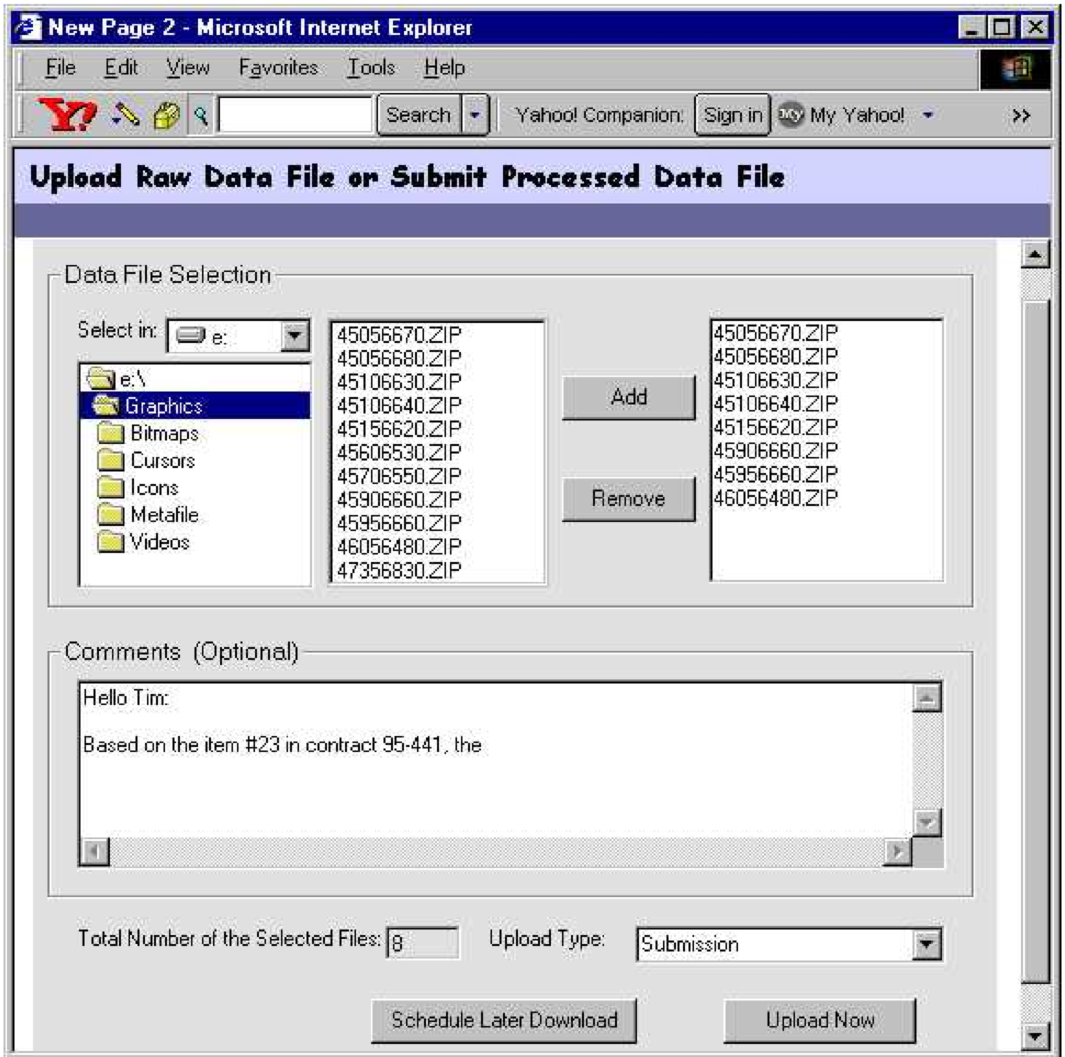


Figure 5-15 Multiple data file uploading (check-in)

The only limitation of using this ActiveX control to check in multiple data files is its reliance on the Internet data transfer performance. When too many data files are selected and uploaded at the same time, the uploading session may be interrupted due to the network performance. This will be further discussed in Section 5.3 and the next chapter.

Function 10: Checkout of Data Files (Data Downloading)

As opposed to Function 9, this ActiveX control supported function allows project participants to download ("check out") multiple data files at the same time from the project data file directory. This function is also limited by the Internet performance in term of the number of data files that can be checked out at the same time. The interface allows creation of a "downloading list" and specifying the local folder where files to be stored.

As discussed in the previous chapters, GIS data files tend to be large in size compared with other types of data files. Downloading multiple GIS data files will sometimes take a long time to finish. It is necessary to take some measures to ensure that files selected to download are the ones actually desired. One of these measures implemented is a data file "Preview" feature which allows users to preview file content before selection. In this case, a map snapshot is created on the server and displayed (the "Data Preview" window in Figure 5-16). The other feature enhancing the usability of this function is the capability to monitor downloading progress and to cancel the downloading process if needed. The progress bar illustrated does not appear at the outset of the function interface. It shows up by replacing the panel holding local folder selection GUI components after the "Download Now" button is pressed.

This capability of monitoring downloading progress is also implemented in the data uploading function described earlier, allowing users to monitor or cancel data uploading process. The interface style is the same as discussed here.

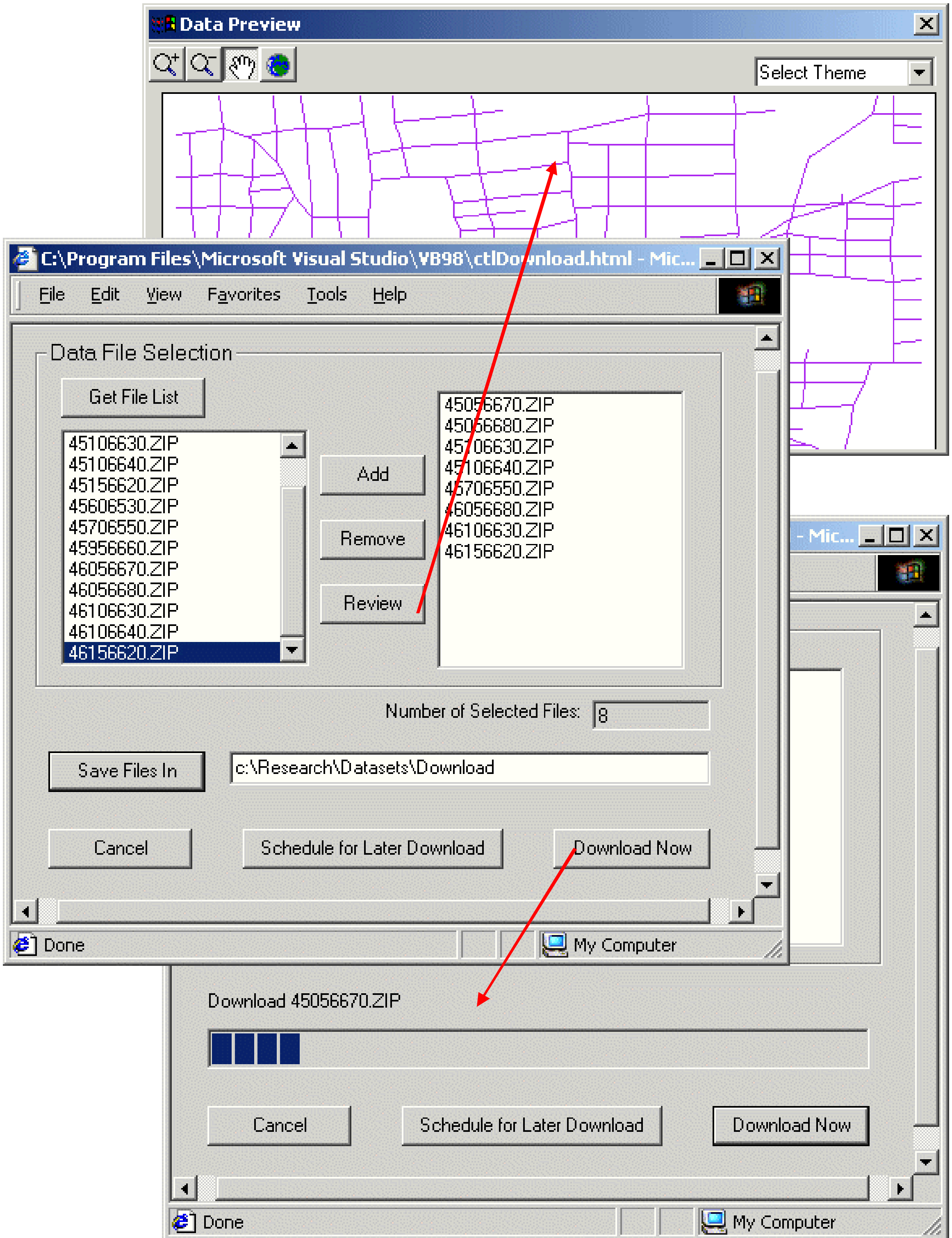


Figure 5-16 Multiple data file downloading (checkout)

There are some special capabilities called “Schedule for Later Upload” in Function 9 and “Schedule for Later Download” in Function 10. These capabilities are especially useful when the Internet file transfer performance is of concern. In this case, transferring of multiple data files may be scheduled at “non-peak” hour, e.g., during later evening or early morning if the transfer is between project participants located in the same time zone. These features, however, were designed but not implemented in this prototype.

5.2.3 Project Collaboration and Communications

Chapter 3 described many functional requirements supporting collaborations among project participants and communications of production issues and problems – both in synchronous and asynchronous modes. This prototype focuses on implementing functions that support asynchronous communications and synchronous collaborations that require less network bandwidth. As discussed in Chapter 2 and Chapter 3, the rationale for doing so is that real-time collaboration among widely distributed project participants is still restricted by the facts that: (1) participants may be located in different time zones; and (2) many small geomatics production firms are still using relatively low speed “Dial-Up” telephone lines for their Internet connections as discussed in Chapter 2. This section describes two examples implemented in the prototype.

Function 11: Discussion of Production Issues through Project Discussion Forum

This function, using the Oracle WebDB form object, implements a threaded discussion forum that allows project participants to informally communicate issues related to project management and production activities. Although web forums are already ubiquitous, the implemented project forum allows structured messages to be directly stored in and retrieved from the same project database.

The function presents three HTML interfaces: an HTML form for posting messages (see Figure 5-17), an HTML page listing all posted messages in a threaded way, and an HTML form for displaying details of a specific posted message and posting reply to that message (see Figure 5-18). To post a message to the forum, similar to many Internet forums, posters have to specify the subject (e.g., data, specifications, procedures, and general issues in this implementation), title and body of the message. The messages posted to the forum are stored in the Oracle database table with a table scheme defined as follows:

Postings [MESSAGEID, SUBJECT, TITLE, POSTER, POSTDATE, MESSAGE, THREADID, ISREPLIED]

Project participants can then browse the posted messages and select any of them to view its message. On the same HTML page where the message can be viewed, an HTML form is included to allow replies to that specific message, as illustrated in Figure 5-18. For both message posting and message replying, after the message is successfully received by the server, the confirmation information is presented

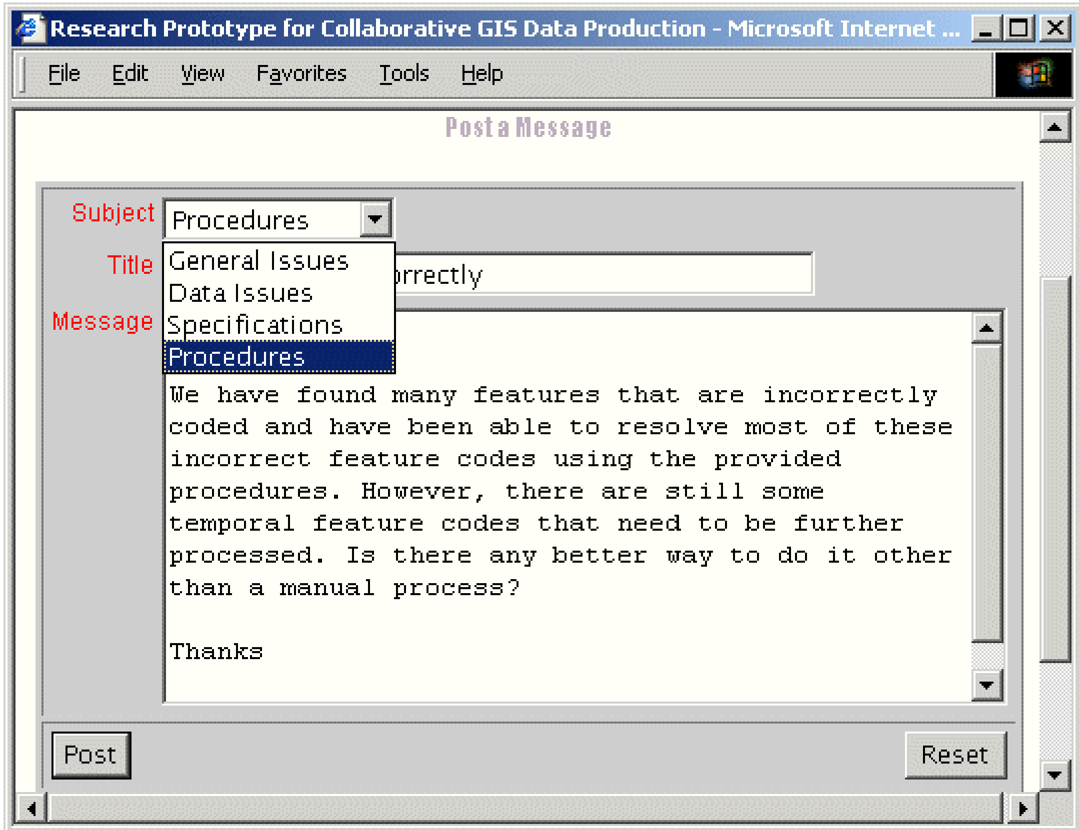


Figure 5-17 Posting a message in the project forum

The project forum is mainly used for informal discussions among project participants. The discussion can be on any issues involved in the project life cycle. In many cases, a project may require more than one forum to better organize this type of informal communications for building part of project memory. However, for demonstration purposes, this prototype only implements one instance of the project forum. With the Oracle WebDB, multiple instances of the forum can be easily duplicated, even by authorized users.

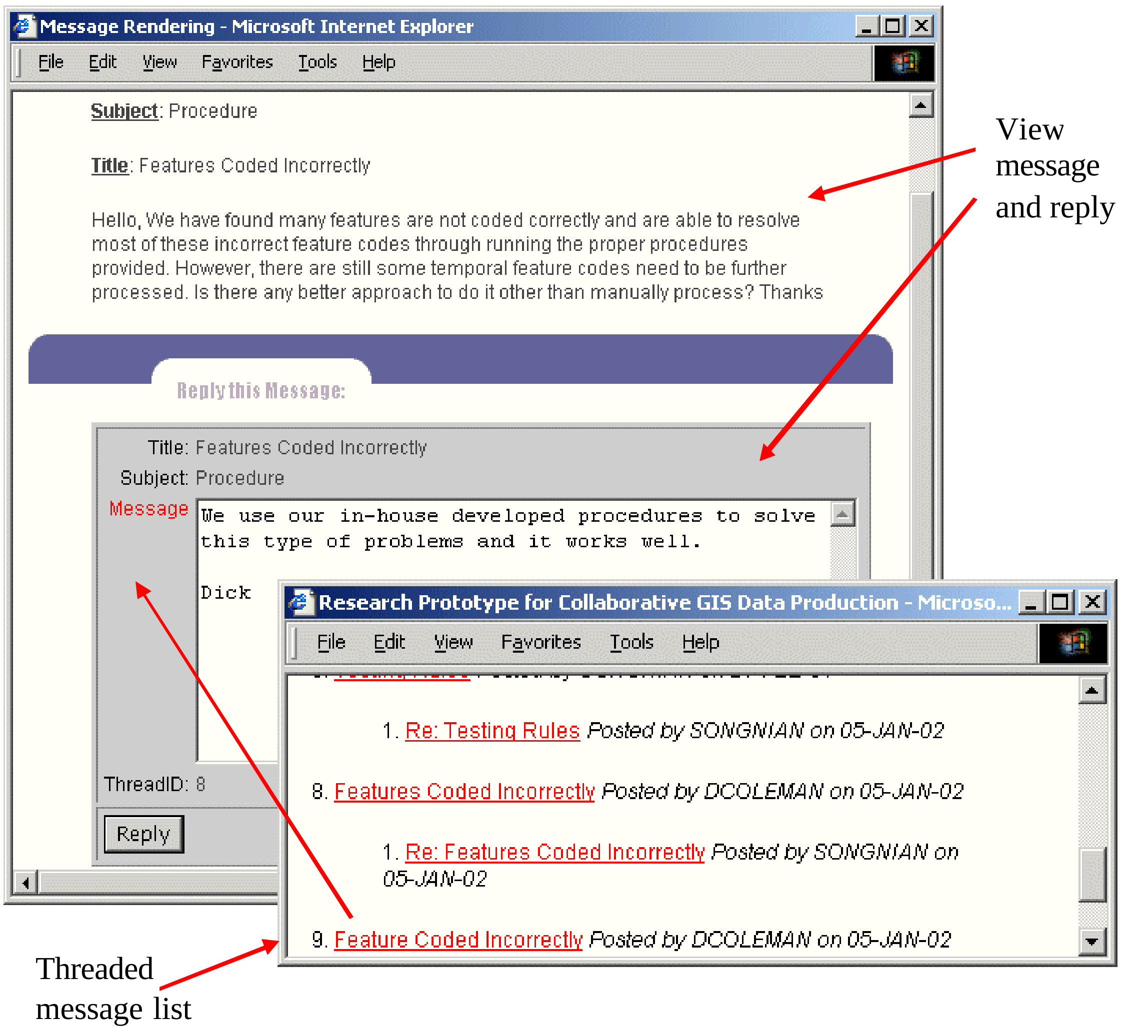


Figure 5-18 Viewing forum and replying to a message

Function 12: Collaborative Data Viewing and Markup

The concept of electronic "whiteboarding" of GIS data file discussed in Chapter 3 implies three aspects: (1) simultaneous viewing and manipulating data files; electronic "markup" of data entities requiring further work during "shared" sessions; and (3) collaborative editing of data files during "shared" sessions.

In this thesis research, the priority is given to simultaneous viewing of data files by two or more users at different locations and electronic "markup" of data entities individually due to the reason described at the beginning of this section. Therefore, only these two capabilities are implemented in the prototype. They are combined into the same ActiveX control and presented with its interface to users, as illustrated in Figure 5-19. However, the overall design of the ActiveX control considers other capabilities described above for future development and implementation.

There are some basic manipulation capabilities to support simultaneous viewing of data, including panning, zooming in and out, and zooming to full extent. The data files viewed can be opened either from local computers or from the project server. The ActiveX control opens up data files in its native format – in this particular implementation – ESRI shape format. While the datasets available for the research are all in CARIS format, they have to be converted into shape file format to be accessed by this function, which reflects a combined effect of both software and data limitations discussed in Section 1.5.

The electronic "markup" of data entities currently uses two symbols for graphic markups and short text annotation for necessary comments, i.e. the ActiveX control includes tools that allow users to add these graphic symbols and annotation text over or beside selected data entities. The two symbols are yellow question marks in graphic form and red circles. Yellow question marks mean the marked entities may have problems and need to be rechecked, while red circles mean the marked entities have errors and must be corrected.

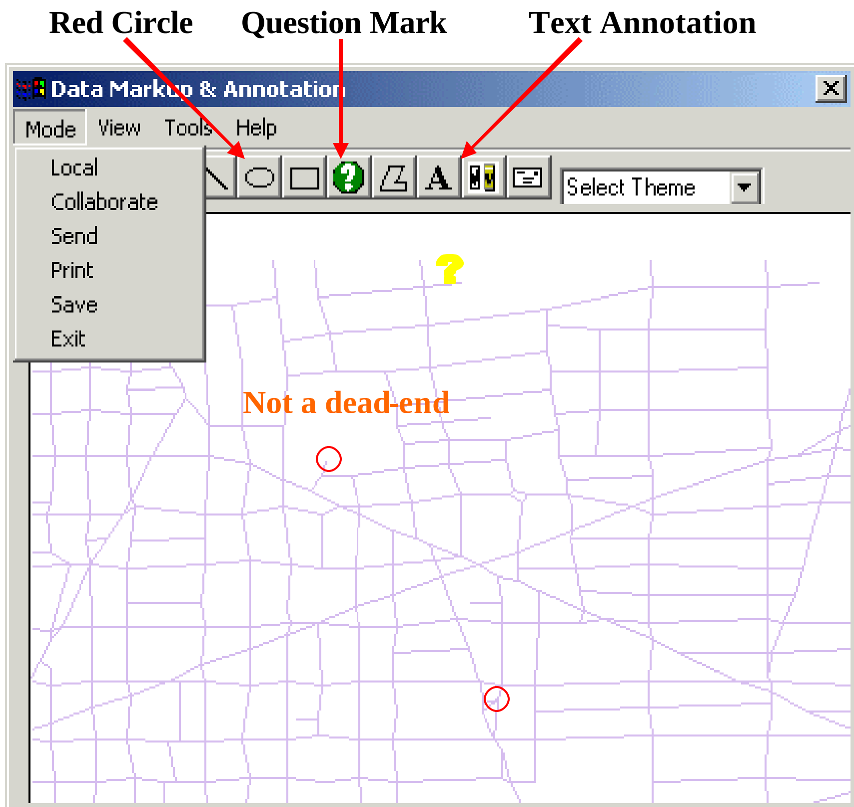


Figure 5-19 ActiveX control interface for electronic data viewing and markup

5.3 Performance Testing

This section discusses testing and analysis considerations of the prototype performance, of which the results contribute to the verification of the collaborative workspace model and the overall evaluation of the system. The original plan of this research was to test the prototype in two steps as described in Chapter 4. However, the actual testing was only done under controlled laboratory conditions due to the unavailability of appropriate GIS data production projects at the time of this thesis work.

The performance testing is based on the prototype collaborative workspace and implemented workflow process in the context of GIS data quality control. Two aspects of the performance are investigated here, i.e. data production process performance and system performance related to improving production performance. While the process performance tests concern sets of performance metrics that can be used to evaluate the improvement of the overall QC process, the system performance tests only focus on obtaining the estimation of the actual Internet performance in transferring data files.

5.3.1 Process Performance

The process performance in this research was compared to a typical paper-based manual QC inspection process (non-CSCW supported process). It is important to note that the comparison here is between processes only and not technologies, since the supporting technologies in this context are only tools to facilitate process gains. The comparison is generally conducted to determine three categories of benefits as discussed in Section 2.3.2: *improved outcomes*, *individual and group gains*, and *efficiency*. The first two groups of benefits were difficult to measure in this thesis due to the lack of data from real world projects. They will however be further discussed in Chapter 6.

For the purposes of this research, the "efficiency" is expressed in terms of a number of performance metrics including duration, elapsed time and resources required for every isolated workflow task. The resources are measured as capitalized costs of involved process workers, consumed materials, and supporting hardware and software shares. The duration and elapsed time are measured in days (or hours) and cost is estimated as dollar values. To make the comparison easy, all measurements of these performance metrics are converted to the equivalent amount required by processing the same number of data files.

The compared workflow tasks are well isolated from the implemented QC workflow as described in Section 5.2.1.1. Since the Oracle Workflow uses a notification mechanism to execute its workflows, most of the workflow tasks are sent to the workflow performers through notifications and are completed “off-line” (i.e. not controlled by the workflow), except those automatically executed tasks. For example, when the QC inspector is notified of some data files to be inspected, the inspector will check out these data files and inspect them off-line in the same way the paper-based QC process does. In this sense, there is no need to compare the performance metrics for actual inspection work. With this in mind, the process performance has been tested against those tasks that are directly supported by the collaborative workspace. Table 5-2 lists those measurable tasks in the implemented QC process.

The process performance also concerns other factors such as the average number of resubmissions, which affects the estimation of the overall performance. In addition, the improvement on project-related communications such as sharing documents and informing specification updates is measured by time, i.e. how much faster compared to a paper-based system.

Table 5-2 Measurable tasks

Tasks	Description
Data file packing and shipping	- Packing data files for submission, returning, resubmission, and final delivery - Shipment delivery via couriers
Shipment receiving and checking	- Unpacking delivered shipment, stamping and data loading
Running Automatic QC programs	- Execute all automatic QC programs against data files
Routine and project reporting	- Weekly or monthly progress reporting - QC report - Final project reporting
Corresponding in process	- Notifying data file receiving, QC approval and data file returning - Reminding approaching deadlines

5.3.2 Performance of Transferring Data Files Electronically

Generally speaking, for the CSCW-supported data QC processes, the less time each software tool spends on performing QC workflow activities and project management tasks, the more time left within the project schedule for more time-consuming tasks requiring human interaction. As discussed in Chapter 2, transferring data materials between project participants through several rounds of submission and resubmission may often delay final project deliveries. The proposed approach uses the Internet as the vehicle to transfer digital data files. To obtain an ***estimated measure*** of the time required to transfer data files to test the process performance, an uncontrolled approach was used to determine the approximate performance of transferring data files over the Internet.

In order to perform this test, a VB program was developed using the VB built-in control – Microsoft Internet Transfer Control 6.0, which allows tracking of the time when the control successfully sent the FTP request and the time when the request has completed and all data has been received (see the Appendix C for the source code). The testing program is used only for downloading data files from the project file directory to the client computer, which involved "reading" file from the server and "writing" the same file on the local computer.

The tests involved thirty data files and were run at four locations in Fredericton, Ottawa and Toronto. Three of these testing locations have an operating environment common to the normal computing/network environments of universities/government agencies. One of them has a cable-based high-speed Internet connection provided by a commercial Internet Service Provider (ISP). Each test downloaded 30 data files in one FTP session in sequence. However, the time required for every data file was recorded. As such, the purpose of FTP performance testing was to obtain estimated indications for: (1) time required transferring vector and raster GIS data files with various sizes; and (2) feasibility of transferring multiple data files in one FTP session.

While this section describes the basic considerations and testing scenarios used in the performance testing, the results and the analysis of these results is presented in the next chapter.

5.4 Summary

The collaboration model presented in Chapter 4 provides a useful framework for developing the research prototype, especially in selecting the software components and/or systems and defining workflow processes in the Oracle Workflow system. Although, the Oracle Workflow system does not currently support direct import of WPDL-based workflow models into its native workflow definitions, the model described with WPDL (see Appendix B) can be easily converted into the Oracle Workflow process definition since they both support the concept of assigning only one action to each activity.

The prototype discussed above represents a partial implementation of the concepts presented earlier in this thesis. The focus was on demonstration of ideas rather than rigorous software development. Therefore, some inconsistencies may exist between different described functional components. Some capabilities are discussed within the most appropriate functional components in order to present their relevance to the prototype. However, they were not actually implemented given the time constraints of this research project. It must be clear that implementing the complete functional workspace is a major undertaking requiring much further investigations and developments. As such, this research only implemented some strategically selected functional requirements based on the requirements specifications described in Chapter 3 to test the hypothesis (see Table 5-3 for a comparative list).

The biggest hurdle encountered in the prototype development was to select an appropriate workflow software system. Although the results of comparative studies on various workflow systems are not described in the thesis, the work actually completed was very time consuming. The most difficult task lies in the assessment of appropriateness given the unavailability of product documents, difficulties in evaluating software and, mostly, the high cost of the systems. Other problems encountered relate to the lack of Internet support of ESRI MapObjects and the lack of document management system for the Oracle Workflow system.

Table 5-3 Functional Implementation

Implemented Functions	Requirements Specified	Implemented Function Descriptions
Function 1	R2 – (c)	Textual query of data file status & summary
Function 2	R2 – (c)	Graphical query of data file status, status map
Function 3	R2 – (c), R2 – (c)	Status summary, report creation and charts
Function 4	R4 – (b), R4 – (c)	Workflow process control and management
Function 5	R4 – (a), R4 – (c)	Task notification, handling, and reassignment
Function 6	R2 – (a)	Document submission and updating control
Function 7	R2 – (a), R2 – (b)	Project document publishing and access
Function 8	R2 – (c), R3 – (b)	Online reporting and correspondence
Function 9	R1 – (b)	Data file uploading and invoking of workflow
Function 10	R1 – (a)	Data file downloading and data preview
Function 11	R3 – (e)	Project discussion forum
Function 12	R3 – (a)	Collaborative data view and markup

Chapter 6 RESULTS AND ANALYSIS

This chapter presents the results from the limited performance testing and some intangible benefits the proposed collaborative workspace approach could bring to distributed GIS data production project management and operations. While the testing results are summarized and analyzed in the first section in terms of time and cost savings, the potential of enhancing service and data quality levels are discussed in the second section. The chapter then finishes with the section discussing some intangible benefits that may be obtained through the proposed approach.

6.1 Time Savings and Cost Factors

Time and cost savings are analyzed comparing with the selected time and cost factors from SNB Topological Structuring of the Digital Topographic Database project for creating an Enhanced Topographic Base (ETB) covering the whole New Brunswick [Doucette, 1998; Castonguay, 1999; Roberts, 1999]. The project, hereafter called ETB'96, was finished in 1996 and handled 1890 data files corresponding to the same number of 1:10000 data windows. These data files were delivered in 14 shipments, in which the numbers of files are different with an average of 135 (see Table 6-1 for a list of shipments). The size of files ranges from a few hundred kilobytes to less than 2 megabytes. Five production contractors were involved. A separate company was acting as both project manager and QC inspector.

Table 6-1 ETB'96 contract shipment summary

SHIPMENT	DATE	FILES IN SHIPMENT
1	13 – Dec – 1995	55
2	10 – Jan – 1996	105
3	24 – Jan – 1996	142
4	7 – Feb – 1996	175
5	21 – Feb – 1996	176
6	6 – Mar – 1996	162
7	20 – Mar – 1996	175
8	3 – Apr – 1996	142
9	17 – Apr – 1996	150
10	1 – May – 1996	139
11	15 – May – 1996	161
12	29 – May – 1996	153
13	12 – Jun – 1996	109
14	26 – Jun – 1996	46
Total Files		**1890**

During the comparison analysis, the cost factors consider both implementation cost of *GeoPM* and actual reduced cost in project management and operations while the time savings reflect the increased speed of individual workflow tasks that contributes to the speed of the overall QC process.

6.1.1 Time Savings

Factors considered in determining time savings include the performance of transferring GIS data files over the existing Internet infrastructure, reduction of duration and elapsed time in selected QC process activities that could be performed through collaborative workspace support, and the time required for project reporting. The output from the FTP performance is used in determining QC process improvements.

Performance of Transferring Data File over the Internet

Sixteen uncontrolled tests were performed to download thirty selected data files stored on the prototype server (TYR) from two different locations[14]. Among them, however, only six tests completed the whole FTP session. The other tests were not completed due to either unexpected network interruptions or FTP error such as "time out" or "connection reset". Table 6-2 presents the results of eighteen selected data files from the six completed tests (see Appendix D for complete testing results). Numbers in bold are the maximum duration required to download the same data file among six tests and numbers in italic bold are the minimum duration required.

Among the 18 selected data files, each of the five ZIP files has the size of 1.0 MB which is close to the average size of CARIS data files (containing vector data) used in the research. The results indicate that the transfer of each data file could be done in less than a minute to the testing sites. A reality check of a recent development of using the Internet to facilitate data file submission also verifies this result (see Survey results in Appendix D). The other 13 data files contain raster images, ranging from 20 to 128 MB in size. Depending on the actual size of those raster image data files, the time required to download them varies. While for data files with a size around 20 MB, 73 percent of file

[14] The two locations are Ottawa and Toronto in Ontario, Canada.

downloads were completed within 5 minutes, the time required to download files with a size around 50 MB were mostly longer than 7 minutes (71%).

Table 6-2 Duration time of transferring data files with different sizes

File Name	Size (MB)	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Average (KB/Sec)
45606530.ZIP	1.0	***0:11***	0:17	0:12	**0:40**	**0:40**	0:20	7.14
45706550.ZIP	1.0	0:15	0:15	0:14	**0:47**	0:39	***0:11***	7.09
45956660.ZIP	1.0	0:11	0:15	0:14	**0:35**	0:28	***0:08***	9.00
46056480.ZIP	1.0	***0:12***	***0:12***	0:16	**0:34**	0:33	0:15	8.20
46406470.tf	128	26:07	39:28	26:35	**59:44**	52:38	***23:13***	9.36
46406480.tf	128	***21:04***	41:43	22:51	**58:01**	49:34	19:58	9.56
46506460.tf	128	20:11	32:18	22:00	**55:35**	49:01	***19:57***	10.81
47356830.ZIP	1.0	**0:18**	0:12	0:11	***0:10***	0:13	***0:10***	13.51
5pointofgis.tif	47	6:34	**10:55**	6:36	8:40	14:38	***6:23***	14.57
Buctdune.tif	128	20:23	28:04	22:47	32:54	**42:49**	***17:09***	13.00
caris_test.cpt	22	***3:13***	4:44	3:19	**10:29**	8:31	3:21	10.91
fredprop.e00	24	***3:34***	5:48	3:36	8:03	**8:18**	3:49	12.07
image007.tif	53	8:29	***4:20***	7:50	16:59	**17:24**	7:45	14.07
image1.cdr	50	7:34	10:34	***6:56***	9:46	**19:09**	7:02	13.66
imageII_07.tif	49	7:07	10:05	6:23	10:01	**11:50**	***6:10***	15.83
property.e00	23	***3:05***	3:50	3:10	3:51	**6:11**	3:25	16.29
redriver.pix	22	***2:58***	3:44	3:04	3:21	**6:09**	3:03	16.43
thomits_30.tif	23	3:1	3:27	5:00	3:26	**6:20**	***2:59***	15.98

Figure 6-1 presents the summarized information extracted from those testing results related to four raster image files that are 128 MB in size. The chart displays minimum, maximum, and average time used to download these four data files to two testing sites.

The average time used to download one 128 MB file to the testing sites connecting to university and government networks is approximately less than 50 minutes and even the maximum required time is still under one hour. This results in an average downloading speed of about 4.3 MB per minute. However, for the ISP-based testing site, the maximum required time to download one 128 MB data file range from 3 to 4 hours.

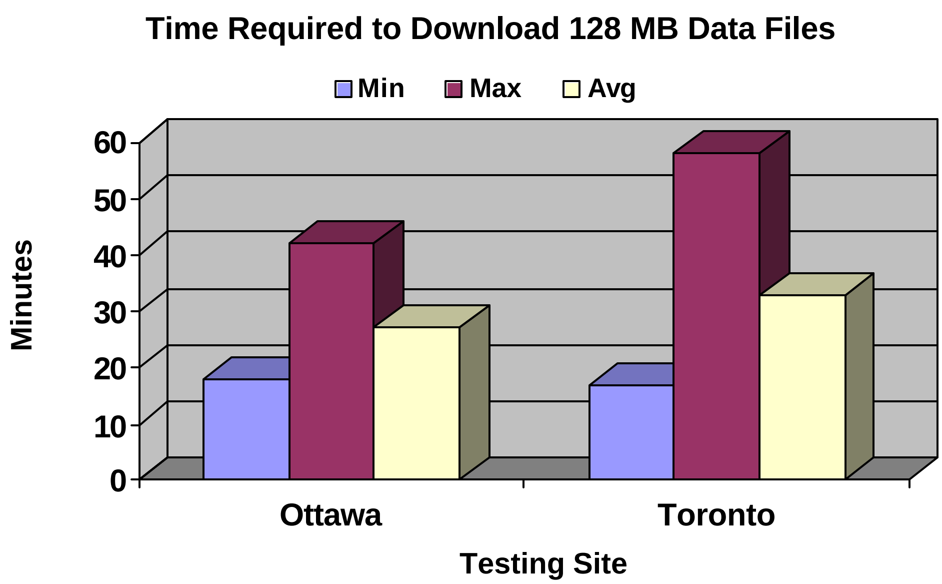


Figure 6-1 Time required to download 128 MB data files

With the proceeding interpretations of results, the following conclusions can be drawn and will be used in the process performance analysis described in the next section:

1. For the projects dealing with vector-based data, data files may be transferred collectively in one session provided the number of data files contained in one "shipment" is not too large because vector-based data files are normally small in size.

As in ETB'96 project, the average size of data files is around 1.0 MB and the average number of data files in one shipment is 135 (see Table 6-1). This would sum the size of each "shipment" to 135 MB, which will require approximately one hour transferring one "shipment" of data files.

2. Image data files are normally ranging from several to a hundred megabytes. While the performance of transferring individual image files may be acceptable compared to traditional approaches of shipping data files, transferring multiple data files in one session is still problematic due to unexpected Internet traffic and reliability. With this in mind, data files should transferred on individual basis to ensure the transmitting sessions are as short as possible. However, this would require the reengineering of the existing data file shipping process where data files are usually packed and shipped in shipment containing many data files.

3. Although the ISP-served connections may well serve the need of transferring small size data files in a timely fashion, the performance of using ISP-served Internet connections, based on either "dial-up" or cable modem, in transferring large size data files involved in GIS data production projects is still difficult to project and justify. Even for high-speed cable modem connections, downloading large size files (with 128 MB) would take as long as over 4 hours. The "dial-up" connections are certainly not sufficient for supporting transfer of data files.

Increased Speed of Executing QC Workflow Process

The total time required to complete QC inspection process for any shipment depends on many factors. Excluding other factors such as delay or process suspending due to unsolved QC problems (e.g., dispute on required corrections), the time periods spent on data files shipping and actual inspection work contribute the most to the overall time elapsed (see Chapter 2 for related discussions). During the overall QC process, data files are shipped between production contactors and QC inspector or between QC inspector and the client as project deliveries. In addition, during the process of inspecting data files at various levels, both manual and automatic QC procedures are performed. With the support of the prototype collaborative workspace, time savings are obtained from two aspects: reduced time for transferring data files and time saved in executing those automatic QC procedures by implementing them as workflow applications.

As discussed in data file FTP performance, the average estimated time for transferring a 128 MB data file over the Internet is about 50 minutes, excluding the case where an ISP-based Internet connection is used. While the typical shipping time using courier services in ETB'96 was at least 24 hours between different cities with an estimated elapsed time of 48 hours, this indicates a potential time saving of almost two days for each shipment. Considering the fairly large number of shipments involved in ETB'96 project[15], the total time savings are very significant.

[15] ETB'96 had 14 shipments in total and each shipment may have been shipped 2 to 4 times, or more depending on the QC results.

Time savings can also result from the execution of those batch QC procedures. With collaborative workspace, all batch QC procedures (often programmed as executable or batch files) may be attached to workflow activities as workflow applications. As such, the workflow engine automatically invokes these executables and results will be stored in separate reporting files. This saves a lot time for the QC inspector to run those programs individually and compile the testing results. In ETB'96, there were approximately 20 QC batch programs and the total time required to manually run batch process on one shipment would be around 10 hours (see the Survey results in Appendix D).

Table 6-3 provides a summary of the estimated times for the tasks related to the above two aspects based on one ETB'96 shipment containing 135 data files on average. The workflow tasks are listed in the same sequence as defined in workflow process discussed in 4.2.2 and implemented in the Oracle Workflow (see the related diagram in 5.2.1). Therefore, the workflow tasks selected do not include all the tasks involved in the overall production workflow and specified in the project-wide workflow process as illustrated in Figure 4-7. Only the portion related to data quality control is included.

For estimates related to the data resubmission process due to QC failure, it is considered that 20% of the total data files involved will fail to pass the QC inspection based on the results from ETB'96 project [Doucette, 2001]. Therefore, these data files need to be returned to the production contractor and resubmitted for further inspection. This returning and resubmission process would require the same time duration in terms of shipping and receiving data files as required in normal submission process.

Table 6-3 Duration and elapsed time of selected QC process tasks

	QC Process Tasks	GeoPM		ETB'96	
		Duration (hour)	Elapsed Time (hour)	Duration (hour)	Elapsed Time (hour)
Production Contractor	Data Submission (Packing & Shipping)	1	1	24	48
Project Manager	Data Receiving	1	1	3	24
Project Manager	Data Dispatching	0*	0*	24	48
QC Inspector	Data Receiving (Checking & Loading Files)	1	1	3	24
QC Inspector	QC Program Invoking & Results Summary	2	2	10	48
QC Inspector	Manual QC Inspection	5 days	5 days	5 days	5 days
QC Inspector	Data Returning (Packing & Shipping)	1	1	24	48
Production Contractor	Data Resubmission (Packing & Shipping)	1	1	24	48
QC Inspector	QC Reporting #	1	1	24	48
Project Manager	Final Data Delivery	1	1	24	48
	Total (without resubmission)	**5.3 day**	**5.3 day**	**9.6 days**	**17 days**
	Total (with resubmission)	**5.4 day**	**5.4 day**	**11.6 days**	**21 days**

* In ETB'96 project, the project manager delivered source data materials to production contractors based on the contract, while using *GeoPM* the project manager notifies the production contractors and the production contractors download data files from the project server.
# QC Reporting task only considers the time required for delivering QC report. The time spent on preparing QC report is not included.

The table shows that, for accomplishing all QC workflow tasks for one shipment including one round of resubmission, the *GeoPM* can help save 6.2 days in duration and 15.6 days in elapsed time. If the data files in the shipment pass all QC tests at the first round, the corresponding time saved would be 4.3 days and 11.7 days. However, the total time savings in data quality control process for both cases is approximately 60% comparing elapsed times.

Reduced Time on Project Reporting

Weekly reports, contract summary reports, and data file status reports are among the most important ones required in project management. Since a lot of project progress information such as data file status is already stored in the project database, the first two kinds of reports can be created automatically (as discussed in the previous chapter) on-the-fly and viewed online almost instantly. Compared with ETB'96 project where reports had to be prepared by a dedicated person working hours per week (2 – 4 hours) and sent through regular mails taking at least two days, the time saved is significant.

Table 6-4 shows the required time for selected project reporting activities. While production of the ETB'96 relied on the use of spreadsheet software and dedicated personnel to prepare project reports, the collaborative workspace provides software tools that allow extraction of status information from the project database and generation of reports on-the-fly. The status and project progress information is normally updated and stored in the project database through the execution of appropriate workflows.

Table 6-4 Time required for selected project reporting

Types of Reporting	GeoPM	ETB'96
Weekly Report on Completion	Anyone Instantly	1 junior technician 2 hours per report
Contact Summary	Anyone Instantly	1 junior technician 1 hour per report
Summary of Data File Status (including Status Map)	Anyone 5 -10 minutes	1 junior technician 4 hour per report

6.1.2 Cost Factors

Cost factors discussed here include both implementation costs of the collaborative workspace and costs incurred in project management and operation activities.

Implementation Costs of a Collaborative Workspace

While GIS data production already involves the extensive use of computers to process digital data using GIS software and other production specific software tools such as automatic QC testing programs, the collaborative workspace requires more than desktop computers. To implement a collaborative workspace that would support the ETB'96 project, a number of computers and Internet connections as well as software support are needed to realize functions discussed in Chapter 5. These required components are presented in Table 6-5 with corresponding costs.

The implementation costs estimated in Table 6-5 are quite conservative and based on the current market values and projected costs of prototype development effort made in this thesis research. While the maintenance cost is usually hard to project [Finley and Coleman, 1999], it may be offset by the lifetime values of the cost components because the ETB'96 project only last for one year and a conservative estimated lifetime of these components would be 3 years. The following further explains the cost calculations:

Table 6-5 Implementation cost of collaborative workspace

Equipment, Software and Internet Services	**Cost Estimation**	**Cost per Component**
System Setup		
Initial Data Input	5 days @ 8 hours / day, $30 / hour	$1,200
Software Acquisition	$3,000 (exclude freeware)	$3,000
Hardware		
GeoPM Server	1 unit @ $6,000 (light server)	$6,000
IICW Client PC	4 unit @ $1,000	$4,000
Internet Access		
Website Hosting	1 unit @ $200 / month	$2,400
IICW Client Access	5 units @ $600 (high-speed cable connection)	$3,000
Maintenance	–	–
Total		**$19, 600**

- ***Initial Data Input:*** Cost incurred for entering process and workflow definitions into the collaborative workspace and populating the project database with setup data.
- ***Software Acquisition:*** Cost of purchasing software licenses and manpower required to customize collaborative workspace including writing programs for supporting customization and modifying web interfaces (HTML pages). However, this cost is

hard to estimate since some software packages used in prototype development were purchased by UNB for multi-purposes. Clearly, software acquisition and maintenance costs would need to be added here, but they could be amortized over many projects carried out during a given year.

- ***GeoPM Server:*** Expense to purchase a light computer server and the required operating system.
- ***IICW Client PC:*** Expense to purchase desktop computers that project participants use to access the project workspace.
- ***Web Hosting:*** Monthly payment to the selected Internet Service Provider (ISP) for maintaining an Internet domain name and the project web site.
- ***IICW Client Access:*** Monthly payment to ISP for Internet connections. The cable-based Internet connections are assumed in this case because the "dial-up" connections through telephone modem cannot sufficiently support GIS data file transfer over the Internet (see discussions in Section 6.1.1).

Since geomatics firms are now ubiquitously handling GIS data in digital formats, desktop computers have become essential facilities in all types of data handling processes, including data production and data quality processes. Taking advantage of these existing computing facilities, the cost projected on "IICW Client PC" in Table 6-5 may be eliminated, which brings down the overall implementation cost to $15, 550. Considering the lifetime values of involved cost components, the total cost may be reduced by re-using them for other purposes, or if the data production is on a program basis, distributed to various phases of the program.

Cost Incurred in Project Management and Operations

It is not feasible to determine the overall cost effectiveness of the proposed approach without testing it in, or parallel to, a real world production project to obtain necessary "hard" numbers. However, some measurable cost savings discussed here can at least contribute to this determination. These measurable costs are incurred on both consumed materials and manpower required to perform workflow tasks.

1. ***Cost incurred in data file shipping and shipment handling*** **–** For each shipment, the estimated cost includes expenses on media used to store data files, materials used for packing, and postal service – usually by courier. Based on the current market values, the cost for each shipment would be around $50. Since each shipment is delivered at least two times (one delivery for inspection and one delivery for payment) in normal QC process, the total number of shipping in ETB'96 would be 28. Considering another 20% increase due to QC failures [Doucette, 2001], the final number is 33. Therefore, the cost spent on data file shipping and handling would be $1650. This cost could be saved by using the collaborative workspace approach.
2. ***Cost spent on manpower*** **–** The number of working hours per workflow performer on selected tasks is used to estimated cost spent on manpower. Table 6-6 summarizes estimated manpower costs incurred in handling shipment (receiving, checking, packing, and shipping) and batch QC programs in *GeoPM* and ETB'96 projects. The estimations are based on performing selected tasks on data files in one shipment, which results in a cost saving of $900.

The proposed approach favors a “paperless” concept, i.e. all documents, reports, project management forms, etc. should be transferred electronically and viewed in electronic forms. Therefore, cost for paper, printing and plotting, photocopying, and faxing should be considered minimal but not “zero” since in some cases signed documents, reports, or map plots in hardcopy are still required. Even though, the cost of paper, postage and manpower are still significantly lower than those incurred in ETB’96 project.

Table 6-6 Cost comparison of required human resources

Cost Breakdown	GeoPM	ETB’96
Preparing Data Submission	0.5 hours @ $40 / hour	2 hours @ $40 / hour
Handling Received Data Shipment	1 hour @ $40 / hour	3 hours @ $40 / hour
QC Program Invoking & Results Assembling	1 hours @ $40 / hour	10 hours @ $40 / hour
Preparing Data Returning Shipment	0.5 hours @ $40 / hour	2 hours @ $40 / hour
Preparing Data Resubmission Shipment	0.5 hours @ $40 / hour	2 hours @ $40 / hour
Preparing Final Data Delivery	1 hour @ $40 / hour	8 hours @ $40 / hour
Total Cost	**$180**	**$1080**

6.2 GIS Data and Service Quality

It is not a trivial task to assess the improvement of data quality the proposed approach brings into the GIS data production projects. Castonguay and Doucette [2000] describe one way to quantitatively measure the quality enhancement gained by introducing an independent QC inspection contractor. Since all data files produced by production contractors contain errors, if an independent QC inspector is not involved, these data files with errors will be delivered to the customer. In this case, quality improvements may be measured based on various statistics such as percent of data files failed to pass QC inspection, percent of features being corrected, etc. Without testing in real world project environments or performing parallel testing, however, it is not possible to quantify quality improvement that the collaborative workspace contributes.

The quality improvement of data files was then be measured qualitatively based on the following facts: (1) the communications capabilities such as forum and whiteboarding improve discussions on technical specifications and reduce misinterpretations of these specifications; (2) the controlled document review/distribution process ensures the availability of the most updated specifications; and (3) easy access to all other production related documents. All of these will help reduce the number of features needing to be corrected in each data file submitted for quality control and the number of data files failing to pass QC inspections. Therefore, the quality of data produced would be improved. Further, the overall float of the QC process may be shortened since the number of data returns-and-resubmissions is reduced.

Services provided through project management usually involve provision of necessary project documents, acknowledgement of submission and resubmission received, reconciliation of project related issues, and regular project progress reporting. In this sense, service may be defined as the ability to produce and provide project information in a timely fashion and are measured based on two factors: time required to produce the information, time required to provide the information, and availability of such information.

One of the major concerns in designing and developing *GeoPM* is to provide sufficient capabilities to support real time project information sharing and communications. While the current *GeoPM* implementation does not provide any support for creating project documents, the document review/distribution workflow can send instant notifications of available documents and/or updates of these documents. Instant acknowledgements for receiving and submitting both documents and data files are also realized through the execution of corresponding workflows. In terms of reconciling project issues, *GeoPM* provides alternative solutions when face-to-face meetings are not possible due to lack of time for travel.

The collaborative workspace is designed and implemented in such a way that all project related information is stored in the centralized project database and organized in various forms (e.g., documents, reports, and structured messages) in the project technical library (see Chapter 5). All this information is then available to project participants 24 hours per day and 7 days a week, except the time required to routinely maintain the system and the project databases. The access to this information is only subject to the predefined permissions. Compared with ETB'96 project operation where project information is available only upon request subject to the availability of responsible individuals, this indicates a significant service improvement.

6.3 Other Intangible Benefits

While both quality and service improvement discussed in the previous section may be considered as significant intangible benefits since they are difficult to assess exactly, there are also some other intangible benefits the proposed approach may bring to the collaborative GIS data production environments. These benefits include at least improved communications, improved decision-making on project issues, improved industry images of participating companies, and an improved project knowledge base.

The collaborative workspace provides a number of Internet-based communication channels which make the communications of project issues easier than before or at least, more options are available. In addition, the communication results are captured by the workspace and stored in the project attribute database so that they can be referenced later. It also becomes easier for project participants to collaboratively make informal or formal decisions on any project related issues because of these supporting channels.

The collaborative workspace is designed in such a way that most information flowing through the workspace is captured and stored, including timely information such as the status of data files at some moment. The information is then transferred into a project attribute warehouse, where all historic data about the project is held. This actually creates a kind of project "memory" allowing lookup of the information and analysis of project performance at any later time. The project knowledge base not only helps project management but also benefits similar projects conducted at a later time.

Finally, applying the proposed approach to GIS data production projects will help improve the industry image of all participated companies, especially the company acting as project manager since it gives clients an impression of possessing strong technological capabilities.

Chapter 7 CONCLUSIONS AND RECOMMENDATIONS

This thesis describes an attempt to integrate functionality stemming from existing CSCW and groupware tools with support from the Internet especially extranet model into current GIS installations to support selected distributed GIS data production processes. The focus is on understanding unique collaboration requirements, and how ad-hoc GIS and groupware functions can be fit into production processes to support real-time decision-making and project coordination and collaboration activities within contract-based production environments. The thesis has resulted in a collaboration model and a research prototype called ***GeoPM***, developed based on the designed model.

This chapter briefly summarizes the work described in the previous chapters with a recapitulation of research objectives defined in Chapter 1, followed by discussion on the results in general and conclusions obtained. The future work which could improve the research outcomes and implications of further usage of collaborative systems in a wider area in GIS environments are then discussed.

7.1 Realization of Research Objectives

In the first chapter of the thesis, the research objectives of the work reported was defined to develop a collaboration model, including a workflow model, an architecture model, and an implementation framework, that would be used for developing a research prototype to verify the research hypothesis restated as follows:

The collaborative workspace implemented based on the proposed model can provide sound performance for improving distributed GIS production operations and management over Internet-based network infrastructure.

To realize the objectives and examine the hypothesis, the research was done in a series of systematic steps which followed the general software engineering approach, while special considerations were given to characteristics of CSCW-based application development. The rest of this section summarizes major research components and analyzes how the research objectives were reached through these components.

1. Identification and Specification of Collaboration Requirements for Developing a Collaboration Model

The design and development of the collaboration model as defined in the first research objective largely depends on a comprehensive understanding of the existing GIS data production work environments as well as project practices based on the contract-based production model. To accomplish this, an object-oriented modeling language – UML – was used in this thesis to model existing project management and production processes and to capture and specify collaboration requirements necessary for supporting collaborative GIS data production. While the modeling effort was made through a process-centric approach, the capture and specification of collaboration requirements was system oriented, i.e. focused on the functional and non-functional requirements of the proposed collaborative workspace.

To ensure an in-depth understanding of existing GIS data production project practices, effort was made to visit several private companies and government organizations, from which a large number of sample forms, work sheets, procedures, and relevant documents were collected. The outputs, as presented in Chapter 3, include a set of UML activity diagrams and class diagrams capturing both management and production processes as well as associated process resources, and more than 13 use cases specifying functional collaboration requirements in a use case model. However, non-functional requirements are not always described within use cases. The collaboration requirements were further verified with a brief review of several major groupware packages. Gaps between the required functionality and available functions were identified. These outcomes plus the knowledge summarized in Chapter 2 provide a sufficient and solid foundation for the design and development of the collaboration model.

2. Design and Development of Collaboration Model Focusing on Both Process and Structure Perspectives

This part of the thesis work was to fulfill the first research objective, i.e. *to design an Internet-based geomatics production collaboration model which includes: (a) a workflow model, (b) an architecture of workspace, and (c) an implementation framework to* *facilitate the development and implementation of the collaborative workspace*. To ensure the accomplishment of this objective, a participatory and prototype-oriented development approach was adopted. UML and workflow process definition language (WPDL) were used as much as possible to ensure the clear representation of the designed model.

The designed model, which was described in Chapter 4, reflects all production processes modeled and collaboration requirements captured in Chapter 3. The accomplished components (sub-models) of the designed model are summarized here:

- ***The Workflow Model*** provided a process view of the collaborative workspace. To ensure the lasting value and generality of this sub-model, the workflow reference model (WRM) and its interface specifications from WfMC were carefully studied and a hierarchical structure for a workflow process repository was developed. UML and WPDL artifacts were created to present model instances both graphically and textually. The model characterizes existing GIS data production processes with respect to organization structures, invoked applications, and a data model, especially the concept of including a QC testing program repository was presented.
- ***The Architectural Model*** presented a structural view of the collaboration workspace. To ensure component-based structure of the collaborative workspace, the overall logic architecture was designed at higher level into three separate, but linked layers, i.e. presentation layer, business logic layer and data access layer. The presentation layer contains IICW interfaces which can be web-based, stand-alone, or API to

ensure wide suitability. The overall architecture was then populated with the most significant components in each layer.

- ***The Implementation Framework*** described a five-step "*technology driven*" process in implementing the collaborative workspace based on the designed workflow and architectural model. In order to design a more feasible framework, effort was made to learn lessons from existing CSCW application implementation practices and problems encountered during design and development stages of the proposed collaborating systems.

3. Development and Implementation of Prototype Collaborative Workspace in a Internet-based Environment

The accomplishment of the first objective led to the framework for *developing a prototype collaborative workspace* which is the second research objective of this thesis, addressed in Chapter 5. The development was based on a hybrid approach where both software components and stand-alone systems were used and integrated into the prototype system.

To select necessary software components or systems appropriate to the prototype development, a large number of groupware components, workflow systems, and GIS toolkits were examined and some of them were evaluated based on the available evaluation versions. The most difficult part in obtaining necessary software support is to find a proper workflow management system, for which an extensive product search was done through Internet search, email and telephone contacts, and product document review. Another difficult task was to obtain the software components that do not rely on any commercial systems and that can be easily programmed into the collaborative workspace. The developed prototype provides a total of twelve (12) major functions, some of which contain sub functions. To integrate or develop these required functions, a substantial amount of programming work was undertaken using Visual Basic, Java, and PL/SQL languages. These programs contain over a thousand of lines of code, although the primary focus of the prototype development was not on programming. The other major effort made was to learn, install, configure and manage the selected software systems supporting the prototype, such as Oracle Database and Workflow systems. Although this type of work is not trivial for a thesis research project, it has proven to be an excellent learning experience for future work and research in the relevant areas.

Developing a collaborating system to support project management and operations in a distributed GIS data production work environment is a huge task requiring extensive research, which can not be accomplished within one thesis. This is not only because of the design and implementation complexity of CSCW and groupware tools in real-world applications, but also due to multi-participation nature of involved projects. The incremental approach of adding functional components adopted in this thesis has been verified as an efficient way in prototype development.

4. Performance Analysis and Testing on Selected GIS Data Production Process with Selected Parameters

Another of the original proposed objectives, as defined in Chapter 1, was *to test the performance of the proposed approach with both an in-lab simulation method and application situations parallel to a real-world GIS data production project*. However, due to the lack of proper projects, the testing only performed with the first method and, even with this method, the performance analysis is limited due to the lack of actual performance data from the existing projects.

The performance testing and analysis, as discussed in Chapter 5 and Chapter 6, focused on two important aspects: time savings and cost factors. In an attempt to verify whether the collaborative workspace can help reduce the production “float” or not, the Internet performance of FTP data files was first tested at the application level using a specially coded testing program, followed by the measurements of elapsed time of several selected workflow tasks. While the cost factors are more difficult to quantify without applying the proposed approach to a real data production projects, some selected cost breakdown items can be still estimated based on the current market values. With all the limitations acknowledged, the results are still believed to be positive in proving the research hypothesis, especially with the consideration of those un-quantified factors and intangible benefits that the proposed approach may bring in.

7.2 Discussion of Thesis Outcomes and Conclusions

The work presented in this thesis represents a new approach for supporting distributed GIS data production project management and operations. The research hypothesis was considered proved in the sense – based on criteria used to measure positive outcomes [Baeza-Yates and Pino, 1997] – that the total reduction of idle time (based on the difference of elapsed times) introduced into the life cycle of the tested QC process by the proposed approach is about 60%. Since testing was not performed in a real project environment, exact time savings and cost effectiveness could not be assessed. However, it was demonstrated that cost savings on limited cost items basically balance the capital investment on implementing the collaborative workspace. Other significant un-quantified cost savings such as savings on office suppliers and communication spending are reasonably considered as indicators of improved cost effectiveness. The research hypothesis, in this sense, was also proven.

The research results indicates that, using the existing Internet infrastructure to transfer data files involved in GIS data production projects, the performance is still limited to the actual network traffic conditions and unexpected network interruptions. While transferring multiple small-size data files in one FTP session may be acceptable, transferring large-size image files in the same way is problematic. This is further verified by DataQC with its recent practice of allowing production contractors to submit or resubmit small-size data files for QC inspection over the Internet [Roberts, 2002]. The related functions were designed in such a way that data files can be optionally transferred on an individual basis, which mostly ensures the successful transfer of both small-size data files and large-size image files. This, however, imposed extra workload on handling data transfer tasks.

The research results also indicated difficulties in finding appropriate software components that are commercially available in the market and can operate independently of other (unwanted) systems or components. This compromises the component-based design principles followed in the model development for this research. As a result, the prototype development was forced to adopt a hybrid approach of integrating both software components and commercially available software systems. The integration of these components and systems is not a trivial task because of the different software interfaces and designs adopted, which required a fairly large amount of programming work to put them together.

The developed collaboration model facilitates the use of emerging technologies such as groupware and workflow to improve the efficiency and productivity of the collaborative GIS data production by ensuring information concurrency, accessibility and availability, and more importantly, a reengineered workflow process which allows better control and execution of the associated production project activities and procedures. The GIS data production processes have been characterized as “data-centric” processes because the “flowing” of data files through the processes actually controls the progress. In this sense, all other project resources should be associated with data files as much as possible to ensure a consistent management manner in the collaborative workspace.

Finally, part of thesis work demonstrates an early effort (as of the thesis writing) in using UML in systematic, comprehensive modeling of geomatics related business processes, although some applications of workflow technology have been found in managing GIS application processes [Medeiros et al., 1996; Weske et al., 1998]. The systematic approach described should be valuable and useful as a framework for any other similar effort.

In summary, this research, funded by the GEOIDE Networks of Centres of Excellence, brought together GIS, Internet, Groupware, and Data Warehousing principles and applied them into a real-world GIS data production environment to provide solutions for better production problem solving, operational process control, project information sharing, and project managing. The outcomes of this research indicate that (1) CSCW technologies such as workflow can be used to effectively facilitate collaborative GIS data production tasks in a distributed work environment if a *"sound"* model is in place; and (2) the performance of the underlying supporting systems are justified at least based on the analytical results from in-lab simulating testing.

7.3 Recommendations for Future Research

Several issues related to the collaboration model and the research prototype to provide alternative solutions to support distributed GIS data production project management and operations need further investigations, developments and evaluations. These issues are discussed as follows:

(1) All data involved in the underlying data production project should be stored and managed in a GIS database and archived in a data warehouse.

The designed architectural model in this research requires all data involved in the production to be stored in the project operational database and archived in project data warehouse. However, the actual prototype development only introduces data warehouse in terms of managing project attributes and uses a hierarchical file directory to manage and store all data in its original data file format. With this implementation, situations such as managing multiple versions of datasets and partially updating datasets that are already in production become very difficult to handle.

Having a GIS operational database and data warehouse will allow a consistent way to store, manage, update, and archive GIS data. During the production process, data can be checked into and check out from the operational GIS database and the versioned data will be transferred into the data warehouse. When it comes to the situation where the versioned data need to be referenced, it will be retrieved from the project data warehouse. While both GIS databases and data warehousing technologies have been already extensively researched [Bedard, 2001; Rivest and Bedard, 2001], further research effort is required to integrate them into the collaborative workspace implementations. In addition, the performance and effectiveness of using them in collaborative GIS data production environments have to be determined.

(2) The production processes controlled by the project participating companies should be integrated with the project-wide process to obtain the maximum gains of time efficiency and cost effectiveness.

To be simple and feasible for one thesis research project, the design and prototype development described in this thesis has been focused on a project-wide process without touching individual production processes controlled by project participating companies. Ideally, however, all these processes should be integrated under the project-wide workflow process either as sub workflows managed by the project workflow server, or as workflows managed by separate workflow servers and integrated together through standard workflow interfaces as defined in the Workflow Reference Model [WfMC, 1995]. The second approach requires the possession of a workflow system at each participating site. While the research provides a useful starting point towards this development, further work must be done to model the complete hierarchy of processes and to evaluate the suitability of the above approaches for workflow process integration.

(3) The overall performance of the collaborative workspace should be tested and analyzed in real-world production environments to obtain better understanding on improved time efficiency and cost effectiveness.

The research indicated that the ultimate performance measures depend on the parallel and, eventually, real-time testing of the prototype in a real production environment. No matter which case, the testing site or project must be carefully selected and the industry commitment of participation must be secured before hand. In addition to the performance factors discussed in the thesis, other things have to be considered in further testing such as the impact of adopting data file compress/decompress and partial transfer mechanisms in data file transmitting and the impact of integrating GIS operational database and data warehouse.

(4) A set of semantics for markup and annotation of shared objects in collaborative GIS should be developed to facilitate the implementation and interpretation of groupware-based GIS functions.

As discussed in Section 2.2.2 and 3.3, electronic "markup" of shared objects during collaborative GIS sessions is an important capability of the collaborative workspace, which facilitates the project communications in such a manner that traditional face-to-face meetings with "marked-up" hardcopy maps in front of meeting attendants may be greatly reduced. One of the problems associated with the capability is the lack of semantics used for specifying "markup" and annotations. The research used three intuitive symbols to present map features in question, in error, and with comments. However, these are only for demonstration purpose and they are far from sufficient in supporting collaborative GIS sessions. A complete set of "collaboration" semantics should be further studied.

(5) A collaborative GIS data production portal should be investigated by extending the research results to handle multiple GIS data production projects over the Internet.

In many cases, a project manager may be managing multiple GIS data production projects at the same time period. These projects may have different starting and ending dates. The need for using the proposed approach to manage multiple projects is perceived based on the personal interviews with the project managers [Doucette, 2001]. The collaboration model designed in this research has potential to be extended to support the concept of a GIS Data Production Service Portal, controlled by a project manager to provide services to and manage many collaborative GIS data projects simultaneously.

While the designed model may be scalable to accommodate requirements from managing multiple projects and their operations, the capacities of individual software components or systems have to be reexamined and the data models used for both GIS data or data files and project attribute information have to be redesigned.

7.4 Implications of Research in Other Geomatics Areas

The significance of this research may be extended, in part or in full, to benefit other geomatics areas beyond distributed GIS data production. The research outcomes relate to formal modeling of processes, decision-making through enhanced communications, information sharing among widely dispersed people, and automatic process control and executions. These may imply potential usage in the following geomatics areas:

(1) Providing Viable Participatory Tools to Support Public Participation GIS

The research has a strong implication in public GIS participation, which has an intention of providing spatially enabled participatory tools, on top of typical non-spatially enabled tools, for supporting public participation in any decision-making efforts affecting public interests. These tools usually provide such capabilities as rating public participation, voting on public issues, sharing spatial views of different arguments, and communicating with decision-makers and other public participants [Khoshafin and Buckiewicz, 1995].

Apparently, the designed and developed artifacts in this research may be used to serve these purposes over the Internet with or without modifications. The forum discussed can be used to facilitate public discussion of decisions to be made. Electronic whiteboarding capability can be used to support shared view of decision scenarios in spatial presentations and, even more, it may allow public participants to express their concerns spatially on the display by "marking up" their ideas. Most importantly, the simple rating and voting processes can be easily modeled and implemented in the workflow management system that allows automatic statistics of voting and rating results and notifying decision-makers of these results.

(2) Enabling Collaborative Spatial Decision-Making

The computer-based coordination and collaboration mechanisms, or CSCW technologies, was initially created to support information sharing among group of people for better and fast decision-making (e.g., group decision support systems) and routing work-related documents in a controlled way (e.g., workflow), mostly in a centralized homogenous environment [Vinze, 1997]. In this sense, decision-making requires not only capabilities of sharing necessary information but also capabilities of supporting good decision-making processes. With the increased number of situations, where decision-makers may be located at various geographical locations, the need for distributing required information and controlling decision-making processes across wide areas becomes significant.

Especially when the decision-making efforts involve spatial representations of scenarios to be selected, capabilities of collaboratively viewing these representations and making comments on top of them offer more facilitations than those offered by typical means such as telephone meetings and electronic meeting systems. Combining electronic whiteboarding capabilities with video/audio conferencing, supported by workflow-enabled decision-making processes (possibly replacing typical mediators), may offer higher application level of spatial decision-making.

(3) Automating Updating Process of Spatial Data Warehouses

Current effort in both academia and industry focuses on methodologies and tools used to integrate, access, query, and analyze spatial data in the geospatial data warehouses with non-spatial information. Attention paid to the management of data integration processes from operational data sources into geospatial data warehouses is still limited. These unattended data integration areas can be better mapped to both data Integrator and Wrapper/Monitor level in the basic architecture of a data warehousing system defined by Widom [1995]. The wrapper component is responsible for translating spatial data from its native format into the format and data model used by the geospatial data warehousing system. The monitor component is responsible for automatically detecting changes of interest in the source data and reporting this to the integrator.

Workflow technologies discussed in the thesis may be considered as a viable solution for managing the extract-transform-load-detect-notify-refresh process to update GIS data warehouses. With the support of a workflow engine, the appropriate scripts or applications for handling individual tasks in a process can be invoked periodically and the process completion ensured. To do so, the updating process has to be carefully modeled in a series of workflow activities that may include extraction, transformation, loading, detection, notification and refreshing, with each activity linked to a separate workflow application to perform corresponding tasks. This potential offers an automatic approach of updating GIS data warehouse based operational databases.

REFERENCES

Aalst, W. M. P. and K. M. Hee (1995) "Framework for business process redesign", *Proceedings of the Fourth Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE'95)*

Amberg, M (1996) "Modelling adaptive workflows in distributed environments" *Proceedings of the First International Conference of Practical Aspects of Knowledge Management*, Basel, Switzerland, October

Action (1997) "Workflow on the Web", a white paper from Action Technologies, Inc., 1301 Marina Village Parkway, Suite 100, Alameda, CA 94501-1028

Baeza-Yates, R. and J. A. Pino (1997) "A first step to formally evaluate collaborative work" *Proceedings of GROUP'97*, Phoenix, Arizona, U.S.A

Bedard, Y., T. Merrett, and J. Han (2001) "Fundamentals of geospatial data warehousing for geo-graphic knowledge discovery", Miller H. and J. Han (Eds.), *Geographic Data Mining and Knowledge Discovery* (Draft), Taylor and Francis, USA

Bédard, Y., M. Nadeau, and M. J. Proulx (2001) "A new tool for user-driven geographic knowledge discovery with application to environmental health indicators", General Annual Meeting of the NCE GEOIDE, Fredericton, June 20-21

Bernhardsen, T. (1999) "Choosing a GIS", *Geographical Information Systems: Management Issues and Applications (Volume 2)*, John Wiley & Sons, Inc., pp. 589 – 600

Bischofberger, W. and G. Pomberger (1992) *Prototyping-oriented Software Development: Concepts and Tools*, Berlin; New York: Springer-Verlag

Boettcher, R. (1999) *Collaborative GIS in a Distributed Work Environment*, Unpublished M. Eng. Report, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N. B., Canada

Booch, G., J. Rumbaugh, and I. Jacobson (1999) *The Unified Modeling Language User Guide*, Addison Wesley Longman, Inc.

Brelsford, K. (2000) "Contracting for Mapping", *Proceedings of ESRI User International Conference*, San Diego, California, USA

Bruce, M., F. Leverick, and D. Littler (1995) "Complexities of collaborative product development" *Tecnovation*, 15(9), pp. 535-552

Cantor, M. R. (1998) *Object-oriented Project Management with UML*, John Wiley & Sons, Inc., USA

Castonguay, R. C. (1999) "Overview of geospatial information activities in New Brunswick." Presentation at the CGDI Roundtable Meetings, Fredericton, N. B., Feb 19th

Castonguay, R and M. Doucette (2000) "Assessing quality of orthmapping – ensuring value for money", *Proceedings of Annual URISA Conference*, Orlando, Florida, USA, August 19 – 23

Chang et al. (2000) "Oracle workflow guide", a Publication of Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065, USA

Chaudhuri, S. and U. Dayal (1997) "An overview of data warehousing and OLAP technology", ACMSIGMOD Record, 26(1), pp. 65 – 74

Churcher, N. and C. Churcher (1996) "GroupARC – A collaborative approach to GIS", Presented at the 8th Colloquium of the Spatial Information Research Centre, University of Otago, New Zealand, July 9 and 11, 1996

Churcher, N., P. Prachuabmoh, and C. Churcher (1997) "Visualisation techniques for collaborative GIS browsers", *Proceedings of GeoComputation '97 & SIRC '97*, Otago, New Zealand, August, pp. 241-250

Cichocki, A., A. Helal, M. Rusinkiewicz, and D. Woelk (1998) *Workflow and Process Automation: Concepts and Technology*, Kluwer Academic Publishers, the Netherlands

Cockburn, A. and S. Jones (1995) "Four principles of groupware design" *Interacting with Computers*, 7(2), pp. 195 - 210

Cockburn, A. (2000) *Writing Effective Use Cases - The Crystal Collection for Software Professionals*, Addison-Wesley Pub Cooperation

Coleman, D. D. and R. Kbanna (1995) *Groupware: Technology and Applications*, Prentice Hall PTR, Upper Saddle River, NJ07458, USA

Coleman, D. J. (1994a) *Geographic Information Systems Performance in a Broadband Communications Environment*, Unpublished PhD. Dissertation, Department of Surveying and Spatial information Science, University of Tasmania, Hobart, Tasmania, Australia

Coleman, D. J. (1994b) "Internal workshop summary: ChartNet workshop on collaborative production.", Summary prepared for the Canadian Hydrographic

Service and Nautical Data International, Ottawa, Ontario, in support of a CANARIE-funded project on application of broadband networks to electronic chart production.

Coleman, D. J. (1999) "Geographical information systems in networked environments", *Geographical Information Systems: Principle and Technical Issues (Volume 1)*, John Wiley & Sons, Inc., pp. 317-329

Coleman, D. J. and J. D. McLaughlin (1988) "The shift to digital mapping: points of impact on management." *Journal of Surveying Engineering*, 114[2], pp. 59 – 70

Coleman, D. J. and S. Li (1999) "Developing a groupware-based prototype to support geomatics production management." *Computers, Environment and Urban Systems*, 23(4), pp. 1 – 17

Coleman, D. J. and R. Brooks (1996) "Applying collaborative production approaches to GIS data collection and electronic chart production" *Proceedings of the 1996 Canadian Conference on Geographic Information Systems*, Ottawa, Ontario (Available on CD-ROM)

Conallen, J. (2000) *Building Web Applications with UML*, Addison Wesley Longman, Inc., 2000

Comer, D. E. (1995) *The Internet Book*, Prentice Hall, Englewood Cliffs, New Jersey 07632, USA

Darnton, G. and M. Darnton (1997) *Business Process Analysis*, International Thomson Business Press, UK

DataQC (1998) "Quality control manual for the production of the coastal softcopy orthophotomap data base (CSODB)", DataQC, Digital Data Quality Control Services, 301 Woodstock Road, Fredericton, N.B., E3B 2H9, Canada, June

Dawe, P. (1996) *An Investigation of the Internet for Spatial Data Distribution*, M. Eng. Report, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B., Canada

Densham, P., M. Armstrong, and K. Kemp (1995) "Scientific report on the specialist meeting for NCGIA initiative 17", Position Paper of National Center for Geographic Information and Analysis, Sata Barbara, California, U.S.A http://www.ncgia.ucsb.edu/research/i17/I-17_home.html

Dewalt, C. (1999) "Business process modeling with UML", Instruction materials, http://www.omg.org/docs/ad/00-02-04.pdf

Dodgson, M. (1993) "Technological collaboration in industry", Routledge, London

Doucette, M (1998) Personal communication on actual QC processes for ETB'96, Watermark Industries, Inc., Fredericton, New Brunswick, Canada, July

Doucette, M (2001) Personal communication on existing project performance, Watermark Industries, Inc., Fredericton, New Brunswick, Canada, July

Dykeman, J. (1998) "The road ahead for workflow" *Managing Office Technology*, 43(3), pp. 28 – 34

Edwards, W. K. and E. D. Mynatt (1997) "Timewarp: techniques for autonomous collaboration." *The Proceedings of CHI'97*, March 22 - 27, Atlantic, Georgia, USA

Eriksson, H. and M. Penker (1999) *Business Modeling with UML*, Addison Wesley Longman, Inc., USA

ESRI (1998) "MapObjects IMS users guide", the product document from Environmental System Research Institute, Inc., Redlands, 380 New York Street, CA 92373-8100, USA http://www.esri.com/software/mapobjects/ims/upgrade.html

Faber, B. G., R. Watts, J. Hautaluoma, J. Knutson, W. Wallace, and L. Wallace (1994) "A groupware-enabled GIS" *Proceedings of the GIS/94 Symposium*, Vancouver, B.C., February, pp. 551 – 561

Faber, B. G., W. Wallace, K. Croteau, V. Thomas, and L. Small (1997) "Active response GIS: an architecture for interactive resource modeling", *Proceedings of the GIS '97 Annual Symposium on Geographic Information Systems*, Vancouver, B. C., GIS World, Inc.

Faber, B. G. (1998) "Community planning: smart places for community modeling and predictive planning", *URISA'98 Conference Proceedings*, Charlotte, N.C., July, pp. 225 – 226

Favreau J. P. and K. L. Mills (1996) "From national information infrastructure to global collaboration infrastructure", *the Proceedings of KITS' 96*, November

Felice, D. (1996) "Business process & workflow", *Computer Conference Analysis Newsletter*, Feb 21, Issue 380, pp.1 – 17

Finley, D. (1997) *Collaborative GIS in a Distributed Work Environment*, Unpublished M. Eng Report, Department of Geodesy and Geomatics Engineering University of New Brunswick, Fredericton, N. B., Canada

Finley, D. and D.J. Coleman (1999) "Introducing groupware to distributed geomatics production environments", *Journal of Surveying Engineering*, ASCE, 125(1), pp. 1 – 17

Finley, D., B. Arseneau, J. D. McLaughlin, and D. J. Coleman (1998a) "The provincial land information infrastructure for New Brunswick: from early visions to design to reality", *Geomatica*, 52(2), pp. 165 – 175

Finley, D., J. D. McLaughlin, B. Arseneau, and D. J. Coleman (1998b) "Serving the public: the New Brunswick land gazette", *URISA '98 Conference Proceedings*, Charlotte, N. C., July 18-22, pp. 268 – 277

Freiden, A. (1997) "A sampling-based method for inspecting third party-produced maps", *Proceedings of ESRI User Conference*, San Diego, California, USA, July 8 – 11

Fry, C. (1999) "GIS in telecommunications", *Geographical Information Systems: Management Issues and Applications (Volume 2)*, John Wiley & Sons, Inc., pp. 819 – 826

Giga Information Group (GIG) (1997) "Business process & workflow", *Conference Analysis*, March 17, 4(404), pp. 1 – 16

Grather, W., W. Prinz, and S. Kolvenbach (1997) "Enhancing workflow by web technologies", *Proceedings of GROUP 97*, Phoenix, Arizona, USA

Greenberg, S. (1991) "Computer-supported cooperative work and groupware: an introduction to the special issues." *International Journal of Man-Machine Studies*, (1991) 34, pp. 133 – 141

Greif, Irene (Eds.) (1998) *Computer Supported Cooperative Work: A Book of Readings*, Morgan Kaufmann Publishers, San Mateo, CA, USA

Grudin, J. (1988) "Why CSCW applications fail: problems in the design and evaluation of organizational interfaces" *Proceeding of CSCW'88*, 85-93, Portland, OR, USA

Grudin, J. (1994) "Groupware and social dynamics: eight challenges for developers." *Communications ACM*, 37(1), pp. 93 – 105

Hardie, A. (1998) "The development and present state of Web-GIS", *Cartography*, 27(2), December, pp. 11 – 26

Hartman, R. (1997) *Focus on GIS Component Software: Featuring ESRI's MapObjects*, OnWorld Press, USA

Hastings, K. L. and M. A. White (1997) "Effective management of GIS projects using quality control and quality assurance standards", the *Proceedings of ESRI International User Conference,* San Diego, California, USA, July 8 – 11

Hills, M. (1997) *Intranet Business Strategies,* John Wiley & Sons, Inc., USA

Hunter, M. (1996) *Quality in Geomatics – A Practical Guide to ISO 9000*, A document under contract to Geomatics Canada, Natural Resources Canada, 613 Booth Street, Ottawa, Ontario, K1A 0E9

IGD: (1995) "CrystalPad: a cooperative annotation environment", http://www.igd.fhg.de/www/igd-a9/projects/CSCW-Tools/CSCW-Tools.html

Jankowski, Piotr (1995) "Collaborative spatial decision-making", Position Paper of the National Center for Geographic Information and Analysis, Santa Barbara, California, U.S.A, http://www.ncgia.ucsb.edu/research/i17/htmlpapers/jankowski/Jankowski.html

Kao, Y. C. and G. C. I. Lin (1998) "Development of a collaborative CAD/CAM system", *Robotics and Computer-integrated Manufacturing,* 14 (1998), pp. 55 – 68

Khoshafian, S. (1998) "WebDeploy: workflow", A White Paper from Technology Deployment International, Inc.

Khoshafian, S. and M. Buckiewicz (1995) "*Introduction to Groupware, Workflow, and Workgroup Computing*", John Wiley & Sons, Inc., USA

Klein, D. (1998) "The Internet and the future of the decision making process", *URSIA'98 Conference Proceedings,* Urban and Regional Information Systems Association, Charlotte, N.C., pp. 682-688

Koch, T. (1999) "XML in practice: the groupware case", *Presented in IEEE 8th International Workshop on Enabling Technologies: Infrastructure for collaborative Enterprises,* Stanford University, California, USA, June 16-18th

Kramer, M. (1997) "Special report: groupware systems are growing up", *PC-Week Online,* http://www.zdnet.com/pcweek/reviews/1117/17group.html

Leverenz et al. (1999) "Oracle *8i* concepts", a Publication of Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065, USA

Li, S., R. Boettcher, and D. J. Coleman (1998) "Geomatics workflow management requirements in a collaborative production environment", *1998 Conference Proceedings of the American Congress on Surveying and Mapping,* March 2-5, pp.269 -

Li, S., D. J. Coleman, and S. M. Easa (1999) “Public utilities" in *Urban Planning and Development Applications of GIS*, Said Easa and Yupo Chan, eds., American Society of Civil Engineers, Reston, Virginia, Chapter 8, pp. 135 – 150.

Lipschutz, R. P., L. Gilbert, K. Heard, J.A. Kent, M. Nguyen, K. Smith, and A. Soofi (1997) *Mastering Netscape SuiteSpot 3 Servers*, Network Press

Manheim, M. (1998) “Beyond groupware and workflow”, J. L. Kellogg Graduate School of Management, Northwestern University, a white paper from the Workflow Management Coalition, http://www.wfmc.org/

Medeiros, C. B., G. Vossen, and M. Weske (1996) “GEO-WASA-Combining GIS technology with workflow management”, *Proceedings of the Seventh Israeli Conference on Computer Systems and Software Engineering*, Herzliya, Israel, June 12 – 13, pp. 129 – 139

Meyers, J. R. (1999) “GIS in the utilities”, *Geographical Information Systems: Management Issues and Applications (Volume 2)*, John Wiley & Sons, Inc., pp. 808 – 818

Muller, M. J. and Kuhn, S. (1993) "Introduction to the special issues on participatory design", *Communications ACM* 36(4), pp. 24-28

Newton, P. W., P. R. Zwart, and M. E. Cavil (1995) *Networking Spatial Information Systems*, Revised Edition, John Wiley & Sons Ltd., England, UK

Nunamaker, J. F. (1997) “Future research in group support systems: needs, some questions and possible directions”, *International Journal of Human-Computer Studies*, 1997(47), pp.357-385

OMG (2000) “Workflow management facility specification”, Version 1.2, the Object Management Group, April 2000, ftp://ftp.omg.org/pub/docs/formal/00-05-02.pdf

Pfleeger, S. L. (1998) *Software engineering: theory and practice*, Upper Saddle River, Prentice Hall, NJ, USA

Plexus (1998) “Workflow for web-based customer service”, A White Paper from Plexus, Division of BancTec, Email: info@plx.com, Tel: 800-999-5910 or 408-747-1210

Plewe, B. (1997) *GIS Online: information retrieval, mapping, and the Internet*, OnWorld Press, USA

Podgorny M., K. Walczak, D. Warner, and G. C. Fox (1998) “Internet groupware technologies – past, present and future”, Northeast Parallel Architectures Center, Syracuse University, 111 College Pl., Syracuse, NY 13244, U.S.A http://trurl.npac.syr.edu/tango/papers/BIS.html

Poltrock S. and G. Engelbeck (1997) "Requirements for a virtual collocation environment", *the Proceedings of GROUP 97*, Phoenix, Arizona, USA

Portfolio, T. (1999) "PL/SQL users guide and reference", a Publication of Oracle Corporation, 500 Oracle Parkway, Redwood Shores, CA 94065, USA

Rivest, S., Y. Bédard, and P. Marchand (2001) "Towards better support for spatial decision-making: defining the characteristics of spatial on-line analytical processing (SOLAP)", Geomatica, Journal of the Canadian Institute of Geomatics, December

Roberts, C (1999) Personal communication on the ETB'96 project, DataQC, Inc., Fredericton, New Brunswick, Canada, July

Roberts, C (2002) Personal communication on the recent practice of using the Internet for QC data submission, DataQC, Inc., Fredericton, New Brunswick, Canada, March

Rojas, E. M. and A. D. Songer (1999) "Web-centric systems: a new paradigm for collaborative engineering", *Journal of Management in Engineering*, 15(1), Jan/Feb 1999, pp. 39 – 45

Sebert, L. M. (1989) "Provincial large-scale topographic mapping", *CISM Journal ACSGC*, 43(1), spring, pp. 39 – 53

Sabourin, M. (1994) "The national topographic data base (NTDB) of Canada", *Proceedings of Decision Support - 2001*, Vol.1, Combined events of the 17th Annual Geographical Information Seminar and the Resource Technology'94 Symposium, Toronto, Ontario, Canada

Sanderson, D. (1992) "The CSCW implementation process: an interpretative model and case study of the implementation of a videoconference system", *CSCW'92 Proceedings*, October 31 to November 4, Toronto, Canada

Seggern, M. (1994) "The enterprise GIS: a client/server approach using distributed relational databases to create a multi location/multi application GIS", *EGIS'94*, http://www.odyssey.maine.edu/

Shah, D., L. Candy, and E. Edmonds (1998) "An investigation into supporting collaboration over the Internet", *Computer Communication*, 20[1998], pp.1458-1466

Shih, H. M. and M. M. Tseng (1996) "Workflow technology-based monitoring and control for business process and project management", *International Journal of Project Management*, 14(6), pp. 373 – 378

Susi-Ross, M. Ramage, and Y. Rogers (1995) "PETRA: participatory evaluation through redesign and analysis", *Interacting with Computers*, 7(4), pp. 335 – 360

Van der Aalst, W. M. P. (1998) "The application of Petri nets to workflow management", *The Journal of Circuits, Systems and Computers,* 8(1), pp. 21 – 66

Vinze, A. S. (1997) "Preface: group support systems", *International Journal of Human-Computer Studies,* (1997) 47, pp. 355 – 356

WaterMark (1995) "Quality Control Manual for the Topological Structuring of the Digital Topographic Database", WaterMark Industries Inc. in association with OPTEX INC., Fredericton, N.B. CANADA, 1996

Weske, M., G. Vossen, C. B. Medeiros, and F. Pires (1998) "Workflow management in geoprocessing applications", *Proceedings of the 6th ACM International Symposium Geographic Information Systems (ACM GIS'98)*, pp. 88 – 93

WfMC (1995) "The workflow reference model", A White Paper from Workflow Management Coalition, PO Box 165, 2 Crown Walk, Winchester, SO22 5XE, United Kingdom, Tel: +44(0) 1962 87340, June, http://www.wfmc.org/

WfMC (1998) "Workflow and Internet: catalysts for radical change", A White Paper from Workflow Management Coalition, PO Box 165, 2 Crown Walk, Winchester, SO22 5XE, United Kingdom, Tel: +44(0) 1962 87340, June, http://www.wfmc.org/

WfMC (1999) "Interface 1: process definition interchange – process model", A Standard Document from Workflow Management Coalition, PO Box 165, 2 Crown Walk, Winchester, SO22 5XE, United Kingdom, Tel: +44(0) 1962 87340, October, http://www.wfmc.org/

Widom, J. (1995) "Research problems in data warehousing", *Proc. of the 1st International Conference on Information and Knowledge Management (CIKM)*, November

BIBLIOGRAPHY

Allen, R. H., S. Nidamarthi, S. P. Regalla, and R. D. Sriram (1999) "Enhancing collaboration using an Internet integrated workbench", *Proceedings of 1999 ASME Design Engineering Technical Conference*, Las Vegas, USA, September 12 – 15

Amberg M. (1999) "The benefits of business process modeling for specifying workflow-oriented application systems", http://www.seda.sowi.uni-bamberg.de/workflow/

Ames, C. K, S. C. Burleigh, and J. S. Mitchell (1999) "WWWorkflow: World Wide Web based workflow", http://www.atinet.org/~steve/workflow.html

Barthelmess, P. and J. Wainer (1995) "Workflow modeling", *Proceedings of CYTED-RITOS International Workshop on Groupware*, Lisbon, Portugal, September

Benford, S., D. Snowdon, A. Colebourne, J. O'Brien, and T. Rodden (1997) "Informing the design of collaborative virtual environments", *Proceedings of GROUP 97*, Phoenix, Arizona, USA

Ben-Shaul, I. Z. And G. E. Kaiser (1996) "Integrating groupware activities into workflow management systems", Paper Presented in the Seventh Israeli Conference on Computer Based Systems and Software Engineering, June

Bentley, R., W. Appelt, U. Busbach, E. Hinrichs, D. Kerr, K. Sikkel, J. Trevor, and G. Woetzel (1997) "Basic support for cooperative work on the World Wide Web", *International Journal of Human-Computer Studies*, (1997) 46, pp. 827 – 846

Calvary, G., J. Coutaz, and L. Nigay (1997) "From single-user architectural design to PAC: a generic software architecture model for CSCW", *Proceedings of CHI'97*, Atlanta, GA, USA, March 22 – 27

Casati, F., S. Ceri, B. Pernici, and G. Pozzi (1995) "Conceptual modeling of WorkFlows", *Proceedings of OO- ER Conference*, Gold Coast, Australia

Clarke, A., B. Soufi, L. Vassie, and J. Tyrer (1995) "Field evaluation of a prototype laser safety decision support system", *Interacting with Computers*, 7(4), pp. 361 – 382

Collaborative Strategies (1997) "Electronic collaboration on the Internet and Intranet", White Paper, http://www.collaborate.com/publications/intranet.html

Cook, S. and G. Birch (1991) "Modeling groupware in the electronic office", *International Journal of Man-Machine Studies*, (1991) 34, pp. 369 – 393

Faber, B.G. (1996) "Extending electronic meetings systems for collaborative spatial decision making: obstacles and opportunities", Position Paper of National Center for Geographic Information and Analysis, Sata Barbara, California, U.S.A

Freiden, A. and M. Johnson (1997) "A sampling-based method for inspecting third party-produced maps", *Proceedings of ESRI International User Conference*, San Diego, California, USA, July 8 – 11

Gains, J. C. (1996) "Project management in the GIS/LIS World", *Proceedings of GIS/LIS'96*, Bethesda: American Society for Photogrammetry and Remote Sensing, 1006 – 1015

Gallo, S. K. (1999) "GIS within electronic meetings: the effective use of spatial data in a joint intellectual effort", *Proceedings of 1996 ESRI User Conference*, Palm Springs, California, May, pp. 20 – 24

Gillner, R. (1998) "Collaborative project management with a web-based database editor", Fh Fulda, Fb Angewandte Informatik, D-36039 Fulda, Germany

Grather, W., W. Prinz, and S. Kolvenbach (1997) "Enhancing workflows by web technology", *Proceedings of GROUP 97*, Phoenix, Arizona, USA

Gronbak, K. M. Kyng, and P. Mogensen (1992) "CSCW challenges in large-scale technical projects – a case study", *CSCW 92 Proceedings*, November

He, P., S. Nag, and E. Chapman (1995) "Quality control in full-scale GIS database development for local government", *GIS/LIS Conference Proceedings*, pp. 407 – 417

Hofte, G. H., H. J. Lugt, and M. A. Houtsma (1996) "Co4: a comprehensive model of groupware functionality", Paper presented in the APTEC conference of Euromedia'96, London, UK, December 19 – 21

Hudson, R. D. (1997) "DT-5: Enabling technologies desktop video Conferencing", http://www.visc.vt.edu/succeed/videoconf.html

Jankowski, P and M. Stasik (1997) "Design considerations for space and time distributed collaborative spatial decision making", *Journal of Geographical Information and Decision Analysis*, 1(1), pp. 1 – 8

Jones, R. M., Copas, C. V., and E. A. Edmonds (1997) "GIS support for distributed group-work in regional planning", *International Journal of Geographical Information Science*, 11(1), pp. 53 – 71

Kao, Y. C. and G. C. I. Lin (1996) "CAD/CAM collaboration and remote machining", *Computer Integrated Manufacturing Systems*, 9(3), pp. 149 – 160

Karhu, V. and P. Lahdenpera (1999) "A formalized process model of current Finnish design and construction practice", *the International Journal of Construction Information Technology*, 7(1), pp. 51 – 71

Intel Corporation (1998) "Collaborative work for small business: working together to get the edge", Intel Corporation Web site, http://www.intel.com/businesscomputing/small/ProductiveComputing/CollaborativeWork.htm

Manhart P. (1999) "A system architecture for the extension of structured information spaces by coordinated CSCW services", *Proceedings of GROUP 99*, Phoenix, Arizona, USA

McCain, M. J. and W. C. Masters (2001) "Preparing for data acceptance: are you ready?" *Proceedings of 2001 ESRI International Users Conference*, July 9 – 13

McConnell, T. and C. White (1999) "Using ArcView GIS as a project management tool", *The Proceedings of 1999 ESRI International Users Conference*, San Diego, California, USA, July 26 – 30

Medina-Mora, R., T. Winograd, R. Flores, F. Flores (1992) "The action workflow approach to workflow management technology", *CSCW'92 Proceedings*, Toronto, Canada, November

Miller, J. A., A. P. Sheth, K. J. Kochut, and D. Palaniswami (1998) "The future of web-based workflows", http://LSDIS.cs.uga.edu

Mills K. L. (1999) "Introduction to symposium on computer-supported cooperative work", *ACM Computing Survey*, March

Mines, J. (1995) "Data capture in the field with the POSSE inspection remote system", *Proceedings of GIS/LIS'95*, Bethesda: ACSM/ASPRS, AAG, URISA, AM/FM, 1995. 2:760-769

Muehlen, M. Z. and R. Allen (2000) "Embedded & autonomous workflow management systems", A Whitepaper from the Workflow Management Coalition, PO Box 165, 2 Crown Walk, Winchester, SO22 5XE, United Kingdom, Tel: +44(0) 1962 87340, June, http://www.wfmc.org/

Netscape Communications Corporation (1996) "Executive summary: a new generation of applications", http://home.netscape.com/one_stop/intranet_apps/summary.html

Ortner, W. and C. Stary (1999) "Virtualization of organizations: consequences for workflow modeling", *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Maui, HI, USA, January 5 – 8

Palen, L. (1999) "Social, individual & technological issues for groupware calendar systems", *Proceedings of CHI'99*, Pittsburgh, PA, USA, May 15 – 20

Radosevich, L. (1999) "What's happening to workflow", InfoWorld, 21(14), pp. 1 – 3

Rezgui, Y., G. Cooper, and P. Brandon (1998) "Information management in a collaborative multiactor environment: the COMMIT approach", *Journal of Computing in Civil Engineering*, 12(3), pp. 136 – 144

Rosenberg, D. and C. Hutchison (Eds.) (1994) *Design Issues in CSCW, Computer Supported Cooperative Work*, Series edited by D. Diaper and C. Sanger, Springer-Verlag London Limited, Great Britain

Swenson, K. D. (1996) "Collaborative development of groupware applications", Fujitsu Open Systems Solutions, Inc., 3055 Orchard Dr., San Jose, CA, 95134

Stewart, J., Bederson, B. B., and Druin, A. (1999) "Single display groupware: a model for co-present collaboration", *Proceedings of ACM CHI'99*, pp. 286-293

Torres-Alejo, I. and Cea, Ileana (2001) "QC and implementation of digital ortho images for Miami-Dade County Government", http://www.GIScafe.com/

Turner, J. and R. Kraut (1992) "CSCW'92 sharing perspectives", *Proceedings of the Conference on Computer-Supported Cooperative Work*, Toronto, Canada, November

Veijalainen, J, A. Lehtola, and O. Pihlajamaa (1995) "Research issues in workflow systems', a Research Report, VTT Information Technology, P.O. Box 1203, FIN-02044, VTT, Finland

Vera, A. H., T. Kvan, R. L. West, and S. Lai (1998) "Expertise, collaboration and bandwidth", *Proceedings of CHI 98*, LOS Angeles, California, USA, April 18 – 23

Wang, J. (1996) "A collaborative GIS decision support model on Internet", Position Paper of National Center for Geographic Information and Analysis, Sata Barbara, California, U.S.A

Weske, M. Goesmann, T. Goesmann, R. Holten, and R. Striemer (1999) "A reference model for workflow application development processes", Proceedings of WACC'99, February, San Francisco, CA, USA

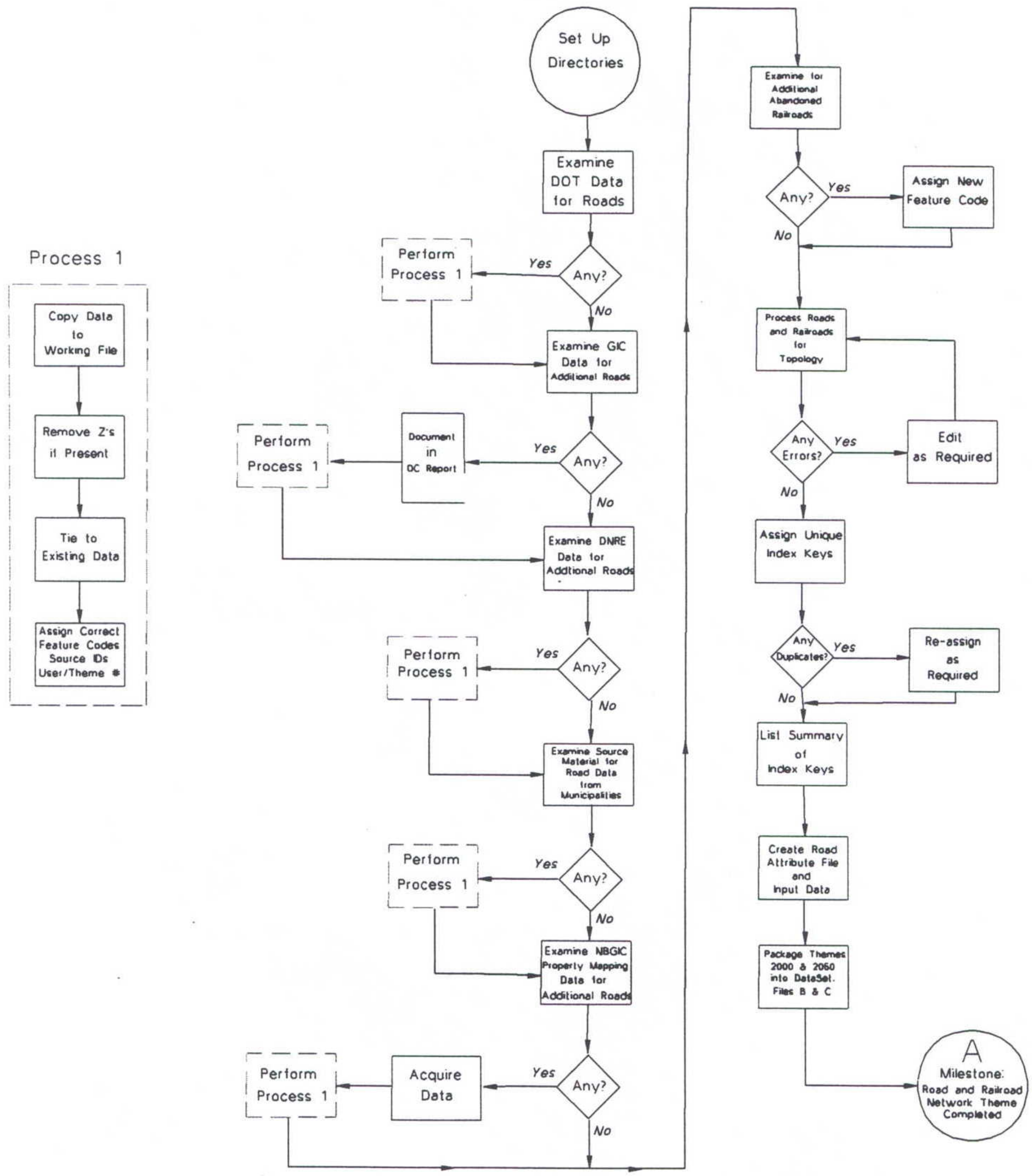
WfMC Survey (1999) "Questionnaire for users and potential users of workflow systems", A Survey from Workflow Management Coalition, PO Box 165, 2 Crown Walk, Winchester, SO22 5XE, United Kingdom, Tel: +44(0) 1962 87340, June, http://www.wfmc.org/

Appendix A Collected Samples

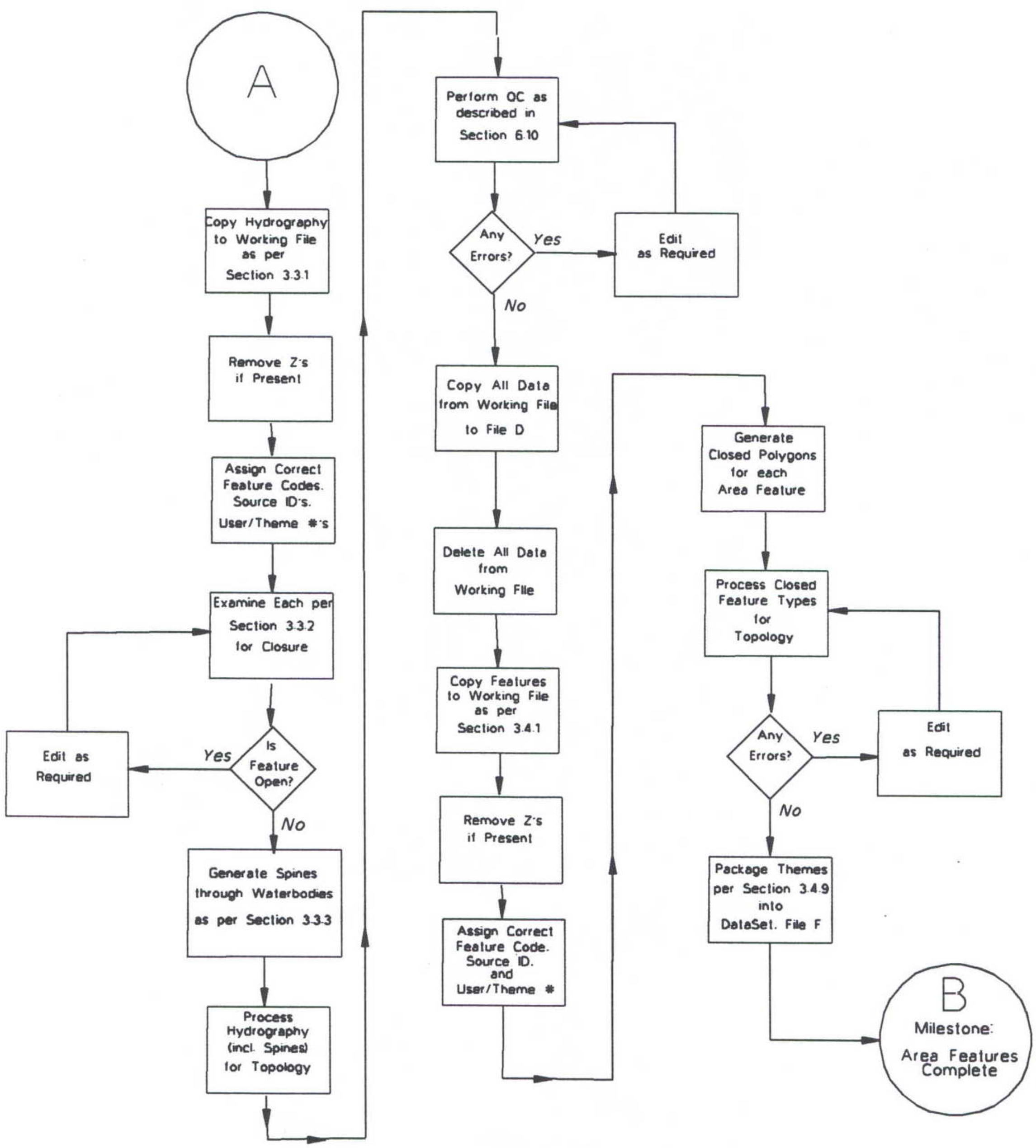
The sample forms, charts and related documents were collected from several Canadian geomatics data production firms and government organizations (as described in Chapter 3) involved in GIS data production contracts. These samples demonstrate some aspects of how current GIS data production projects are executed through paper-based manual systems.

1. Flow charts of data production and/or project processes (Sample 1 – 5);
2. Forms used for procedural and task related scheduling, assignment and management (Sample 6 – 9);
3. Hardcopy status maps showing the project progress or data file processing status information (Sample 10 – 12);
4. Sample batch quality assurance procedure
5. Report templates used for project reporting purpose.

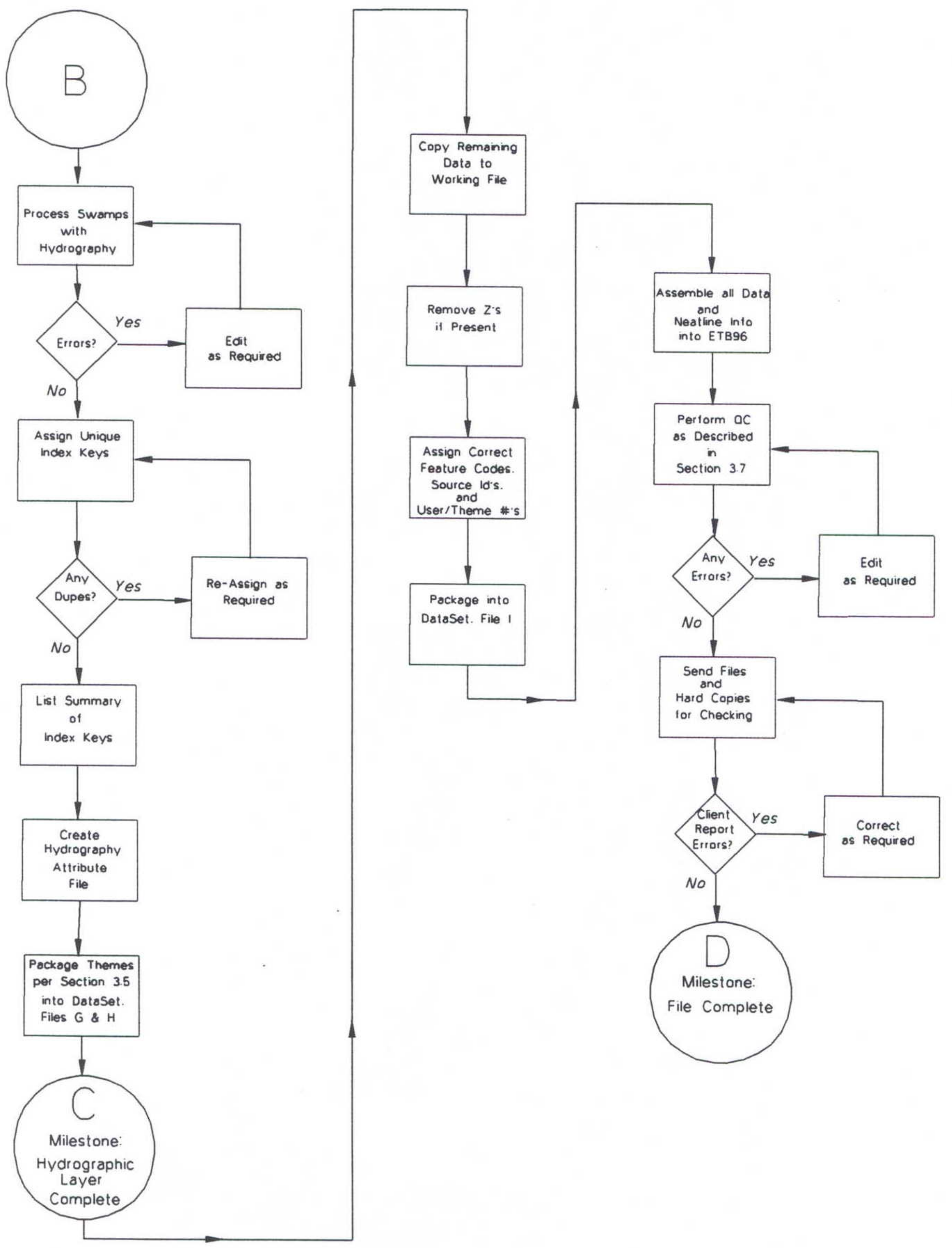
Sample 1: The production flowchart adopted from ETB '96 (Part A)



Sample 2: The production flowchart adopted from ETB '96 (Part B)



Sample 3: The production flowchart adopted from ETB '96 (Part C)



Sample 4: Process connection flow chart from InterMap

DOCUMENT NAME:	Processing of Digitally Complied Data	PAGE:	8 of 13
DOCUMENT NUMBER:	MAP.QPD.0017 V1.4	DATE:	August 5, 1998

5.0 PROCESS FRAMEWORK

5.1 Process Connections Flow Chart (PCFC)

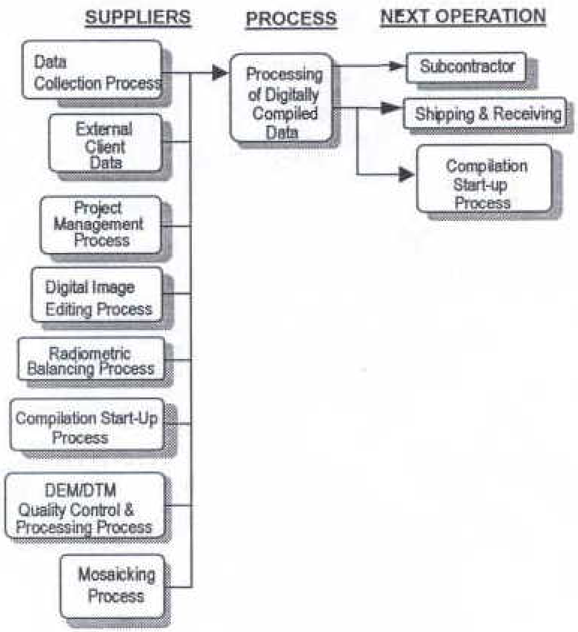


Figure 1. Process Connections Flow Chart (PCFC)

This flow chart shows the relationship of the Processing of Digitally Compiled Data, to the Client supplied data, DOI data, and information supplied by the Project Management Process, DEM/ DTM Quality Control & Processing, *Compilation Start-up Process,* and the internal and external clients which are the process customers.

Sample 5 Process flow chart from InterMap

DOCUMENT NAME:	Processing of Digitally Complied Data	PAGE:	10 of 13
DOCUMENT NUMBER:	MAP.QPD.0017 V1.4	DATE:	August 5, 1998

6.0 PROCESS DESCRIPTION

6.1 Process Flow Chart (PFC)

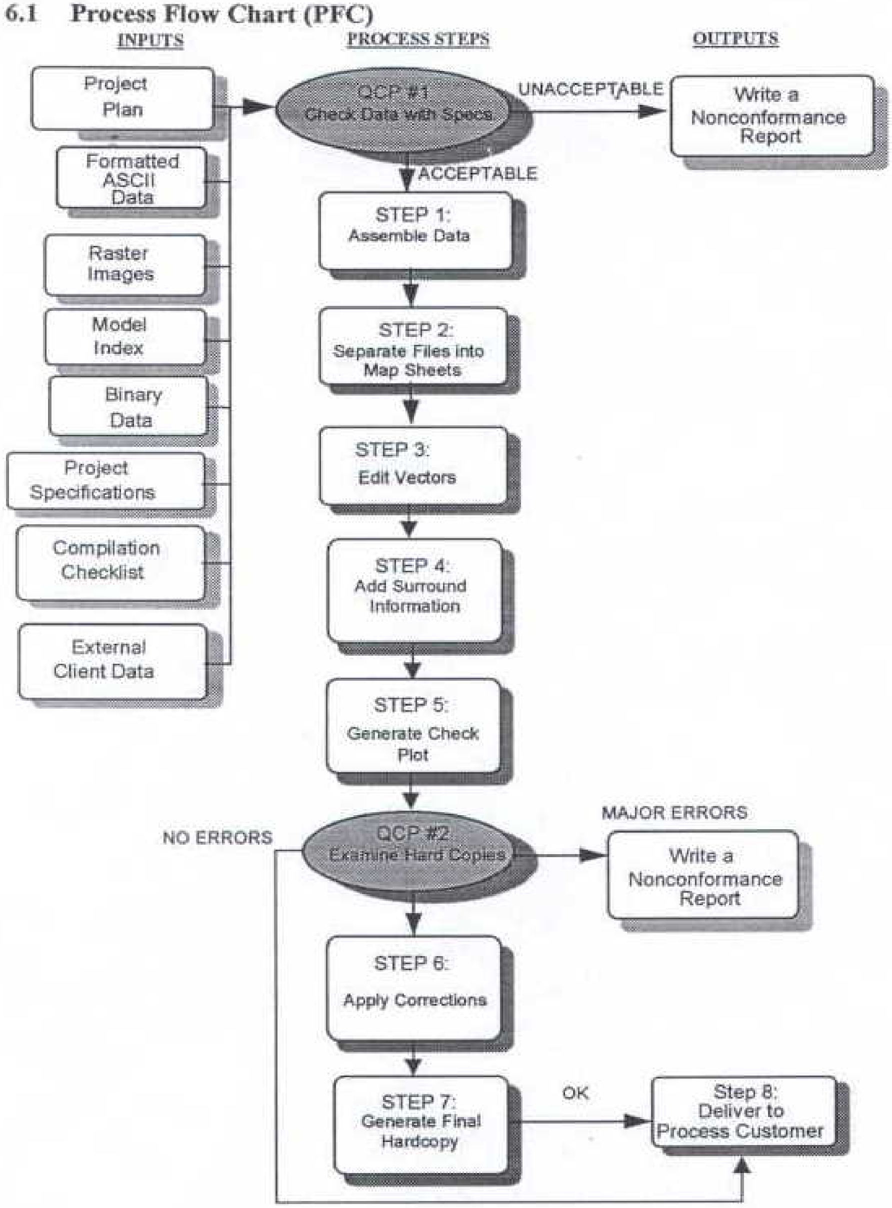


Figure 2. Quality Check Points (QFC) Within the Process

Sample 6 Process checklist used in InterMap

PROCESS CHECKLIST

Applicable	Process	Complete	Date	Comments

Sales Process

	Proposal Definition Form			
	Proposal Review			
	Update Sales Database			
	Budget Estimate Summary			
	Contract Review			
	Update Sales Database			

Project Management

Project Number:		Project Name:	

	Project Plan			
	Project Plan Sign off			
	Production Review Meeting			
	Project Plan Sign off by Production Staff			
	Identify Client Supplied Data			
	Quality Records			
	Final Review Report			
	Invoice			
	Project Archived			

Production Processes

	A/T			
	Compilation			
	Cartography/Editing			
	DOI/Imaging			
	SARMAP – DEM Editing			
	Product Development			
	STAR-3i Processes (ASG)			

Intermap Technologies
INT.QRF.0015 V2.1
R:\STDFORMS\GLOBAL\SALESDOC\PCHKLST.DOC 1 of 1

Company Confidential
Controlled Document
June 8, 1999

Sample 7 Process improvement/nonconformance report form used in InterMap

PROCESS IMPROVEMENT / NONCONFORMANCE REPORT FORM

1. NONCONFORMANCE: ☐	2. PROCESS IMPROVEMENT ☐	3. LOG NUMBER:
4. PROCESS:		5. DATE:
6. SUGGESTION OR PROBLEM DESCRIPTION:		
7. SUGGESTED CORRECTIVE ACTION (SHORT TERM): CORRECTIVE ACTION COMPLETION DATE- PLANNED :		ACTUAL:
8. SUGGESTED PREVENTIVE ACTION (LONG TERM):		
9. IMPACT (cost, time lost, consequences, savings):		
10. ORIGINATOR:	11. ASSIGNED TO: 12. ASSIGNED DATE:	
13. UPDATED BY: 14. UPDATED DATE:	15. CLOSED BY: 16. CLOSED DATE:	
17. DISTRIBUTION:		

DO NOT WRITE IN SHADED AREAS

Intermap Technologies
INT.NCR.0001 V1.8
R:\STDFORMS\GLOBAL\NONCONFO\PIFNCR.DOC

PI/NCR FORM

1 of 1

May 13, 1998
Controlled Document
Company Confidential

Sample 8 Sheet creation & general processing form (F-11A) used in Terra Surveys

terra
surveys ltd

Sheet Creation & General Processing Form — F-11 A

PROJECT NAME: [redacted]

JOB NUMBER: 76-1315

SCALE: 1:10,000

DEADLINE:

Sheet Name to Create:			Directory:		
Models	Operator Initials	Date	Models	Operator Initials	Date

Notes: LOWER LEFT CORNER: X=
Y=

	DATE	INITIAL (please print)
SHEET CREATE		
FIRST SUBMISSION		
PLOT MADE		
QC		
RESET		
EDIT *(76-1315B)*		
SECOND SUBMISSION		
QA Certificate (Form #F-50)		

SHIPPED TO CLIENT	Date	Initials	Delivery Method	

THIS FORM SHALL FOLLOW THE MAP SHEET THROUGH PROCESSING, THEN RESIDE IN THE PROJECT FILE (OR SPECIFIED PROJECT BINDER)

October 19, 1999, Version 1 *C:\ISO 9002\ISO Forms\F-11A Sheet Create (Sweden).wpd*

Sample 9 Sheet creation & general processing form (F-11B) used in Terra Surveys

terra surveys ltd

Sheet Creation & General Processing Form F-11 B

PROJECT NAME: PUERTO RICO - 99113ent

JOB NUMBER: 76-1320

SCALE:

DEADLINE:

Sheet Name		Directory:	
Notes:			
DEPT	TASK	DATE	INITIAL
COMP	Comp-Complete (including 250m overedge)		
MF	Pre-Processing		
ORTHO	Single Frame Ortho		
MF	Plot		
QC	QC		
COMP	Resets		
MF	3DCAR		
QA	Edgematching & QC		
QA	Finalization		
MF	MetaData & Final Chk Plot	N/A	N/A
MF	Translations	N/A	N/A
Quality Assurance			

SHIPPED TO CLIENT	Date	Initials	Delivery Method	

THIS FORM SHALL FOLLOW THE MAP SHEET THROUGH PROCESSING, THEN RESIDE IN THE PROJECT FILE (OR SPECIFIED PROJECT BINDER)

November 24, 1999 Version 2

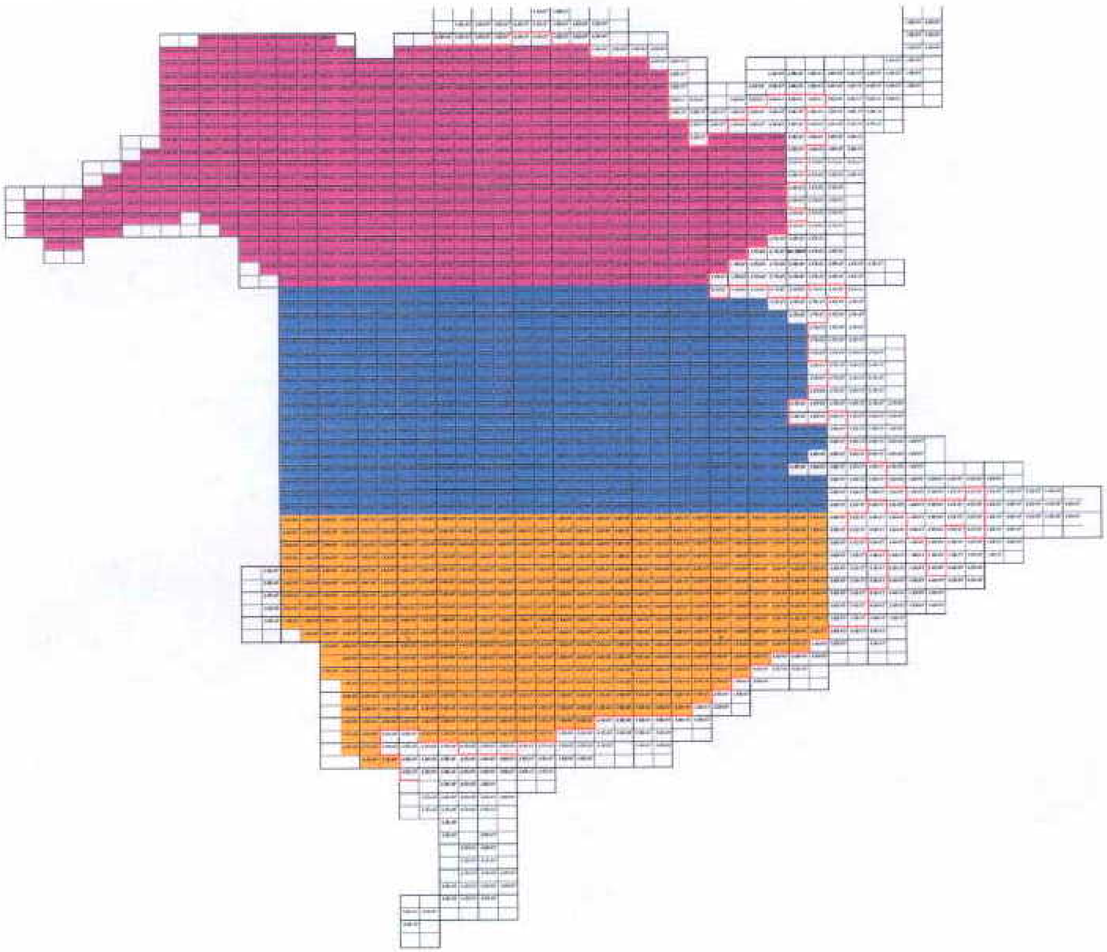
C:\ISO 9002\ISO Forms\F-11B Sheet Create (Puerto Rico).wpd

Sample 10 Block map used in SNB data production project

January 3, 2000 Contract 99-053

APPENDIX A

Map of Blocks 6 / 7 / 8



Block 6

Block 7

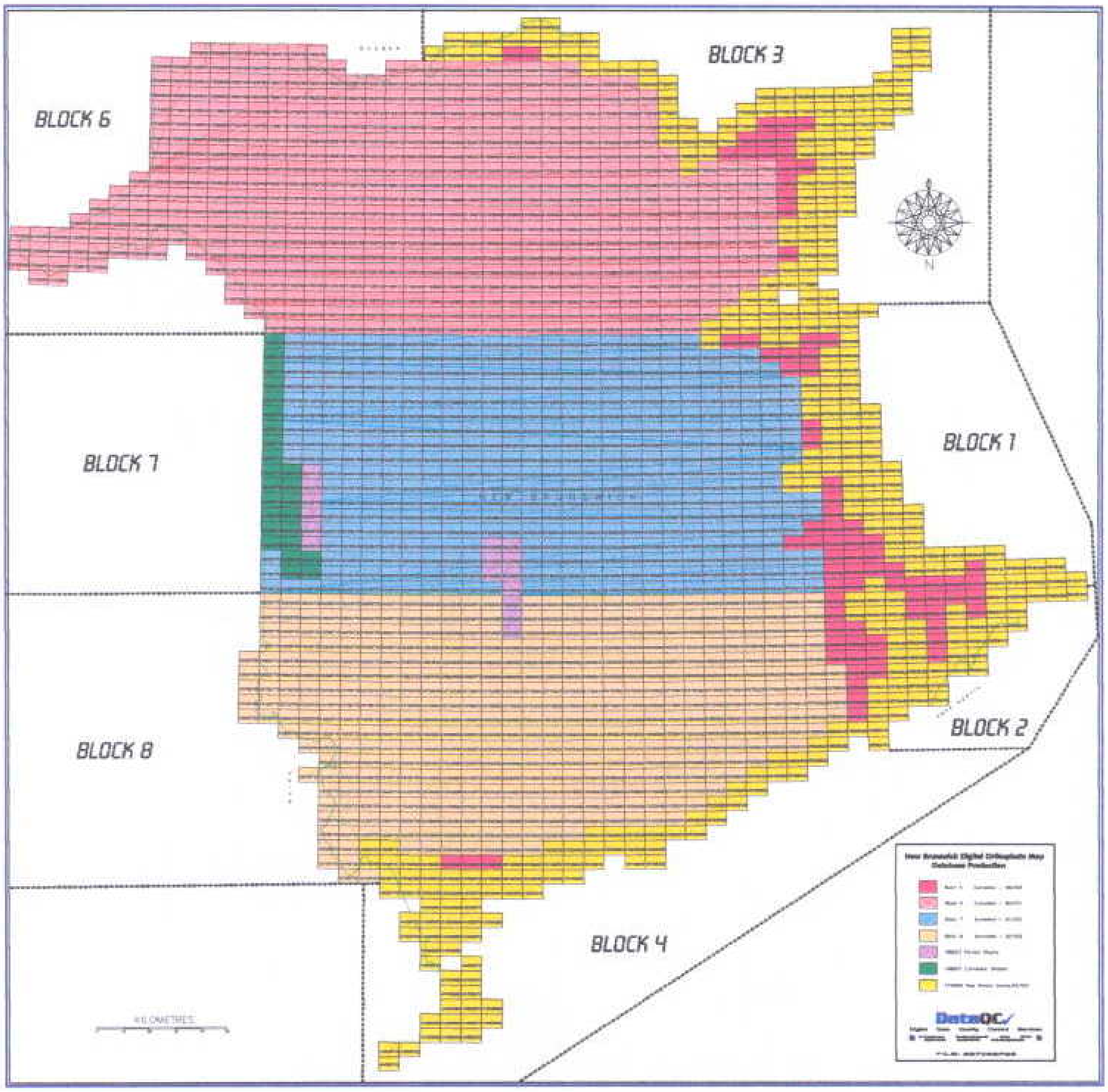
Block 8

17

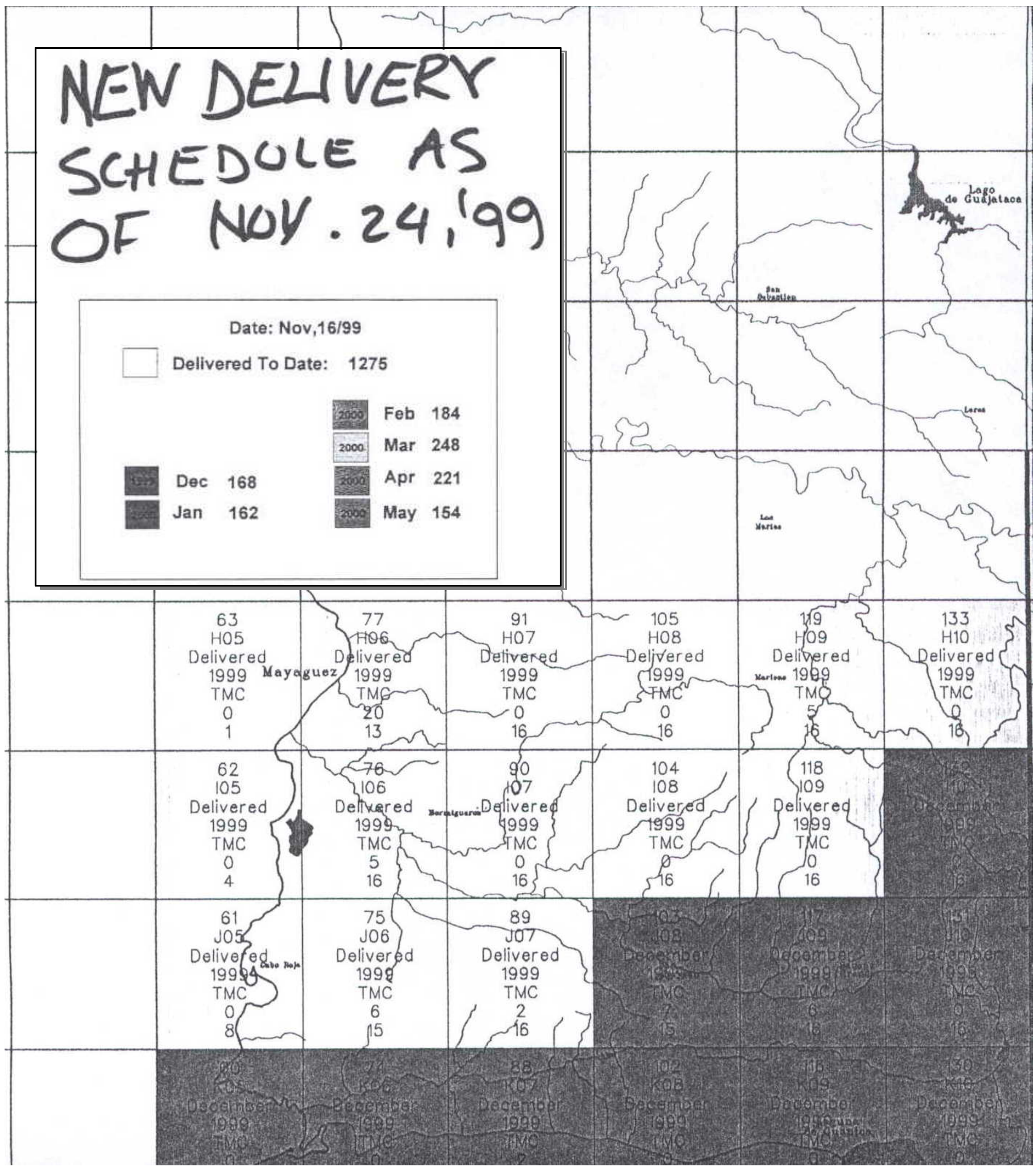
New Nouveau Brunswick

Service New Brunswick

Sample 11 Data quality control status illustrated on an index map (DataQC)



Sample 12 Status map showing project delivery schedules (Terra Surveys)



Sample 13 Batch process for quality assurance adopted by ETB'96

Quality Assurance

Processing

ETB level 1

METHOD:

A list of ETB filenames (ETB1_LIS) is given to batch command procedure ETB1.COM.

A new .ETB file is generated.

Several processing log files are generated.

PROCEDURES:

ETB1.COM
NTXNAM1.COM
NTXNET.COM
NTXSUM.COM
NTXHDR.COM
NTXCHDR.COM

VERIFICATION:

Listing VFY1.LIS is given to VFY1.COM

ARCHIVE LOG:

For Each File:

@ETB_LOG 'filename' 1

Produces: 'filename'.LOG1

Sample 14 Data quality report template used in ETB'96

Sample Quality Control Report

Dates

Contractor ___________ Submitted Returned
Sheet Window #_________
Date Accepted _________

Summary of Errors and Omissions from Supplied Data (per Sect. 3.1.2)

NBDOT

Supplied ETB

NB DNR&E

Other

List of NBGIC Roads Not Present in NBDOT Data Source, and Corrective Action Taken (per Sect. 6.3)

List of Errors in Road/Railroad Network (per Sect.6.6)

List Range of Keys for Road/Railroad Network (per Sect. 6.7)

Lowest = Highest = Total = any Duplicates?

yes__ no__

List of Errors in Hydrographic Network (per Sect. 6.10)

...page 2

Sheet Window #______

List Range of Keys for Hydrographic Network (per Sect. 6.7)

Lowest = Highest = Total = any Duplicates?

yes__ no__

List of Errors in Area Features, by Theme/User Number (per Sect. 6.12)

List of Errors in Hydrography Layer (Including Swamps)(per Sect. 6.13)

Visual Check Against Cartobase (per Sect. 3.7.3)

__ completed, no errors detected

__ completed, errors detected and corrected

File Contains Only Valid CARIS Attributes & Valid combinations of CARIS Attributes

___Yes
___No

Signature of Contractor ________________________

Appendix B Designed Artifacts

This section of the appendix includes the following designed artifacts:

1. Formal descriptions of selected use cases to specification requirements of the collaborative workspace; and
2. Sample WPDL description of the designed workflow model.

USE CASE – Upload Data File

Name	***Upload Data File***
Goal	To check in data files into the central repository
Actor	Client, Contractors and QC Inspector
Scenarios	***Primary Scenario*** 1. Actor browses through local directory and selects data files to upload, and specifies upload type (e.g., For Submission). 2. System displays estimated uploading time and actor goes to upload. 3. System confirms successful uploading right away. 4. System sends a formal confirming email to actor. 5. System assigns version and unlocks data file if it was locked when checked out. ***Alternative***: Preview Selected Data File At step 3, if actor's uploading is not properly received, system sends a message to actor and allows actor to re-upload data files without reselecting them. Then return to step 3. ***Alternative***: Schedule Later Uploading At step 2, after reading estimated time, actor wants to schedule a later time for uploading selected files. 1. System presents scheduler interface. 2. Actor selects specific time and goes for later uploading. 3. Return primary scenario at step 3.
Constraints	1. Later uploading can be scheduled only within 12 hours. 2. Actor does have to zip data files after selection. When actor goes to upload, the system automatically zips all selected data files. 3. Step 3 to 5 can be processed in parallel.

USE CASE – Download Data File

Name	***Download Data File***
Goal	To checkout data files from the central repository
Actor	Client, Contractors and QC Inspector
Scenarios	***Primary Scenario*** 1. Actor browses through catalog and selects data file to download, and specifies download type (e.g., For Inspection) and version numbers. 2. System displays selected data file in central directory and estimated downloading time. 3. Actor goes to download (saves file in local directory). 4. System requests confirmation for successful downloading. 5. System locks downloaded data file and makes appropriate changes of data file attribute information. ***Alternative***: Preview Selected Data File At step 1, actor may want to preview the content of a selected data file. System creates and presents both graphical view of the content and related textual attribute information. ***Alternative***: Schedule Later Downloading At step 2, after reading estimated time, actor wants to schedule a later time for downloading selected files. 1. System presents scheduler interface. 2. Actor selects specific time and goes for later downloading. 3. Return primary scenario at step 3.
Constraints	1. Later downloading can be scheduled only within 12 hours. 2. Actor does have to unzip data files after selection. When actor goes to download, system automatically unzip downloaded data package on local drive.

USE CASE - Query Data File Status

Name	***Query Data File Status***
Goal	To get current status information of data files in process
Actor	Client and project manager
Scenarios	***Primary Scenario*** 5. System presents the query form with default values to form elements leading to query status of all data files. 6. Actor enters query criteria (e.g., data files contracted by contractor XYZ or list of data files that have passed QC inspection). 7. Actor goes to query. 8. System performs query and presents results. ***Alternative***: Print Status Report At step 3, actor may want to print a status report. Actor selects report format options and goes to print. Then use case finishes. ***Alternative***: Query Status Map At step 2, actor may query graphically. System displays a status map showing query results. Actor can pan, zoom, identify, hyperlink, and plot the displayed status map.
Constraints	2. Status information should be updated regularly, normally on daily or weekly basis. However, the time interval depends on individual project implementation.

USE CASE - Monitor Project Progress

Name	***Monitor Project Progress***
Goal	To track and report the overall project progress
Actor	Client and Project Manager
Scenarios	***Primary Scenario*** 1. Actor specifies view options (e.g., completion view or timeline view) and goes to request. 2. System creates and displays specified view of current progress in textual format. ***Alternative***: View and Plot Progress Chart At step 2, actor may want to see a view of progress chart (e.g., Gantt chart). System creates and displays chart. Actor may also plot the chart. ***Alternative***: View Previous Progress Report At step 1, instead of specifying view options for current progress, actor may need to see previous progress report up to a past date.
Extension Points	1. Data File Status Query The actor decides to see the current status for individual data files.
Constraints	2. Rules must be in place to make sure that the progress report is regularly (e.g., weekly or monthly) created and archived into project metadata warehouse by the system.

USE CASE - Participate Project Forum

Name	***Participate Project Forum***
Goal	To discuss project-related issues
Actor	All project participants
Scenarios	***Primary Scenario*** 1. Actor browses through forum message listing and selects interested message title. 2. System displays selected message. ***Alternative***: Post New Message At step 1, instead of selecting any message, actor may want to post a new message. Actor goes to post and fills in content of new message (title, subject, message body, etc.). System confirms receiving of new post. ***Alternative***: Reply Posted Message At step 2, after reading selected message, actor wants to send a reply. Actor goes to reply and composes a reply message. System confirms receiving of reply.
Constraints	1. All replied messages are threaded with the original message. Each message only has one thread. 2. There is no limitation for the number of replied message to one specific message. 3. Posted or replied messages are to be seen on the forum right away.

USE CASE - Markup Map (Data Content)

Name	***Markup Map***
Goal	To graphically discuss or convey problems of data elements
Actor	Contractors, QC Inspector and Project Manager
Scenarios	***Primary Scenario*** *Asynchronous Mode:* 1. Actor opens a graphical view (data image) of data content and reviews any markups and annotations if exist (e.g., marked errors by inspector). 2. Actor adds his/her own markups and annotations using tools provided by system. 3. Actor specifies parties who should see this marked-up data view and goes to save. 4. System saves marked-up data view. 5. System notifies specified parties. *Synchronous Mode:* 1. Actor specifies parties to be involved and calls an electronic map whiteboard session. 2. System establishes sessions among specified parties and presents electronic whiteboard. 3. Actor selects and opens a graphical view of data content. 4. System sends this graphical view to other parties. 5. Actor adds markups or annotations using tools. 6. System updates other party's view in whiteboards. 7. Step 5 and 6 iterate until actor goes to close session. 8. System saves final marked-up data view as image files in the central repository. ***Alternative***: Use Audio/Video Conferencing At step 1 in synchronous mode, actor may also want to invoke an audio/video session. Actor goes to invoke. System establishes audio/video session at step 2.
Constraints	1. Actor should be able to select and open graphical view of data content from both central repository and local machine. 2. In synchronous mode, each session should allow more than two parties to participate. 3. In synchronous mode, system should be able to update all other party's whiteboards in less than 5 (or no worse than 10) seconds, based on one party's inputs.

Sample WPDL Description of the Designed Workflow Model

```
// Workflow Model for Geospatial Data Processing

MODEL                                   'GEODPWM'
     NAME                               "Geospatial Data Processing Workflow Model"
     DESCRIPTION                        "A workflow model for GIS data processing processes"
     WPDL_VERSION                       "Research Prototype"
     CREATED                            September 1st, 2000
     AUTHOR                             "SONGNIAN LI"
     STATUS                             UNDER_REVISION

// Workflow Participant List

   PARTICIPANT              'Client'
            NAME            "Client/Customer"
     DESCRIPTION            "delivers raw data and receives final data product"
            TYPE            ORGANIZATIONAL_UNIT
   END_PARTICIPANT

   PARTICIPANT              'Inspector'
            NAME            "Inspector"
     DESCRIPTION            "inspects the processed data and submits QC reports"
            TYPE            ORGANIZATIONAL_UNIT
   END_PARTICIPANT

   PARTICIPANT              'Production_Contractor'
            NAME            "Production Contractor"
     DESCRIPTION            "processes raw data materials and/or performs internal data Q/C"
            TYPE            ORGANIZATIONAL_UNIT
   END_PARTICIPANT

   PARTICIPANT              'Project_Manager'
            NAME            "Project Manager"
     DESCRIPTION            "manages/controls overall data production project and delivers final
                              data products"
            TYPE            ORGANIZATIONAL_UNIT
   END_PARTICIPANT

// Workflow Application List

   APPLICATION                          'compress_data'
            NAME                        "compress data"
            TOOLNAME                    UNDEFINED
            DESCRIPTION                 "perform data compression before packing it "
            IN_PARAMETERS               'data_file'
            OUT_PARAMETERS              'compressed_data'
   END_APPLICATION

   APPLICATION                          'decompress_data'
            NAME                        "decompress data"
            TOOLNAME                    UNDEFINED
            DESCRIPTION                 "perform data decompression after unpacking data"
```

```
                IN_PARAMETERS       'compressed_data'
                OUT_PARAMETERS      'decompressed_data'
END_APPLICATION

APPLICATION                         'zip_data'
                NAME                "zip data"
                TOOLNAME            UNDEFINED
                DESCRIPTION         "Use ZIP utilities to pack one or more data files"
                IN_PARAMETERS       'data_file'
                OUT_PARAMETERS      'zipped_data'
END_APPLICATION

APPLICATION                         'unzip_data'
                NAME                "unzip data"
                TOOLNAME            UNDEFINED
                DESCRIPTION         "Use WINZIP or other ZIP utilities to unpack a data package"
                IN_PARAMETERS       'zipped_data'
                OUT_PARAMETERS      'data_directory'
END_APPLICATION

APPLICATION                         'upload_data'
                NAME                "upload data"
                TOOLNAME            UNDEFINED
                DESCRIPTION         "Perform data checking-in into the central repository"
                IN_PARAMETERS       'zipped_data'
                                    'upload_type'      // raw delivery or submission/re-submission
                OUT_PARAMETERS      'version_number' // program checks and assigns a number
                                    'packing_list'     // describe proper packing information
END_APPLICATION

APPLICATION                         'download_data'
                NAME                "download data"
                TOOLNAME            UNDEFINED
                DECSRIPTION         "Perform requested data checking-out from the repository"
                IN_PARAMETERS       'zipped_data'
                                    'download_type' // processing, inspection, or final delivery
    END_APPLICATION

APPLICATION                         'validate_data'
                NAME                "validate data"
                TOOLNAME            UNDEFINED
                DESCRIPTION         "Check data package contents and data usability"
                IN_PARAMETERS       'data_directory'
                OUT_PARAMETERS      'validation_report'
                                    'validation_status'          // indicate pass or failed
END_APPLICATION

APPLICATION                         'update_status'
                NAME                "update status"
                TOOLNAME            UNDEFINED
                DESCRIPTION         "Update data processing status in project metadata database"
                IN_PARAMETERS       'data_ID'
                                    'data_status'
END_APPLICATION
APPLICATION                         'unload_data'
```

```
            NAME                    "unload data"
            TOOLNAME                UNDEFINED
            DECSRIPTION             "Perform kind of rollback to clear uploaded data"
            IN_PARAMETERS           'data_directory'
   END_APPLICATION

APPLICATION                         'send_notification'
            NAME                    "send notification"
            TOOLNAME                UNDEFINED              // any built-in email component
            DECSRIPTION             "Send email notification to workflow participants"
   END_APPLICATION
```

// Workflow Process Relevant Data List

```
DATA                                'data_file'
            TYPE                    REFERENCE
            NAME                    "data file"
            DEFAULT_VALUE           ""
            DESCRPTION              "A file containing data covering certain geographic area"
END_DATA

DATA                                'zipped_data'
            TYPE                    REFERENCE
            NAME                    "zipped data"
            DEFAULT_VALUE           ""
            DESCRIPTION             "A ZIP file containing one or more compressed data files"
END_DATA

DATA                                'upload_type'
            TYPE                    STRING
            NAME                    "upload type"
            DEFAULT_VALUE           ""
            DECRIPTION              "A type of Raw Delivery, Submission, or Resubmission"
END_DATA

DATA                                'validation_report'
            TYPE                    REFERENCE
            NAME                    "validation report"
            DEFAULT_VALUE           ""
            DESCRIPTION             "A file containing summary of results from validating data"
END_DATA

DATA                                'validation_status'
            TYPE                    STRING
            NAME                    "validation status"
            DEFAULT_VALUE           "Pass"
            DESCRIPTION             "A STRING variable indicating Pass/Fail of data validation"
END_DATA

DATA                                'download_type'
            TYPE                    STRING
            NAME                    "download type"
            DEFAULT_VALUE           ""
            DESCRIPTION             "A type of For Processing, For Inspection, or Final Delivery"
```

```
END_DATA

DATA                        'data_ID'
        TYPE                STRING
        NAME                "data identifier"
        DEFAULT_VALUE       ""
        DESCRIPTION         "An unique identifier for each data unit "
END_DATA

DATA                        'data_status'
        TYPE                STRING
        NAME                "data status"
        DEFAULT_VALUE       ""
        DESCRIPTION         "A string indicating current production status of a data file"
END_DATA
```

// Workflow Process Definitions
// Process Definition – "From Raw Data to Final Data Product"

```
WORKFLOW                    'GEOPWM_PM_OVERALL'
        NAME                "Overall Process Model"
        DESCRIPTION         "The overall data handling process"
        DURATION_UNIT       DAY
        VERSION             "Research Prototype"
        STATUS              UNDER_REVISION
        CLASSIFICATION      "Production"
        LIMIT               UNDEFINED    // DURATION_UNIT to complete
        DURATION            UNDEFINED    // expected DURATION_UNIT to complete
        WORKING_TIME        UNDEFINED    // actual DURATION_UNIT to complete
```

// Activities within Overall Process Model

```
    ACTIVITY                        'Raw_Data_Delivery'
            NAME                    "Raw Data Delivery"
            DESCRIPTION             "raw data material delivery and validation"
            IMPLEMEMTATION          WORKFLOW SYNCHR
    END_ACTIVITY

    ACTIVITY                        'Raw_Data_Processing'
            NAME                    "Raw Data Processing"
            DESCRIPTION             "data processing and submits processed data"
            IMPLEMEMTATION          WORKFLOW SYNCHR
    END_ACTIVITY

    ACTIVITY                        'Data_Inspection'
            NAME                    "Processed Data Inspection"
            DESCRPTION              "A sub flow deals with data quality control activities"
            IMPLEMEMTATION          WORKFLOW SYNCHR
            XOR JOIN                T0_2, T0_5
            XOR SPLIT               T0_3, T0_4
    END_ACTIVITY
    ACTIVITY                        'Failed_Data_Reworking'
```

```
                NAME                 “Failed Data Reworking”
                DESCRIPTION          “processing data failed to pass QC inspection”
                IMPLEMEMTATION       WORKFLOW SYNCHR
  END_ACTIVITY

  ACTIVITY                           ‘Final_Data_Delivery’
                NAME                 “Final Data Delivery”
                DESCRIPTION          “delivering final data to client/customers”
                IMPLEMEMTATION       WORKFLOW SYNCHR
  END_ACTIVITY
```

// Transition Information for the Overall Process Model

```
        TRANSITION                   ‘T0_1’
                FROM                 ‘Raw_Data_Delivery’
                TO                   ‘Raw_Data_Processing’
                CONDITION
        END_TRANSITION

        TRANSITION                   ‘T0_2’
                FROM                 ‘Raw_Data_Processing’
                TO                   ‘Data_Inspection’
                CONDITION
        END_TRANSITION

        TRANSITION                   ‘T0_3’
                FROM                 ‘Data_Inspection’
                TO                   ‘Final_Data_Delivery’
                CONDITION            ‘QC_result’ = “Pass”
        END_TRANSITION

        TRANSITION                   ‘T0_4’
                FROM                 ‘Data_Inspection’
                TO                   ‘Failed_Data_Reworking’
                CONDITION            ‘QC_result’ = “Fail”
        END_TRANSITION

        TRANSITION                   ‘T0_5’
                FROM                 ‘Failed_Data_Re working’
                TO                   ‘Data_Inspection’
                CONDITION
        END_TRANSITION
```

// Workflow Relevant Data for Overall Process Model

```
  DATA                               ‘QC_result’
                TYPE                 STRING
                NAME                 “QC result”
                DEFAULT_VALUE        “Pass”
                DESCRIPTION          “A STRING variable indicating Pass/Fail of QC”
  END_DATA
END_WORKFOW                 // GEOPWM_PM_OVERALL
```

// Process Definition – “Raw Data Delivery”

```
WORKFLOW                    ‘GEOPWM_PM_RDD’
        NAME                “Raw Data Delivery Process Model”
        DESCRIPTION         “process handles raw data material delivery and validation”
        DURATION_UNIT       DAY
        VERSION             “Research Prototype”
        STATUS              UNDER_REVISION
        CLASSIFICATION      “Production”
        LIMIT               UNDEFINED     // DURATION_UNIT to complete
        DURATION            UNDEFINED     // expected DURATION_UNIT to complete
        WORKING_TIME        UNDEFINED     // actual DURATION_UNIT to complete
```

// Activities within Raw Data Delivery Process Model

```
ACTIVITY                    ‘Raw_Data_Packing’
        NAME                “Raw Data Packing”
        DESCRIPTION         “Pack data files with predefined directory structure”
        IMPLEMEMTATION      APPLICATION
                              TOOL_LIST
                                  ‘compress_data’
                                  ‘zip_data’
                            END_TOOL_LIST
        PERFORMER           ‘Client’
        START_MODE          MANUAL
        FINISH_MODE         AUTOMATIC
        DURATION            UNDEFINED
        WORKING_TIME        UNDEFINED
END_ACTIVITY

ACTIVITY                    ‘Raw_Data_Uploading’
        NAME                “Raw Data Uploading”
        DESCRIPTION         “Transfer, unpack and validate”
        IMPLEMEMTATION      APPLICATION
                              TOOL_LIST
                                  ‘upload_data’
                                  ‘unzip_data’
                                  ‘validate_data’
                              END_TOOL_LIST
        PERFORMER           SYSTEM                      // resources
        START_MODE          AUTOMATIC
        FINISH_MODE         AUTOMATIC
        DURATION            UNDEFINED
        WORKING_TIME        UNDEFINED
END_ACTIVITY

ACTIVITY                    ‘Raw_Delivery_Approval’
        NAME                “Raw Delivery Approval”
        DESCRIPTION         “Review validation and decide if raw data should be relayed”
        IMPLEMEMTATION      NO
        PERFORMER           ‘Project Manager’
        START_MODE          AUTOMATIC
        FINISH_MODE         MANUAL
        DURATION            UNDEFINED
```

```
            WORKING_TIME        UNDEFINED
            XOR SPLIT           T1_3, END_WORKFLOW
      END_ACTIVITY

      ACTIVITY                  'Failed_Raw_Delivery_Handling'
            NAME                "Failed Raw Delivery Handling"
            DESCRIPTION         "Handle the situation where raw data delivery failed"
            IMPLEMEMTATION      APPLICATION
                                  TOOL_LIST
                                       'unload_data'
                                       'send_notification'
                                  END_TOOL_LIST
                                  PROCEDURE_LIST
                                       ABORT_PROCESS
                                  END_PROCEDURE_LIST
            PERFORMER           SYSTEM | "Project Manager"
            START_MODE          AUTOMATIC
            FINISH_MODE         MANUAL
            DURATION            UNDEFINED
            WORKING_TIME        UNDEFINED
      END_ACTIVITY
```

// Transition Information for Raw Data Delivery Process Model

```
            TRANSITION          'T1_1'
            FROM                'Raw_Data_Packing'
            TO                  'Raw_Data_Uploading'
            CONDITION
     END_TRANSITION

            TRANSITION          'T1_2'
            FROM                'Raw_Data_Uploading'
            TO                  'Raw_Delivery_Approval'
            CONDITION
     END_TRANSITION

            TRANSITION          't1_3'
            FROM                'Raw_Data_Approval'
            TO                  'Failed_Raw_Delivery_Handling'
            CONDITION           REJECTED
     END_TRANSITION
  END_WORKFLOW             // GEOPWM_PM_RDD
```

// Process Definition – "Raw Data Processing"

```
  WORKFLOW                      'GEOPWM_PM_RDP'
            NAME                "Raw Data Processing Process Model"
            DESCRIPTION         "processes raw data materials and submits processed data"
            DURATION_UNIT       DAY
            VERSION             "Research Prototype"
            STATUS              UNDER_REVISION
            CLASSIFICATION      "Production"
            LIMIT               UNDEFINED      // DURATION_UNIT to complete
            DURATION            UNDEFINED      // expected DURATION_UNIT to complete
```

```
        WORKING_TIME          UNDEFINED      // actual DURATION_UNIT to complete
```

// Activities within Raw Data Delivery Process Model

```
ACTIVITY                      'Raw_Data_Receiving'
        NAME                  "Raw Data Receiving"
        DESCRIPTION           "Download raw materials into directory structure"
        IMPLEMEMTATION        APPLICATIONS
                                      TOOL_LIST
                                              'download_data'
                                              'unzip_data'
                                              'decompress_data'
                                              'validate_data'
                                      END_TOOL_LIST
        PERFORMER              "Shipment_Staff"
        START_MODE            AUTOMATIC
        FINISH_MODE           MANUAL
        DURATION              UNDEFINED
        WORKING_TIME          UNDEFINED
        XOR JOIN              START, T2_3
END_ACTIVITY

ACTIVITY                      'Check_Receivings'
        NAME                  "Check Receivings"
        DESCRIPTION           "Check downloaded materials – completeness and usability"
        IMPLEMENTATION        NO
        PERFORMER              "Shipment_Staff"
        START_MODE            AUTOMATIC
        FINISH_MODE           MANUAL
        DURATION              UNDEFINED
        WORKING_TIME          UNDEFINED
        XOR SPLIT             T2_2, T2_3
END_ACTIVITY

ACTIVITY                      'Internal_Data_Processing'
        NAME                  "Internal Inspection"
        DESCRIPTION           "A sub flow handles actual raw data processing"
        IMPLEMEMTATION        WORKFLOW SYNCHR   // at the production contractor
        START_MODE            AUTOMATIC
        FINISH_MODE           AUTOMATIC
        DURATION              UNDEFINED
        WORKING_TIME          UNDEFINED
        XOR JOIN
END_ACTIVITY

ACTIVITY                      'Internal_Inspection'
        NAME                  "Internal Inspection"
        DESCRIPTION           "performs internal quality control of processed data"
        IMPLEMEMTATION        WORKFLOW SYNCHR   // at the production contractor
        START_MODE            AUTOMATIC
        FINISH_MODE           AUTOMATIC
        DURATION              UNDEFINED
        WORKING_TIME          UNDEFINED
        XOR SPLIT
```

```
  END_ACTIVITY

  ACTIVITY                    'Data_Submission'
      NAME                    "Data Submission"
      DESCRIPTION             "Submit processed data for inspection"
      IMPLEMEMTATION          APPLICATIONS
                                  TOOL_LIST
                                      'compress_data'
                                      'zip_data'
                                      'upload_data'
                                      'validate_data'          // server-side
                                  END_TOOL_LIST
      PERFORMER               'Shipment_Staff'
  END_ACTIVITY
```

// Transition Information for Raw Data Processing Process Model

```
TRANSITION                    'T2_1'
      FROM                    'Raw_Data_Receiving'
      TO                      'Check_Receivings'
      CONDITION
  END_TRANSITION

TRANSITION                    'T2_2'
      FROM                    'Check_Receivings'
      TO                      'Internal_Data_Processing'
      CONDITION               COMPLETE
  END_TRANSITION

TRANSITION                    'T2_3'
      FROM                    'Check_Receivings'
      TO                      'Raw_Data_Receiving'
      CONDITION               INCOMPLETE
  END_TRANSITION

TRANSITION                    'T2_4'
      FROM                    'Internal_Data_Processing'
      TO                      'Internal_Inspection'
      CONDITION
  END_TRANSITION

TRANSITION                    'T2_5'
      FROM                    'Internal_Inspection'
      TO                      'Data_Submission'
      CONDITION               'QC_result' = "Pass"
  END_TRANSITION

TRANSITION                    'T2_6'
      FROM                    'Internal_Inspection'
      TO                      'Internal_Data_Processing'
      CONDITION               'QC_result' = "Fail"
  END_TRANSITION
```

```
    // Workflow Participant within Raw Data Processing Process Model

      PARTICIPANT                 'Shipment_Staff'
            NAME                  "Shipment Staff"
            DESCRIPTION           "Handle incoming data at the Primary Contractor"
            TYPE                  ROLE
      END_PARTICIPANT

  END_WORKFLOW          // GEOPWM_PM_RDP

// Process Definition – "Data Inspection"

  WORKFLOW                        'GEOPWM_PM_DI'
            NAME                  "Data Inspection Process Model"
            DESCRIPTION           "The process handles quality control of processed data"
            DURATION_UNIT         DAY
            VERSION               "Research Prototype"
            STATUS                UNDER_REVISION
            CLASSIFICATION        "Production"
            LIMIT                 UNDEFINED      // specified DURATION_UNIT to complete
            DURATION              UNDEFINED      // expected DURATION_UNIT to complete
            WORKING_TIME          UNDEFINED      // actual DURATION_UNIT to complete

      // Activities within Data Inspection Process Model

      ACTIVITY                    'Processed_Data_Receiving'
            NAME                  "Processed Data Receiving"
            DESCRIPTION           "Download processed data for inspection"
            IMPLEMEMTATION        APPLICATIONS
                                        TOOL_LIST
                                              'download_data'
                                              'unzip_data'
                                              'decompress_data'
                                              'validate_data'
                                        END_TOOL_LIST
            PERFORMER              "Shipment_Staff"
            START_MODE            AUTOMATIC
            FINISH_MODE           MANUAL
            DURATION              UNDEFINED
            WORKING_TIME          UNDEFINED
            XOR JOIN              START, T3_3
      END_ACTIVITY

      ACTIVITY                    'Check_Receivings'
            NAME                  "Check Receivings"
            DESCRIPTION           "Check downloaded materials – completeness and usability"
            IMPLEMENTATION        NO
            PERFORMER              "Shipment_Staff"
            START_MODE            AUTOMATIC
            FINISH_MODE           MANUAL
            DURATION              UNDEFINED
            WORKING_TIME          UNDEFINED
            XOR SPLIT             T3_2, T3_3
```

```
  END_ACTIVITY

  ACTIVITY                    'Data_Inspecting'
        NAME                  "Data Inspecting"
        DESCRIPTION           "A sub flow performs data quality control"
        IMPLEMEMTATION        WORKFLOW SYNCHR            // at the inspector
        START_MODE            AUTOMATIC
        FINISH_MODE           AUTOMATIC
        DURATION              UNDEFINED
        WORKING_TIME          UNDEFINED
  END_ACTIVITY

  ACTIVITY                    'Inspection_Reporting'
        NAME                  "Inspection Reporting"
        DESCRIPTION           "Review inspection report/marked data file and send report"
        IMPLEMEMTATION        APPLICATIONS
                                    TOOL_LIST
                                          'review_report'   // inspection results
                                          'send_report'
                                    END_TOOL_LIST
        PERFORMER             'Inspection_Manager'
        START_MODE            AUTOMATIC
        FINISH_MODE           MANUAL
        DURATION              UNDEFINED
        WORKING_TIME          UNDEFINED
  END_ACTIVITY
```

// Transition Information for Data Inspection Process Model

```
TRANSITION                    'T3_1'
        FROM                  'Processed_Data_Receiving'
        TO                    'Check_Receivings'
        CONDITION
  END_TRANSITION

TRANSITION                    'T3_2'
        FROM                  'Check_Receivings'
        TO                    'Data_Inspecting'
        CONDITION             COMPLETE
  END_TRANSITION

TRANSITION                    'T3_3'
        FROM                  'Check_Receivings'
        TO                    'Processed_Data_Receiving'
        CONDITION             INCOMPLETE
  END_TRANSITION

TRANSITION                    'T3_4'
        FROM                  'Data_Inspecting'
        TO                    'Inspection_Reporting'
        CONDITION
  END_TRANSITION
```

```
// Workflow Participant within Data Inspection Process Model

PARTICIPANT            'Shipment_Staff'
    NAME               "Shipment Staff"
    DESCRIPTION        "Handle incoming data at the Inspector"
    TYPE               ROLE
END_PARTICIPANT

PARTICIPANT            'Inspection_Manager'
    NAME               "Inspection Manager"
    DESCRIPTION        "Manage inspection results reporting at the Inspector"
    TYPE               ROLE
END_PARTICIPANT

// Workflow Application within Data Inspection Process Model

APPLICATION            'review_report'
    NAME               "review report"
    TOOLNAME           UNDEFINED      // web browser or spreadsheet
    DESCRIPTION        "Review the inspection report and marked data files"
END_APPLICATION

APPLICATION            'send_report'
    NAME               "send_report"
    TOOLNAME           UNDEFINED            // FTP or email attachment
    DESCRIPTION        "Submit inspection report to project manager"
END_APPLICATION

END_WORKFLOW          // GEOPWM_PM_DI

// Process Definition – "Failed Data Reworking"

WORKFLOW               'GEOPWM_PM_FDR'
    NAME               "Failed Data Reworking Process Model"
    DESCRIPTION        "The process reprocesses data failed to pass inspection"
    DURATION_UNIT      DAY
    VERSION            "Research Prototype"
    STATUS             UNDER_REVISION
    CLASSIFICATION     "Production"
    LIMIT              UNDEFINED      // specified DURATION_UNIT to complete
    DURATION           UNDEFINED      // expected DURATION_UNIT to complete
    WORKING_TIME       UNDEFINED      // actual DURATION_UNIT to complete

// Activities within Failed Data Reworking Process Model

ACTIVITY               'QC_Report_Receiving'
    NAME               "QC Report Receiving"
    DESCRIPTION        "Get QC reports and included marked data for reworking"
    IMPLEMEMTATION     APPLICATIONS
                           TOOL_LIST
                               'download_report'
                               'identify_ambiguity'
                           END_TOOL_LIST
```

```
        PERFORMER               'Shipment_Staff"
        START_MODE              AUTOMATIC
        FINISH_MODE             MANUAL
        DURATION                UNDEFINED
        WORKING_TIME            UNDEFINED
    END_ACTIVITY

    ACTIVITY                    'Data_Correcting'
        NAME                    "Internal Correcting"
        DESCRIPTION             "Correct all marked or reported errors in processed data"
        IMPLEMEMTATION          NO
        PERFORMER               'Production_Group'
        START_MODE              MANUAL
        FINISH_MODE             MANUAL
        DURATION                UNDEFINED
        WORKING_TIME            UNDEFINED
        XOR JOIN                T4_1, T4_4
    END_ACTIVITY

    ACTIVITY                    'Internal_Inspection'
        NAME                    "Internal Inspection"
        DESCRIPTION             "A sub flow performs internal quality control"
        IMPLEMEMTATION          WORKFLOW SYNCHR   // at the production contractor
        START_MODE              AUTOMATIC
        FINISH_MODE             AUTOMATIC
        DURATION                UNDEFINED
        WORKING_TIME            UNDEFINED
        XOR SPLIT               T4_3, T4_4
    END_ACTIVITY

    ACTIVITY                    'Data_Resubmission'
        NAME                    "Data Resubmission"
        DESCRIPTION             "Submit corrected data files"
        IMPLEMEMTATION          APPLICATIONS
                                        TOOL_LIST
                                                'compress_data'
                                                'zip_data'
                                                'upload_data'
                                                'validate_data'          // server-side
                                        END_TOOL_LIST
        PERFORMER               'Shipment_Staff'
        START_MODE              AUTOMATIC
        FINISH_MODE             MANUAL
        DURATION                UNDEFINED
        WORKING_TIME            UNDEFINED
    END_ACTIVITY
```

// Transition Information for Failed Data Reworking Process Model

```
  TRANSITION                    'T4_1'
        FROM                    'Inspection_Report_Receiving'
        TO                      'Data_Correcting'
        CONDITION
    END_TRANSITION
```

```
    TRANSITION              'T4_2'
        FROM                'Data_Correcting'
        TO                  'Internal_Inspection'
        CONDITION
    END_TRANSITION

    TRANSITION              'T4_3'
        FROM                'Internal_Inspection'
        TO                  'Data_Resubmission'
        CONDITION           'QC_result' = "Pass"
    END_TRANSITION

    TRANSITION              'T4_4'
        FROM                'Intemal_Inspection'
        TO                  'Data_Correcting'
        CONDITION           'QC_result' = "Fail"
    END_TRANSITION

    // Workflow Participant within Failed Data Reworking Process Model

    PARTICIPANT             'Shipment_Staff'
        NAME                "Shipment Staff"
        DESCRIPTION         "Handle incoming data at the Primary Contractor"
        TYPE                ROLE
    END_PARTICIPANT

    PARTICIPANT             'Production_Group'
        NAME                "Production Group"
        DESCRIPTION         "Process raw data materials at the production contractor"
        TYPE                ORGANIZATIONAL_UNIT
    END_PARTICIPANT

    PARTICIPANT             'Inspection_Group'
        NAME                "Inspection Group"
        DESCRIPTION         "Perform internal data Q/C at the production contractor"
        TYPE                ORGANIZATIONAL_UNIT
    END_PARTICIPANT

    // Workflow Application within Failed Data Reworking Process Model

    APPLICATION             'download_report'
        NAME                "download report"
        TOOLNAME            UNDEFINED          // web browser (FTP or HTTP)
        DESCRIPTION         "Download the inspection report and marked data files"
    END_APPLICATION

    APPLICATION             'identify_ambiguity'
        NAME                "identify ambiguity"
        TOOLNAME            UNDEFINED          // GIS
        DESCRIPTION         "Identify ambiguous markups, annotations and errors"
    END_APPLICATION
END_WORKFLOW            // GEOPWM_PM_FDR
```

// Process Definition – "Final Data Delivery"

```
WORKFLOW                    'GEOPWM_PM_FDD'
        NAME                "Raw Data Delivery Process Model"
        DESCRIPTION         "process handles raw data material delivery and validation"
        DURATION_UNIT       DAY
        VERSION             "Research Prototype"
        STATUS              UNDER_REVISION
        CLASSIFICATION      "Production"
        LIMIT               UNDEFINED     // specified DURATION_UNIT to complete
        DURATION            UNDEFINED     // expected DURATION_UNIT to complete
        WORKING_TIME        UNDEFINED     // actual DURATION_UNIT to complete
```

// Activities within Final Data Delivery Process Model

```
    ACTIVITY                'Final_Data_Approval'
        NAME                "Final Data Approval"
        DESCRIPTION         "Approve the final data delivery"
        IMPLEMEMTATION      NO
        PERFORMER           "Project Manager"
        START_MODE          MANUAL
        FINISH_MODE         MANUAL
        DURATION            UNDEFINED
        WORKING_TIME        UNDEFINED
    END_ACTIVITY

    ACTIVITY                'Notify_Client'
        NAME                "Notify Client"
        DESCRIPTION         "Notify client of the availability of final data products"
        IMPLEMEMTATION      APPLICATION
                                  TOOL_LIST
                                        'send_notification'
                                  END_TOOL_LIST
        PERFORMER           SYSTEM
        START_MODE          AUTOMATIC
        FINISH_MODE         AUTOMATIC
        DURATION            UNDEFINED
        WORKING_TIME        UNDEFINED
    END_ACTIVITY

    ACTIVITY                'Final_Data_Receiving'
        NAME                "Final Data Receiving"
        DESCRIPTION         "Download final data files and related report"
        IMPLEMEMTATION      APPLICATIONS
                                  TOOL_LIST
                                        'download_data'
                                        'unzip_data'
                                        'decompress_data'
                                        'validate_data'
                                  END_TOOL_LIST
        PERFORMER           "Client"
        START_MODE          AUTOMATIC
        FINISH_MODE         MANUAL
        DURATION            UNDEFINED
```

```
        WORKING_TIME        UNDEFINED
    END_ACTIVITY

    // Transition Information for Final Data Delivery Process Model

   TRANSITION               'T5_1'
        FROM                'Final_Data_Approval'
        TO                  'Notify_Client'
        CONDITION           APPROVED
    END_TRANSITION

   TRANSITION               'T5_2'
        FROM                'Notify_Client'
        TO                  'Final_Data_Receiving'
        CONDITION
    END_TRANSITION

   END_WORKFLOW             // GEOPWM_PM_FDD
END_MODEL
```

Appendix C Sample Code of Programs Developed

The VB code included here was used for performance testing of data file transfer. It consists of five segments (four sub-procedures): (1) defining "global" variables; (2) loading program; (3) quitting program; (4) testing; and (5) tracing file transfer (downloading) status. A brief explanation for each segment is included starting with double backslashes.

```
// All global variables are defined in this segment.  These variables are visible in every
// sub procedures

Option Explicit
Private ftpErr As Boolean
Private f_time As Date
Private f_name As String
Private errMess As String
Private intCtrlState As Boolean

// This procedure is called when a "Quite Program" request is received.  The program
// simply closes itself.

Private Sub cmdQuite_Click()
   End
End Sub

// This procedure is called when a "Get List of Files" request is received.  The program
// uses "Execute" method the Internet Control to get a name list of the files stored on the
// server and add them into a ListBox control by calling another sub procedure
// "Inet_StateChanged"

Private Sub cmdGet_Click()
   ftpErr = False
   intCtrlState = True

   With InetUpload
      .AccessType = icDirect  'value is 1
      .Protocol = icFTP   'value is 2
      .URL = "FTP://131.202.134.29"
      .RequestTimeout = 30
```

```
        f_name = "Get the list of files on FTP server"
        .Execute , "LS " & "/Public/"
        If ftpErr Then
           MsgBox errMess
           GoTo cancel
        End If

        'wait until job done
waitftp:
        DoEvents
        If .StillExecuting Then
           GoTo waitftp
        End If
cancel:
        .Execute , "close"
    End With
    'MsgBox fileLst.ListCount
End Sub
```

// This procedure is called when a “Testing” request is received. The program uses VB
// Internet Control and calls its “Execute” method to fetch all data files stored on the
// server to the local testing machine, using a For-Next loop structure. The
// “.StillExecuting” method of the Internet Control is used to ensure that every fetch is
// completed before looping to another.

```
Private Sub cmdTest_Click()
    Dim i As Integer

    ftpErr = False
    intCtrlState = False

    With Inet
        .AccessType = icDirect  'value is 1
        .Protocol = icFTP   'value is 2
        .URL = "FTP://131.202.134.29"
        .RequestTimeout = 30
        .UserName = "XXXXX"
        .Password = "XXXX"
        For i = 0 To fileLst.ListCount - 1
           f_name = fileLst.List(i)
           .Execute , "get " & "\Public\" & fileLst.List(i) & " " & App.Path & "\Temp\" &
fileLst.List(i)
           If ftpErr Then
              MsgBox errMess
              GoTo cancel
```

```
        End If

        'wait until downloading job done
waitftp:
        DoEvents
        If .StillExecuting Then
           GoTo waitftp
        End If
      Next
cancel:
      .Execute , "close"
      MsgBox "Testing task has been successfully completed!"
   End With
End Sub
```

// This procedure is called when the testing program is first started. Using VB File
// System object, the program creates an empty text file to store testing data during the
// testing session. The file overwrites itself every time the procedure is called.

```
Private Sub Form_Load()
   Dim fs As Scripting.FileSystemObject
   Dim a As TextStream

   Set fs = CreateObject("Scripting.FileSystemObject")
   Set a = fs.CreateTextFile(App.Path & "\report.txt", True)
   a.Close
End Sub
```

// The procedure is called when the “Execute” method of the VB Internet Control is
// called. The procedure traces three states of the execution session: icRequestSent,
// icResponseCompleted, and icError. The start time of the data file downloading is
// recorded when icRequestSent is detected and the end time is recorded when the
// icResponseCompleted is detected.

```
Private Sub Inet_StateChanged(ByVal State As Integer)
   Select Case State
   ' other case statements here, if needed
   Case icRequestSent
      f_time = Time
   Case icResponseCompleted
      Dim fs As Scripting.FileSystemObject
      Dim a As TextStream
      Set fs = CreateObject("Scripting.FileSystemObject")
      Set a = fs.OpenTextFile(App.Path & "\report.txt", ForAppending, True)
      a.WriteLine (f_name & "," & Date & "," & f_time & "," & Time)
```

```
        a.Close

        If intCtrlState = True Then
            ' make sure the case only executed when click on "Get File List" by using
intCtrlState
            Dim strData As String
            Dim vtData As Variant
            Dim bDone As Boolean: bDone = False

            Do While Not bDone
                ' Get chunk.
                vtData = Inet.GetChunk(1024, icString)
                strData = strData & vtData
                DoEvents

                If Len(vtData) = 0 Then
                    bDone = True
                End If
            Loop

            'the delimiter between file names is a set of "carriage return" plus "line feed"
            Dim inCounter, inFoundPos, i As Integer

            inCounter = 1
            inFoundPos = InStr(inCounter, strData, Chr(10))
            i = 0
            While inFoundPos <> 0
                fileLst.List(i) = Mid$(strData, inCounter, inFoundPos - inCounter - 1)
                inCounter = inFoundPos + 1
                inFoundPos = InStr(inCounter, strData, Chr(10))
                'the reason for this if-then is that strData has two sets of "carriage return" plus
"line feed"
                If inFoundPos = Len(strData) Then
                    inFoundPos = 0
                End If
                i = i + 1
            Wend

        End If
    Case icError ' value is 11
        ftpErr = True
        errMess = "ErrorCode: " & Inet.ResponseCode & " : " & Inet.ResponseInfo
    End Select
End Sub
```

Appendix D Testing and Survey Results

This section of the appendix includes the recorded results of the FTP testing and sample results of surveys on some performance measures of the existing GIS data production projects:

1. Recorded time duration of sixteen FTP tests for thirty data files;
2. Performance questionnaire of ETB'96 project; and
3. Email response to the telephone survey on the current use of the Internet for data submission.

Recorded Time Duration of Sixteen FTP Tests (1-8) for Thirty Data Files

File	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8
f1	0:02	0:06	0:07	0:07	0:17	0:32	0:13	0:04
f2	0:06	0:06	0:11	0:08	0:15	0:23	0:08	0:04
f3	0:06	0:05	0:11	0:12	0:17	0:34	0:11	0:04
f4	0:07	0:06	0:07	0:10	0:20	0:24	0:07	0:05
f5	0:05	0:07	0:10	0:08	0:23	0:23	0:09	0:05
f6	0:11	0:10	0:21	0:17	0:36	0:38	0:12	0:09
f7	0:15	0:14	0:18	0:15	0:34	0:40	0:14	0:10
f8	0:11	0:11	0:15	0:15	0:36	0:48	0:14	0:08
f9	0:12	0:13	-	0:12	0:38	0:50	0:16	0:09
f10	26:07	24:33	-	39:28	76:03	147:58	26:35	21:51
f11	21:04	17:45	-	41:43	17:34	-	22:51	24:55
f12	20:11	24:23	-	32:18	16:35	-	22:00	-
f13	0:18	0:17	-	0:12	0:37	-	0:11	-
f14	6:34	8:54	-	10:55	28:53	-	6:36	-
f15	20:23	34:19	-	28:04	15:56	-	22:47	-
f16	3:13	5:30	-	4:44	13:24	-	3:19	-
f17	1:50	-	-	2:38	7:15	-	1:49	-
f18	3:34	-	-	5:48	13:59	-	3:36	-
f19	8:29	-	-	4:20	35:06	-	7:50	-
f20	7:34	-	-	10:34	29:44	-	6:56	-
f21	3:55	-	-	5:39	19:01	-	4:15	-
f22	7:07	-	-	10:05	31:42	-	6:23	-
f23	21:41	-	-	26:41	34:07	-	22:26	-
f24	9:31	-	-	11:47	-	-	10:16	-
f25	3:05	-	-	3:50	-	-	3:10	-
f26	2:58	-	-	3:44	-	-	3:04	-
f27	10:26	-	-	12:51	-	-	10:58	-
f28	1:28	-	-	1:50	-	-	2:26	-
f29	3:1	-	-	3:27	-	-	5:00	-
f30	1:32	-	-	1:51	-	-	2:38	-

Recorded Time Duration of Sixteen FTP Tests (9-16) for Thirty Data Files

File	Test 9	Test 10	Test 11	Test 12	Test 13	Test 14	Test 15	Test 16
f1	0:07	0:14	0:09	0:07	0:06	0:33	0:43	0:47
f2	0:18	0:18	0:08	0:09	0:04	0:31	0:26	0:48
f3	0:14	0:25	0:08	0:08	0:05	0:31	0:35	1:03
f4	0:19	0:23	0:12	0:06	0:06	0:48	0:43	0:57
f5	0:17	0:26	0:10	0:09	0:05	0:41	0:50	1:03
f6	0:40	0:40	0:20	0:12	0:10	1:17	2:26	1:57
f7	0:47	0:39	0:11	0:12	0:09	1:18	1:36	1:58
f8	0:35	0:28	0:08	0:21	0:10	1:16	1:23	2:06
f9	0:34	0:33	0:15	0:12	0:11	1:15	1:07	2:48
f10	59:44	52:38	23:13	28:24	20:51	189:36	198:52	244:49
f11	58:01	49:34	19:58	28:48	-	-	-	-
f12	55:35	49:01	19:57	25:20	-	-	-	-
f13	0:10	0:13	0:10	0:16	-	-	-	-
f14	8:40	14:38	6:23	7:50	-	-	-	-
f15	32:54	42:49	17:09	28:04	-	-	-	-
f16	10:29	8:31	3:21	4:48	-	-	-	-
f17	5:30	4:25	2:01	2:09	-	-	-	-
f18	8:03	8:18	3:49	4:55	-	-	-	-
f19	16:59	17:24	7:45	10:01	-	-	-	-
f20	9:46	19:09	7:02	8:31	-	-	-	-
f21	5:50	11:15	4:44	4:42	-	-	-	-
f22	10:01	11:50	6:10	8:00	-	-	-	-
f23	26:23	39:55	24:40	28:27	-	-	-	-
f24	12:06	26:15	9:36	13:23	-	-	-	-
f25	3:51	6:11	3:25	4:40	-	-	-	-
f26	3:21	6:09	3:03	3:21	-	-	-	-
f27	12:07	22:45	10:41	-	-	-	-	-
f28	2:29	2:22	1:26	-	-	-	-	-
f29	3:26	6:20	2:59	-	-	-	-	-
f30	2:32	2:27	1:35	-	-	-	-	-

Performance questionnaire of ETB'96 project
(Provided by Mark Doucette)

The estimations of the performance factors listed in the following table are based on one shipment which contains 100 (window) ETB data files.

Task	Number of People	Duration (hours)	Elapsed Time (day)	Cost ($)
Data File Packing for Shipping				
Data File Shipping	1	3	1	400
Shipment Receiving and Checking	1	2	0.5	400
Weekly QC Reporting from Inspector - status for individual data files - status per QC level				
Weekly Reporting to Client - status for individual data files - status per QC level	1	1	0.5	400
Running All Automatic QC Programs	1	8	1	400
Data Files Packing for Returning	1	1	0.5	200
Preparing Final QC Report	2	8	1	400
Packing for Final Data Delivery	1	3	0.5	200
Total Cost for Communications Related to the Shipment - paper consumed, postage, and telephone/fax charges - manpower needed				
Hardcopy Map Plotting	1	3	0.5	200

The average number of resubmissions of data files for QC inspections: 20%

The number of hardcopy plots required: 1 per file x 100 x 120% = 120

Email Response to the Telephone Survey on Using the Internet for data submission (Chris Roberts, DataQC, Inc.)

Subject: Timings on Web QC
Date: Fri, 15 Mar 2002 08:18:58 -0400
From: chris.roberts@dataqc.com
To: snli@ryerson.ca
CC: mark.doucette@dataqc.com

As per our telephone call here are some times:

1. Time to upload a shipment of 50 files - 10 minutes
2. Time to run through QC routines 2.5 hours

No human hands touch the files.

Replaces the time to:

1. create a CD (contractor)
2. open delivery package
3. check contents
4. load CD
5. build batch files
6. run batch process
7. send results

Guessing it would be at-least 2 hrs of operator time -- 2 @ $45=90.00

Hope this helps

Chris Roberts, DataQC

Chris Roberts, DataQC Inc.

301 Woodstock Road, Fredericton, N.B Canada E3B 2H9
tel: 506 444 8142
fax: 506 444 8125
email: chris.roberts@dataqc.com
website : www.dataqc.com

CURRICULUM VITAE

Candidate's full name: Songnian Li

Universities attended:

Wuhan Technical University of Surveying and Mapping
Wuhan, Hubei, P. R. China
1979 – 1983
B. Eng. (1983)

University of New Brunswick
Fredericton, N.B., Canada
1997 – 2002
Ph. D. Candidate

Publications (from 1997 to 2002):

Papers in Refereed Publications

Coleman, D. J. and S. Li (1999) "Developing a groupware-based prototype to support geomatics production management", *Computers, Environment and Urban Systems*, 23(4), pp. 1 – 17

Easa, S. M., S. Li, and Y. Shi (1998) "GIS technology for civil engineering education", *Journal of Professional issues in Engineering Education and Practic*e, April, 124(2), pp. 40 – 47

Book Chapters

Li, S., D. J. Coleman, and S. M. Easa (1999) "Public utilities" in *Urban Planning and Development Applications of GIS*, Said Easa and Yupo Chan, eds., American Society of Civil Engineers, Reston, Virginia, Chapter 8, pp. 135 – 150.

Easa, S. M., S. Li, and Y. Shi (1999) "Civil engineering education" in *Urban Planning and Development Applications of GIS*, Said Easa and Yupo Chan, eds., American Society of Civil Engineers, Reston, Virginia, Chapter 14, pp. 241 – 257

Papers in Refereed Conference Proceedings

Coleman, D. J. and S. Li (1999) "Developing a groupware-based prototype to support geomatics production management", *Proceedings of the First International Workshop on TeleGeoprocessing*, Lyon, France, May

Li, S., R. Boettcher, and D. J. Coleman (1998) "Geomatics workflow management requirements in a collaborative production environment", *1998 Conference Proceedings of the American Congress on Surveying and Mapping*, March 2-5, pp.269 –

Easa, S. M., S. Li, and M. Hamed (1997) "GIS applications in infrastructure management: state of the art", *Proceedings of the International Conference of Rehabilitation and Development of Civil Engineering Infrastructure Systems*, June 9-11, Beirut, Lebanon

Conference and Workshop Presentations

Li, S. (2002) "Collaborative GIS for Water Resources Management", ISEIS 2002 Conference on Information Systems for Water Resources Management, September 10 – 13, Beijing, China (accepted)

Li, S. and D. J. Coleman (2002) "Results of CSCW Supported Collaborative GIS Data Production: An Internet-based Solution" accepted for CIG Annual Conference, July 8-12, Ottawa, Canada (Accepted)

Li, S, Coleman, D. J., and Bedard, Y. (2001) "Applying workflow principles to geoprocessing over the Internet", Poster Presentation at the Second International Symposium on Digital Earth, Fredericton, N.B., Canada, June 20-22

Li, S, Coleman, D. J., and Bedard, Y. (2001) "Towards CSCW-supported collaborative GIS data production project management", Poster Presentation in the Third GEOIDE Annual Conference, Fredericton, N.B., Canada, June 24-28, 2001

Coleman, D. J. and S. Li (1999) "Developing a groupware-based prototype to support geomatics production management", Paper presented in the First International Workshop on TeleGeoprocessing, Lyon, France, May 6-7th

Boettcher, R. and S. Li and D. Coleman (1998) "GIS workflow management in an Intranet/Extranet environment", Paper presented in SDI'98 Canadian Conference on Geomatics, Ottawa, Ontario. June 11th

Easa, S. M., S. Li, and M. Hamed (1997) "GIS applications in infrastructure management: state of the art", Paper presented in the International Conference of Rehabilitation and Development of Civil Engineering Infrastructure Systems, Beirut, Lebanon, June 9-11