CANADIAN PATENT OFFICE

AUTOMATION PLAN

Contract No. 54416-7-1499/03-ER

Prepared for:

The Patent Automation Project
Automated Systems Branch
Canadian Patent Office
Consumer and Corporate Affairs Canada

Prepared by:

RES Policy Research Inc.

Authors: Rolf E. Schliewen
John R. Thomas
Ted Eastman
Viktor T. Toth

March 31st, 1988

Table of Contents

Preface 1

Executive Summary 2

1. Introduction 13
    1.1. Background 13
        1.1.1. The Canadian Patent Office 13
        1.1.2. The CPO Needs for Automation 13
        1.1.3. General Automation Objectives 15
        1.1.4. Automation Planning Studies - Major Projects 16
    1.2. Project Objectives 19
    1.3. Development and Implementation Principles 20
    1.4. Purpose of this Document 21
    1.5. The Logic of the Plan Development 22
    1.6. Methods 26
    1.7. Participants and Responsibilities 27

2. Analysis of the Environment 29
    2.1. Patent Application Processing 30
        2.1.1. Process Description 31
        2.1.2. Volumetrics 31
    2.2. Patent Search/Examination Process 32
        2.2.1. Description 32
        2.2.2. Volumetrics 33
    2.3. Information Dissemination 33
        2.3.1. Description 34
        2.3.2. Volumetrics 34

3. Generic Logical Design 36
    3.1. Patent Information Processing (PIP) System 37
        3.1.1. Document Management Subsystem (DMS) 37
        3.1.2. Document Processing Subsystem (DPS) 38
        3.1.3. Patent Search Subsystem (PSS) 38
        3.1.4. Patent Publication Subsystem (PPS) 38
        3.1.5. Integrated Archive Subsystem (IAS) 38
        3.1.6. Remote Access Subsystem (RAS) 39
    3.2. Management Information Reporting (MIR) System 39
    3.3. Financial and Administrative Support (FAS) System 39
    3.4. User Support and Training (UST) System 39

4. The Feasibility of an Automated Solution 41
    4.1. General Design Assumptions and Objectives 41
    4.2. Technical Solutions and Economic Feasibility 43
        4.2.1. Data Representation and Data Conversion 43

4.2.2. Full-text Search Tools and Methods 45
4.2.3. Alternative Architectures 45
4.2.4. Workstations and LANs 46
4.2.5. Remote Access Facilities 47
4.2.6. Software 48

5. Choice Among Alternatives 49
5.1. Selection Among Solutions 49
5.1.1. The PIP (Patent Information Processing) System 49
5.1.1.1. Database Contents 49
5.1.1.2. Proposed Data Holdings 50
5.1.1.3. Data Representation 51
5.1.2. MIR (Management Information Reporting) System 53
5.1.3. FAS (Financial and Administrative Support) System 53
5.1.4. UST (User Support and Training) System 53
5.2. The Preferred Solution - Alternatives Set Aside 53
5.3. Definition of the Preferred Solution 55
5.3.1. The Preferred Automated System 55
5.3.2. Automation Options - Implementation Strategy 58
5.3.2.1. Option One: The Index Database 60
5.3.2.2. Option Two: Separate Text and Image Files 60
5.3.2.3. Option Three: Functionally Integrated Text and Image System 60
5.3.2.4. Option Four: The Preferred General Design (A Fully Integrated Text and Image System) 61
5.4. Conclusion 61

6. Preferred Solution - Costs and Options 63
6.1. Project Cost Estimates 63
6.2. Options - Comparison of Cost and Benefits 70

7. Implementation 73
7.1. Preliminary Action Plan for 1988/89 73
7.1.1. ASB Organization 1988/89 - Principal Responsibilities 73
7.1.2. Project Plan 1988/89 74
7.2. Project Implementation Schedule 75
7.2.1. Organizational Activities 76
7.2.2. Planning 76
7.2.3. Contracting Activities 77
7.2.4. Backfile Conversion 77
7.2.5. Installation 78
7.2.6. Document Management and Processing Systems 78
7.2.7. Search Systems 78
7.2.8. Publication Systems 79
7.2.9. User Support and Training 79
7.2.10. Remote Access Systems 79

7.2.11. Management and Financial Systems 80
7.3. Organization and Staffing Plan 80
7.3.1. Project Management Functions 80
7.3.1.1. Planning and Administration Functions 80
7.3.1.2. Contract Administration 81
7.3.1.3. Facilities and Equipment Planning and Provisioning 81
7.3.1.4. User Support Services 81
7.3.2. Development and Systems Support Functions 81
7.3.2.1. Systems Development 82
7.3.2.2. Systems Operations and Systems Management 82
7.4. Security Plan 82
7.4.1. Familiarization 82
7.4.2. Definition of Security 83
7.4.3. Criteria for Establishing Secure Systems 83
7.4.3.1. Access Privileges and Processing Capabilities 83
7.4.3.2. Classified Information 83
7.4.3.3. Detailed Security Requirements 84

8. Benefits Assessment and Recommendations 87
8.1. General Benefits of Automation 87
8.2. Summary of Tangible Benefits 87
8.3. Fundamental Benefits 88
8.3.1. Strategic Maneuverability 88
8.3.2. Removal of Uncertainty 88
8.3.3. Internal Standardization - External Harmonization 89
8.3.4. Improved Quality and Efficiency 89
8.3.5. Broader Dissemination Capabilities 90
8.3.6. Benefits to Patent Professionals 91
8.3.7. Secure Backup 91
8.3.8. Space Savings 91
8.3.9. Bill C-22 91
8.3.10. Avoidance of Costs of Essential Improvements 91
8.3.11. Derived Benefits 92
8.4. Conclusions and Recommendations 92

Appendices 94

Appendix A: Volumetrics And Throughput Calculations 95
A.1. Tables and Assumptions 95
A.2. Probabilistic Method for Peak-load Calculations 109

Appendix B: Cost Comparison of the Options 117

Appendix C: Detailed Implementation Plan 121
C.1. Implementation Plan: Task and Milestone List 121
C.2. Gantt Chart 124

C.3. Pert Chart 146
C.4. Detailed Cost Estimates 147
C.4.1. Cost Estimates 147
C.4.2. Summary of Costs 150
C.5. Detailed Staffing Plan 169

Appendix D: Wide Area Network Requirements 171

BIBLIOGRAPHY 186

List Of Illustrations

Figure 1.1: The Logic of the Plan Development
Figure 2.1: Major Databases and Processes
Figure 3.1: Major Subsystems
Figure 5.1: Major Database Components
Figure 5.2: The Automated System: Schematic Diagram
Figure 5.3: Automation Options
Figure 7.1: ASB Organization - 1988/89

List Of Tables

Table 5.1: Automation Alternatives - Cost Comparison
Table 6.1: CPO Automation Project Cost Estimates
Table 6.2: Automation Project Costs - Comparison of Options
Table 7.1: Project Plan 1988/89 - GANTT Chart
Table 7.2: Project Plan 1988-96 - GANTT Summary
Table 8.1: Summary of Tangible Benefits

PREFACE

"Historically, the patent system was designed as an information gathering tool"

"Patent files can be regarded as a technical information bank"

"There are enormous differences between countries and between patent offices, in terms of both their awareness of the informational value of patents and the steps taken to exploit it"

(*The Bottom Line, Technology, Trade, and Income Growth. The Economic Council of Canada, Ottawa, 1983, p. 59*)

These quotations from a major study on technology policy in Canada are an apt summary of the situation of the Canadian Patent Office - historically, in terms of opportunity, and with respect to the international challenge. After a careful assessment of its strengths and weaknesses, and detailed studies of solutions, the Canadian Patent Office presents its Automation Plan.

The document forms the basis for a decision of the Government of Canada to commit resources to continuing a broad program of upgrading the strength of the Patent Office through automation, and presents a plan to achieve the objectives considered as essential with minimum technical and financial risk during an eight-year project period.

The study team benefited from consistent advice and direction of the Automated Systems Branch of the Canadian Patent Office, that of senior management, and the Senior Consultant. The staff of the Patent Office assisted the study and supports its objectives, in a time of growing discrepancy between resources and the job to be done. The professional community of patent agents across the country supported the planning process directly through advice, and by their valued response to efforts to assess their future demand for automation-supported services. A number of private companies performed essential preparatory studies. Most importantly, a large number of professionals and entrepreneurs in small and medium size businesses contributed valuable data on the extent and direction of future demand for technological information, which can only be met by an ambitious, well-managed effort to make technical information available to all, and interpret its meaning.

RES Policy Research Inc.
Ottawa, March, 1988.

EXECUTIVE SUMMARY

1. Introduction

This document forms the basis for a decision of the Government of Canada to commit resources to continuing the program of upgrading the strength of the Canadian Patent Office through automation.

Background

The mandate of the Canadian Patent Office (CPO) is to grant equitable and exclusive intellectual property (patent) rights and to facilitate technological information dissemination. The Long Range Informatics Plan (LRIP) adopted in March, 1986, identified a number of weaknesses of the CPO. These weaknesses, especially if compared to projects of technical innovation in other jurisdictions, lead to concerns about the continued ability of the CPO to discharge its mandate as required by law and the best interests of Canadian industry and the Canadian communities of patent agents and professional users of patent information. The key factors internal to the CPO are these: The Canadian classification system has not been maintained; there has been a reduction of the number of patent examiners despite a consistent growth in the number of patents processed; the CPO has new duties deriving from amendments to the Patent Act (adopted in November, 1987); the continued degradation of the Canadian system may have adverse effects on foreign trade and investment; Canadian industry is not making adequate use of Canadian patent holdings; and, recent major increases of fees charged by the CPO were tied to promises of increased service levels which must be honoured. Internationally, the key need is that of harmonization of Canadian data files and procedures with those emerging in other major countries; the U.S. Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the Japanese Patent Office (JPO) have joined forces and are committed to major expenditures (two billion dollars over ten years) to automate their patent operations in a compatible form. In order to promote co-operation, the 'trilateral group' adopts and develops EDP standards, particularly in storage, retrieval, and communication of data, text and graphic images. As any other patent office, the CPO is highly dependent on foreign patent data. Finally, Canada has signed the international Patent Co-operation Treaty, Part I of which will take effect on January 1, 1989. This will result in certain new requirements for applications processing and publication. The internal needs and international developments require action from Canada if the CPO is to claim credible continued membership in the international community devoted to protecting intellectual property.

Logic of the Feasibility Study

This document presents results of the Feasibility Study, Business Case Analysis, and Implementation Plan for the automation of the CPO. The study forms the last in a major work program in 1987/88 under direction of the Automated Systems Branch (ASB) of the CPO. Other studies, planned to mesh with the present project, developed needs of the CPO as such, researched external requirements and technology, and assessed problems and costs of converting

the large paper file holdings of the CPO to machine readable form. In addition, new legislative needs had to be converted into operational requirements, a step which affected the timing and content of preparatory studies in the work plan of the Automated Systems Branch for 1987/88. Current developments in other jurisdictions were monitored as well so as to provide the best possible basis for well-informed investment decisions of the Government.

After an iterative process of assessment - of needs, requirements, estimates of external demand for patent data, and technical factors relating to costs and capacities -, general objectives of the automation project were identified. These are formulated to be specific to automation, attainable in the project period (eight years), and are measurable. Each functional area of the CPO has its objectives; for example, in the processing function, the key objective is to reduce the rising volume of pending applications. Another major objective is to reduce the incidence of prior art missed in the search function. General principles of system development are adopted; their purpose is to limit the technical and financial risk of the CPO during the project period.

The methods of the feasibility study were geared to examining in an orderly progression the needs of the CPO, the potential demand from other users, convert these into mandatory and desirable requirements, and develop alternatives for automation against these requirements under the statement of objectives. Low priority alternatives were discarded in favour of examining in depth the feasibility of truly promising solutions. The key problem thoughout has been that of technical feasibility. This was addressed by developing detailed models for workloads in a future automated system, and assembling estimates, devoting the greatest effort to those components that are large or subject to risk. Then options were narrowed down to the preferred group of solutions, which was further analysed in the form of four options. Computer vendors confirmed estimates of the project team through several independent quotations. Thus the method of cost development was shown to be sound, technical feasibility is assured, and the overall project cost estimate is not subject to significant risk. An implementation plan was elaborated in some detail so as to control the timing of major investments and costs by staff category. All figures are supplied in spreadsheets and costing models which support sensitivity analysis, or scaling of the project to different sizes of data holdings.

2. Environment

The Patent Office is a production organization which processes about 25,000 applications per year, performs the key step of search and examination, grants patents for applications that are not abandoned, publishes newly granted patents, and responds to enquiries for technology search and supply of documents under its dissemination mandate. The volume of work is expected to continue growing at about three percent per year.

The process within each of the three basic functions of the CPO is understood and documented in depth, and these processes are not likely to change under automation, since they are largely determined by law and regulations, and evolved under national and international professional practices. For example, if a claim in an application is modified by agreement between the examiner and the agent or inventor, then the entire page containing this claim is replaced in the

document, while the old page is held in the patent file together with related correspondence. Detailed statistics exist on the numbers of documents, their length and frequency of occurence. These form the basis for estimating volumes and transaction workloads in the automated patent systems. Changes in the environment are expected due to the implementation of provisions of Bill C-22, Canada's adherence to the Patent Co-operation Treaty, and the impact of automation in partner countries, principally in the trilateral group.

3. Generic System Design

Based on general requirements developed through other studies performed in the planning phase, the statement of objectives, and knowledge of other automation efforts underway, the components of a generic patent information system are developed. The Patent Office requires automation in its application processing function for more effective, efficient, and timely operation; it requires facilities for storage in electronic form of all Canadian in force patents and certain foreign documents, with advanced tools for index-based search and full-text search; this includes a facility for data conversion of the backfile and future data conversion; and the Office requires systems to support the external user community through document delivery, search and access to other document holdings in order to meet the dissemination mandate set out in the Patent Act. During the development process and thereafter, the CPO must ensure protection of the integrity of Canadian patent holdings and protection of the commercial confidential property represented by pending applications. Since the present operation of the CPO is almost entirely manual, the automation project is not constrained by features of previously installed systems, as is the case in other patent offices ('legacy').

4. Feasibility of an Automated Solution

Technical feasibility and performance of the automated system depend heavily on how patent texts and drawings are represented, the size of data holdings for rapid access, and the speed of document delivery to workstations. This is the area of technical uncertainty in other patent automation projects which started earlier. Other, less important areas of design concern local communications and workstations as well as software for retrieval and search, and arrangements for external access.

A patent document contains text ('tombstone' data, disclosure, claims, and abstract), and graphic data (about three drawings). Occasionally, there are images embedded in the text. The older the patent, the greater will be the incidence of embedded images, or unrecognizable characters. The search of a patent against a pending application should be assisted by software tools. In the simplest case, such tools permit retrieval through an index, after which the document must be read. Other techniques exist to analyse all or part of the document by content, either through key words set up as an index for search purposes, or through free-form text search, where the choice of words for search is not restricted. The preferred method for document storage is in a form where the text portions are held in coded, or searchable form, and where drawings or embedded images are tied to the text by appropriate codes. This is referred to as integrated text and image form. On the other end of the spectrum, one may store patents in image form alone, where only

the header, or tombstone data provide a handle for retrieval of documents from the database. Given this trade-off, it is desirable to store the greatest number of documents that can be accommodated within cost constraints and performance constraints of the local area network which delivers documents to workstations. Other requirements relate to harmonization and international standards, for example, in that images should be stored at 300 pixels per inch.

A key fact for technical feasibility is the wastefulness of image storage. Advances in methods of image compression are not expected to change this fact significantly. The 400,000 Canadian in force patents, held in image form, would consume about 1,600 billion bytes (1.6 Terabytes), or the equivalent of 500 million pages in text format. That is roughly twenty times the volume needed for integrated text and image format. The greatest part of that volume is consumed by the text of patents in image representation. Image storage can be supported with high turnaround requirements (desired: 10 seconds to first document delivery, and a 'flip rate' of pages of not more than 1 second). However, the cost of such a solution is extremely high, not only due to the cost of storage media, but equally because other system components (local networks, processors, and workstations) grow in terms of required capacity and cost. Telecommunications delivery of documents in image form to external users is considered not feasible; no other patent office is planning such a service, regardless of their budgets.

The integrated text and image form for Canadian patents will use 75 Gigabytes of optical disk space, an amount that does not require slow data storage devices (jukeboxes of optical disks) and where the performance standards can be met with many computers and networks. On-line magnetic disk for patent applications, the index database, and overhead allowances raise the total proposed storage to 150 Gigabytes. In system architecture, a main question in other patent automation projects has been the choice between centralized and decentralized solutions, such as connected minicomputers. If large image holdings were planned, the decentral approach would be superior. If the size of the database is moderate, which can be accomplished for Canadian needs by holding data in integrated text and image form, then each approach is equally feasible. The main criterion for vendors will be whether their solutions provide the throughput required, given a proposed organization of the patent and applications databases. If documents are held in integrated text and image form, then workstations and local area networks of moderate cost are available. Concerning remote access, documents in this form can be delivered for a communications cost of about $0.20 per page.

In sum, technically feasible solutions are available for various forms of data representation and size of databases. These differ greatly in terms of cost and certain ancillary system problems. There are high benefit-cost advantages to solutions that do not require large holdings of documents (for rapid access) in image form.

5. Selection and Definition of the Preferred Solution

The preferred solution is selected in two steps. First, major alternatives are compared which differ greatly in terms of cost, and a preferred solution is identified. Major alternatives range from $65 million to $200 million in total estimated project cost. They differ in the technology for

data representation and the extent of holdings of foreign patent files. Benefit-cost comparisons show that marginal additions of patent documents useful for search carry very high costs, without proportional increments in benefits. Among major alternatives initially considered, only that solution survives which involves Canadian in force patents and U.S. title pages in integrated text and image form, at an estimated total project cost of $65 million.

The technical concept of the preferred solution is based on four incremental options in order to limit the development risk and provide the CPO with assurance that a working system that yields benefits can be put in place in the event that the international process of standardization of integrated text and image data should be delayed.

The options of the preferred solution are displayed symbolically in Figure 1 (next page), and described as follows.

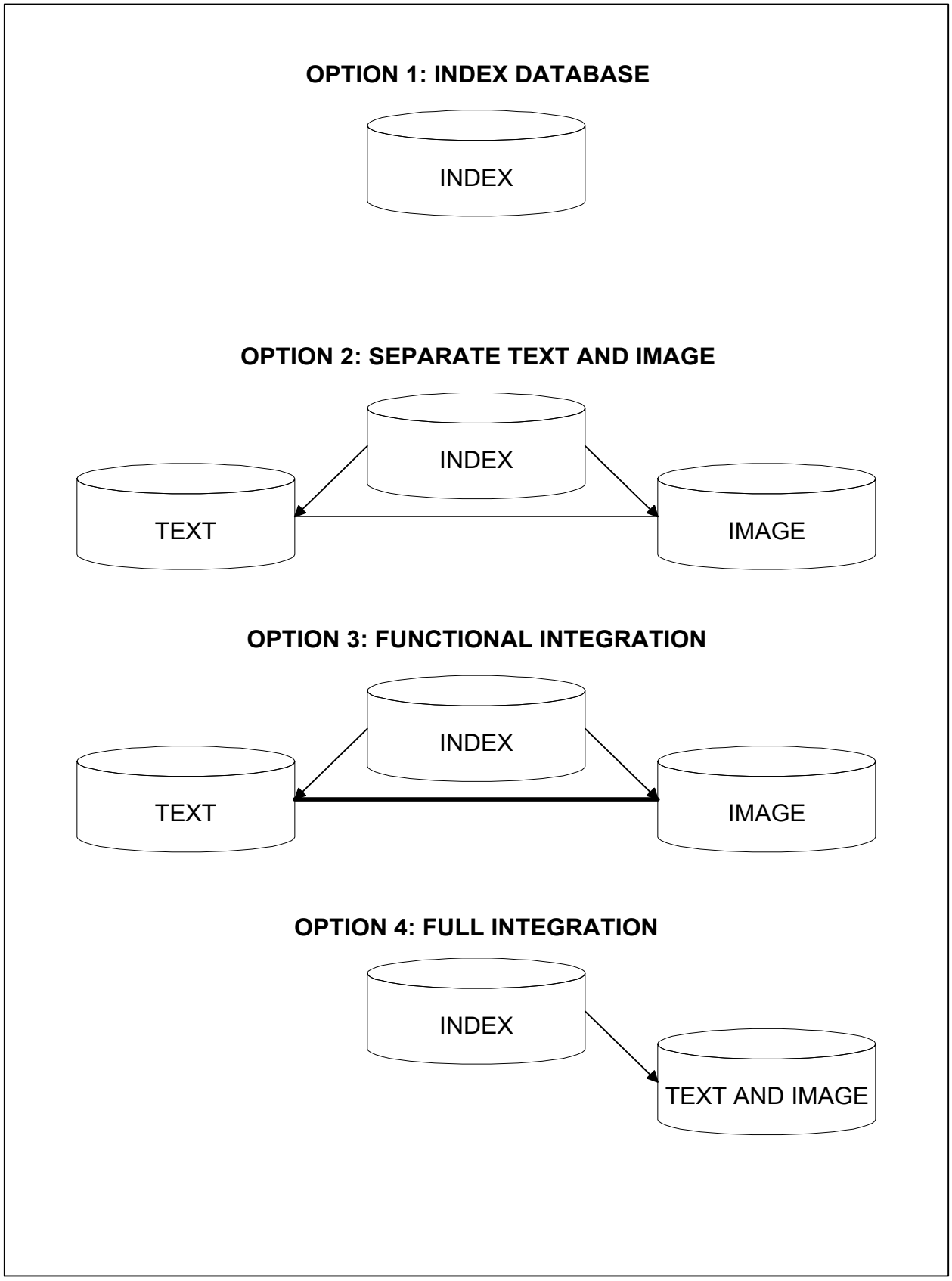

OPTION 1: INDEX DATABASE
INDEX
OPTION 2: SEPARATE TEXT AND IMAGE
INDEX
TEXT
IMAGE
OPTION 3: FUNCTIONAL INTEGRATION
INDEX
TEXT
IMAGE
OPTION 4: FULL INTEGRATION
INDEX
TEXT AND IMAGE

Figure 1.

Option 1 consists only of an index database, to be used for classification search and management of the application documents. All applications and patent documents remain in paper form. Option 2 implements text and image databases which are controlled by the index database. Text and image holdings are handled separately. Functional integration is achieved in Option 3. With this system, documents can be re-created on workstation screens which closely resemble the originals. Option 4 provides full integration of text and image, to a standard such that the exact original can be re-created electronically. Whereas in options 2 and 3, the archive may have to include image forms of documents which are not accessible for rapid retrieval, in Option 4 strict boundaries between text and image holdings will dissappear. International standards for this form of data storage and exchange are expected within the next three years. This strategy of system development ensures protection of investment. The recommended solution carries low, and easily controlled technical and financial risk. At the same time, it is an advanced solution in view of its utilization of expected advances in standards for electronic document interchange and search technology.

The database of the full system is planned as follows. Canadian in force patents (from 1978 on) and title pages of U.S. patents will be stored on-line for full-text search and other retrieval purposes. Beginning with the start of new procedures under Bill C-22, all applications will be converted to integrated text and image format and held on-line. Conversion of older documents, foreign documents, and correspondence for on-line storage is not a requirement. However, retrieval of these document through an automated retrieval system is required. This may occur in analog form through a microfilm storage and retrieval facility linked to the system, by digitization on demand with slower turnaround. In sum, the databases will be modest in size, dynamic, and will grow into the newest forms of data representation as the automation project matures.

Databases for rapid access and search will be backed up by full copies on slower media. All data holdings will be secure through the following measures:

- Final archive of documents
- Storage of all changes to documents
- Off-site backup of the entire file holdings.

Patent documents protected under the Official Secrets Act will not be handled by the automated system. The system contains security measures to protect pending applications against unauthorized access.

Estimates of file sizes for different storage media are given in number of pages held and Gigabytes. The proposed databases require about 150 Gigabytes of on-line storage. The basic estimates are incremented by 30 percent to allow for system overhead.

A full-text search system is proposed. Canada has a special reason for avoiding key-word search: the holdings of patent in the official language of submission would render the task of defining

useful key words nearly impossible. Full-text search will be employed in the language of submission. All user functions of the system will be available in any one of the official languages at the option of the user. A schematic overview of the proposed system is provided in Figure 2 on the next page.

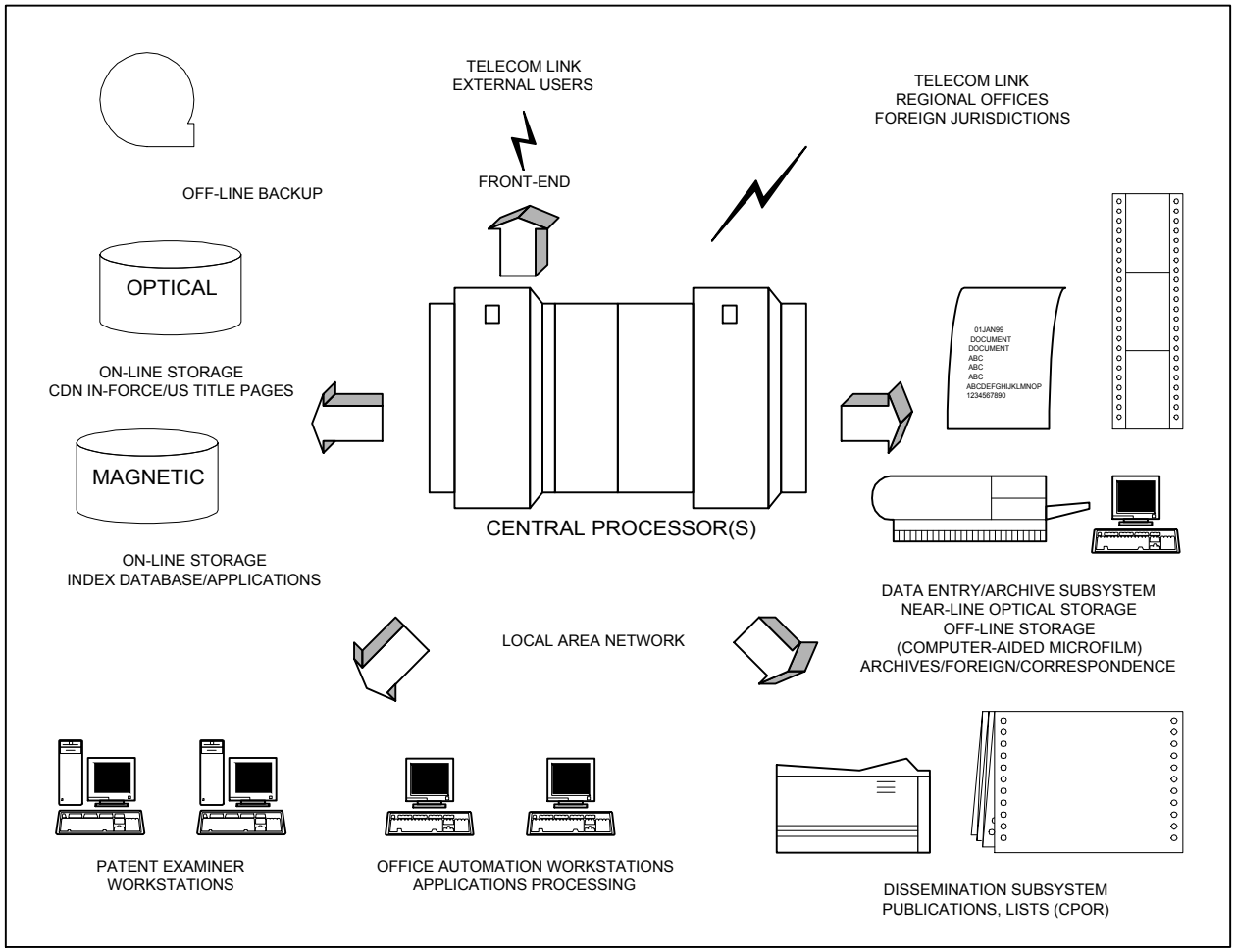


Figure 2.

6. Preferred Solution - Costs and Options

Costs of hardware, workstations, data conversion, software and operation over the expected project period of eight years have been estimated in detail, including calculations of system capacity and throughput. Costs of hardware and telecommunications have been verified with vendors in the form of non-binding quotations. Costs of telecommunications for servicing the external user community are included in the project costs, and are based on estimates of demand for automated patent office services obtained in a market demand assessment study. Costs are based on conservative assumptions regarding file sizes and future reductions in the cost of storage media and workstations. Full costs are used, including costs of an Automated Systems Branch of sufficient size to ensure prudence in contract administration, proper project control, and protection of the databases of the CPO. Estimated total project costs are as follows.

	1988/89	1989/90	1990/91	1991-96	Total
Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	23680	37100
Total:	2220	15995	12535	34180	64930

A comparison of costs and benefits of the four options within the preferred solution leads to the following conclusions. Option 1, the index database, can be implemented for about $23 million. While acceptable as an interim solution, it does not meet certain mandatory requirements, such as full-text search of patent documents, and it does not support delivery of documents to external users. Option 2, providing separate image and text databases, is estimated at $61 million, and Option 3, for functional, but not full integration, at $63 million. Option 2 is the least technically ambitious option that meets mandatory requirements. It can be upgraded without loss of investment. Option 3 is provided as an insurance against the event that international factors prevent the adoption of standards for full restorability of the original form of documents. Option 4, with fully integrated text and image databases, should be developed as the target design, since the other options either provide unacceptable compromises in functionality, or provide less than full functionality at a price which is not sufficiently lower to warrant their choice as the target system.

7. Implementation

A detailed plan for implementation is developed. Its guiding principles are to limit the technical and financial risk, and ensure full control of the project in the hands of CPO management. At the same time, it seeks to provide early tangible benefits for the CPO operation, and ensures that the system is made available for outside use only after it is stable and secure. Project planning tools are in place for continued monitoring of progress against plan, and to support future updates of the plan. The initial organization of ASB and activities in the first year are defined. A summary of major activities and events is provided in the GANTT chart which follows.

PROJECT PLAN 1988/96

#	ACTIVITY	88 89 90 91 92 93 94 95 96
1	Organization	==o==o..1..o..o..o..o..o.
2	Planning	==2..o.=o..o..o..o..o.=o.
3	Contracting	==o==3..................
4	Backfile Conversion	==o======4
5	Installation	==5.........=o....=o
6	Document Management and Processing	==o==o==6......==o
7	Search Systems	==o==o==o==7...==o
8	Publication Systems	==o==o==8.....o
9	User Support and Training	==o==o...==o==o==o
10	Remote Access Systems	==o==o==o..9
11	Management and Financial Systems	..o..o..o..o..o==o==o..o

Legend:
=== : task
... : task suspended
=o= : milestone
=3= : major milestone

8. Benefits

Benefits of patent office automation accrue to the CPO itself, the community of patent agents, and Canadian business. Estimates of internal savings and external benefits are conservative.

Overall, there are quantifiable benefits through cost avoidance of $290 million over the eight years of the project, and $80 million per year thereafter. External economic benefits are obtained through increased productivity of the patent community, and have been estimated at $163 million for the first five years after project completion (current dollars). A point form summary of other benefits follows.

- Automation secures Canada's position in the group of countries which are currently investing in automation.

- A committment to progress in automation will benefit the CPO staff itself by removing the effects of indecision and declining effectiveness.

- Standardization of documents and processing will reduce the backlog; external harmonization of formats for submission, including growing shares of electronic submission and data exchange with other patent offices, must be accomodated, or the CPO runs the risk of losing a large part of its future patent activity.

- With automation, the quality of retrieval and examination will be higher, and can be verified. The examiner can check more, and more efficiently selected documents.

- Dissemination benefits are mainly expected for small and medium firms. Estimates of avoidable waste in industrial R and D that may be eliminated through better technical information are in the order of $200 million for the project period, and $80 million per year thereafter.

- Modest estimates of the cost of patent litigation avoided through better patents and databases are in the order of $10 million per year.

- Benefits to patent professionals are significant, as was demonstrated by positive survey results.

- A long term benefit comes from treating patent data as a national asset with secure backup storage.

- Space savings, although not expected during the project period, will occur in later years; moreover, no additional file space will be used.

- Bill C-22 implementation will be aided, but savings are not identified.

- Other essential improvements (costing $10.9 million during the project period) can be avoided.

Conclusions and Recommendations

Patent office automation is technically and economically feasible. The timing of the project is favourable, as it enables Canada to benefit from recent advances in technology and standards, without, on the other hand, being significantly out of step with automation efforts underway in key partner countries. However, the risk incurred by not proceeding, or the cost of all but minor delays, are substantial. It is recommended to proceed promptly with a well-controlled project of development, and participate actively in international efforts to improve co-operation between jurisdictions in matters of data exchange.

Credit is due to participants in the planning studies and market assessment studies leading to the present plan, for their outstanding co-operation and support of the initiative of the Government of Canada.

1. INTRODUCTION

1.1. Background

1.1.1. The Canadian Patent Office

The mandate of the Canadian Patent Office (CPO), is to grant equitable and exclusive intellectual property (patent) rights and to facilitate technological information dissemination. The granting of patents is recognized as being of fundamental importance to the fostering of technological development in this country by protecting inventions through the granting of patents and by disseminating information on new patents to attract investment and development.

The effectiveness of the CPO in carrying out its mandate depends on the efficiency with which it can execute its functions, which can be broadly classified as follows:

a. patent application processing and examination relating to the granting of patents and their maintenance;

b. patent search on Canadian and foreign patents maintained by the CPO and by external search systems to which the CPO has access; and

c. patent information dissemination.

1.1.2. The CPO Needs for Automation

In March 1986, the CPO completed a Long Range Informatics Plan (LRIP). This plan identified critical factors that influence the degree to which the CPO can exercise its mandate effectively and efficiently. It identified the needs of the CPO for automation given the following areas of concern:

a. CPO Search Files

The lack of maintenance of the Canadian classification system and the lack of maintenance of the file integrity of CPO search files will eventually erode the usefulness of the CPO as Canadian patent files may well become unsearchable.

b. Quality Control

The reduction in the number of patent examiners implies that each examiner must: become an expert in a greater number of fields of technology, spend less time assessing the quality of each application and spend less time searching and assessing prior art. These changes will shift the burden of quality control to the applicant and increase the

probability of issuing poor quality patents and of the prospects for more litigation in Canadian courts.

c. Amendments to the Patent Act

Recent amendments to the Patent Act (Bill C-22) necessitate the development of new CPO practices, procedures and processes.

d. Harmonization with Foreign Patent Offices

Automation projects have been completed or are underway in a number of patent offices abroad. These initiatives followed the evolution of image scanning and compression and optical character recognition technologies which placed within reach the technical and economic feasibility of computerized document text and image management and search systems on very large databases. In particular, major automation projects have been undertaken by the United States Patent Office (USPTO), the European Patent Office (EPO), the Japanese Patent Office (JPO) and the West German Patent Office (GPO). The estimated $2 billion (Canadian) being invested by these countries to patent system automation is solid evidence as to the importance they attach to these initiatives. The first three of the above patent offices have formed a Trilateral Agreement for Cooperation in terms of setting standards, exchanging their respective patent documents in electronic form, and other patent matters. At the other end of the international scale, small patent offices are automating their patent systems in anticipation of world communication of patent information in electronic form in the near future.

e. System Deterioration

Further deterioration of CPO facilities and the patent system which it supports may well send a signal to our trading partners and foreign investors that Canada is no longer interested in providing adequate protection for patents and the R&D and investment associated with those patents. This may, therefore, have a negative impact on foreign investment in Canada.

f. Canadian Use of The Patent System

The lack of awareness, understanding and use of the patent system by Canadian industry relative to other industrialized countries may affect future competitiveness. Part of the problem is due to lack of regional access to adequate patent information services.

g. Fee Increase

Fees of the CPO for patent services were raised in 1985. The Government justified this fee increase with reference to increased service levels and future automation.

1.1.3. General Automation Objectives

The Long Range Informatics Plan identified emerging operational problems of the Canadian Patent Office and matched these against technical opportunities as well as aspects of the international situation in patent office automation. In addition, the priorities of the Government had to be considered before terms of reference for automation planning studies could be set, and general objectives to be met by a major automation initiative could be defined. The Automated Systems Branch formulated primary and intermediate, or operational objectives and obtained approval for them from senior management of Consumer and Corporate Affairs. In this manner, there is a demonstrable linkage of the objectives of the automation project as such to the mission of the Department and priorities of the Government. The following primary objectives were set out. Each is specified further by indicating how it can be addressed through automation.

1. To stimulate small and medium size businesses

    - By fully exploiting the potential for access to and dissemination of Canadian and foreign patent information.

2. To enhance trade, research, investment and technology transfer opportunities

    - By providing a modern, efficient and effective patent system with improved standards.

3. To provide equitable access to, and distribution of, patent information throughout the regions of Canada

    - By means of effective use of computer and communications techniques.

Based on a more specific assessment of conditions noted in the LRIP and opportunities seen in other countries, the following intermediate, or operational objectives were defined.

A. To be able to deal with an increasing volume of patent applications while maintaining the integrity of patent files.

B. To be able to search an ever expanding document base in a manner consistent with performance standards.

C. To be able to raise service levels and improve the standard of the patent system without significant increases in employment levels.

D. To be able to maintain harmonious interfaces with foreign patent offices.

E. To be able to meet the operational requirements arising from the revised Patent Act (Bill C-22).

F. To be able to provide communications links to at least the five major regional centres in Canada so as to ensure more equitable access.

G. To be able to minimize the duplication of disseminated information and to utilize shared resources where possible in the dissemination of information.

Specific project objectives of the automation project were developed during the feasibility study. Before these are introduced, a brief synopsis of results of several of the planning studies performed in preparation of this Automation Plan will be given.

1.1.4. Automation Planning Studies - Major Projects

The Automation Plan is an outcome of a systematic program of studies which were designed and executed in 1986/87 and 1987/88 by the Automated Systems Branch in response to approval by the Cabinet of the Long Range Information Systems Plan. Overall, sixteen studies were undertaken to address various internal and external aspects of automating operations of the Patent Office. This program of work was completed on time and within the budget allocated for this purpose. Most studies were performed by consulting firms selected through competition; others were conducted by staff of the CPO. A list of titles of all projects is included in the Bibliography. Thus, the strategy and plan of automation outlined in this document rest on a base of prior research and investigation.

The following paragraphs are a summary of objectives and results of four major studies of immediate interest for the Automation Plan.

Market Demand Assessment (CPO Work Plan, Task 4)

This study was a first full-scale marketing research project to address the present service offering of the CPO to outside users, and assess the likely future demand for patent-related services which can be met by an automated patent office. The study sought to segment the market and obtain qualified information on the nature of patent information demanded by various groups of users

across the country. It also attempted to derive estimates on price sensitivity of future services which rely on automation. Methods consisted of enquiry among specialists in the profession, and surveys by telephone and mail.

Among existing services, the Technical Information Search Program (TIS) and the Public Information Program of the CPO received strong backing from the user community. Similarly, the patent document distribution service which is handled by a facilities management firm, was rated highly, however, there are indications that demand for microfilm copies of patents is beginning to shift in favour of U.S. documents. Other products, such as the Canadian Patent Office Record (CPOR), received low rates of approval, which indicates that there is a need for a better method of publishing new technological information. The market research among potential users disclosed a strong, widely based demand for patent information. Calculations indicate that between one and three million 'uses of patent information' per year can be expected from various user groups. How much of this potential business volume the CPO can attract depends on the service timing, extent, quality (in terms of value added), and price of its service offering. Various user groups are identified and ranked by priority. These range from patent agents to chemists and R and D firms, to engineers, large businesses, to the general population of technically interested users. These differ in the amount of equipment they would be prepared to maintain, their knowledge and need for assistance. The study concludes that the automated patent retrieval system should be built and made accessible to external users. The high potential demand for services from an automated patent system indicates that the proposed investment will carry a low financial risk.

General Requirements (Task 5)

This study developed a high level statement of the design alternatives open to the CPO in introducing information technology. Based on an analysis of present operations, needs of the CPO identified earlier, and future expectations, a design of a 'future logical system' was developed. The data flows and data structures that must be handled by any future automated system were described. A set of common design assumptions was developed and accepted by management. Their purpose was to establish boundaries within which the search for general options for solution of the automation problem should proceed. Then four options were defined as general scenarios.

It was found that these can all be conceived as increments to a base design. The base design consists of a stand-alone Canadian solution to the problem of application processing and storage and search of Canadian patents. Other files would be processed as before - mainly through manual searching of paper documents and microfilm. Other alternatives differed essentially in their extent of on-line holdings of partial or full copies of documents from other jurisdictions, and the degree of telecommunications access to foreign files. Each alternative was assessed, but no choice was proposed. A general path of implementation was laid out, and preliminary estimates of costs were prepared. The project period was estimated as requiring up to eight years, depending on the alternative, and likely total project costs were estimated in the order of $80 million to $170 million.

Technology Assessment (Task 6)

This study was commissioned to evaluate present and estimated future costs and performance for various technological alternatives - paper, microform, magnetic, optical - with respect to storage, retrieval, data and document conversion, telecommunications, workstations, and electronic publishing, and to provide supporting data for the cost/benefit analysis of general design options. The project consisted of a large survey of the status, developments and trends in the subjects indicated. Interim results from other patent offices were also assessed, and a number of practical lessons for the implementation strategy of the CPO project were derived. A few selected conclusions and cautions derived from this study are summarized.

It was found that optical storage of the image version of large volumes of patent documents requires jukeboxes of optical disk which, in the Canadian situation, is a problematic solution. Instead, the storage of text and imagery in more compact form, preferably in conformance with emerging international standards, was identified as a promising solution, one which may avoid a long-term commitment to large holdings of optical disk. It was found that in view of advances in desktop publishing systems, the CPO should consider optical character recognition (OCR) methods for data conversion, especially if these are accompanied by a requirement to submit applications in a consistent layout.

The study predicted drops in price over the next five years (in the order of thirty to fifty percent) in three main items - storage media, workstations and telecommunications. In the latter area, it was found that delivery of patent documents in image form to external users will be extremely costly, so that this method of dissemination cannot be supported. In terms of practical advice, the technology assessment concluded that the CPO should plan to accommodate manual methods of processing for some time, that the Office should always have a fall-back solution in case of technical or organizational problems, and that special attention should be devoted to data protection and security of access. Overall, the study concluded that there are no overriding areas of technical risk, or technical advances worth waiting for, which would warrant a decision to delay the automation project.

CPO Pilot Studies & Evaluations (Task 8)

This study assessed the file and data integrity of Canadian indices and search files to ensure that data and document bases contain accurate basic bibliographic information on ownership, technology and industrial classification or country of origin of patents, in order to determine the scope and cost of data and document conversion work, and it determined the feasibility of using foreign sources of data to build the CPO database.

One main result concerning the equivalence of foreign and Canadian patent documents was that data conversion of Canadian patents must be done from Canadian sources. A demonstration data conversion facility was set up and samples of documents were converted, using various sources

of equipment and software. Estimates of person years and equipment needs for conversion of the backfile of Canadian patents, the file of in-force patents, and the pending file were derived. The study recommended that the backfile should be converted to image form, and the file of in-force patents by OCR methods with manual edits. Depending on options and assumptions, the estimate of data conversion costs of the total Canadian holdings is in the order of ten million dollars. The study concludes that there are no significant technical risks in a project to produce a machine-readable file of Canadian patent documents of high quality.

1.2. Project Objectives

As a basis for the orderly introduction of automation functions, the following objectives have been established in conjunction with CPO staff. The objectives are stated in order of priority and are justified with respect to obligations of the CPO under law, departmental objectives and objectives of the Government. The objectives are stated to be specific to the automation project, attainable in the project period and finally measurable.

Each project objective encompasses one or more requirements which were derived from previous studies. Not all of these requirements are of equal importance. In order to show how potential automation solutions were assessed against requirements and proposed project objectives, each objective is identified as reflecting either mandatory or desirable features of the target solution (M or D).

Processing Function

Efficiency and Productivity

- To reduce the rising backlog of patent applications and reduce the patent pending period (M)
- To manage the rising workload of the Canadian Patent Office (M)

Quality of Service

- To maintain the quality of patent examination (M)
- To improve the quality of patent examination (D)
- To meet Patent Co-operation Treaty standards for publication (typesetting) (D)
- To improve liaison with clients and foreign patent offices (D)

Effectiveness

- To improve management planning and control of workflow (M)
- To reduce the risk to document and data holdings overall, and maintain the protection of commercial intellectual property (M)

Search Function

- To reduce the incidence of prior art missed in searching by increasing the holdings of readily accessible data (M)
- To increase the quality of search by introducing improved search tools (M)
- Develop the capacity to exchange patent data with other jurisdictions (harmonization) (M)
- Increase the availability of patent data to the public for better protection of intellectual property (M)
- Increase the level of protection of Canadian data holdings (D)

Dissemination Function

- Improve the response to public demand for state-of-the-art searches (M)
- Make patent data available through telecommunications access in five regions (M)
- Provide access to CPO holdings of foreign patent data, where not excluded by licence agreements (M)
- Provide public access facilities in response to demand (M)
- Provide automation support to a growing number of intermediaries (D)
- Distribute machine-readable data to end-users and intermediaries (D)

1.3. Development and Implementation Principles

In addition to these substantive project objectives, several elements of a strategy for development were adopted. These may be considered as principles to be observed during the process of feasibility assessment, development, and implementation. They are derived from prior planning studies, reflect lessons learned from other jurisdictions (USPTO, EPO, and GPO), and are summarized as follows.

- The system development and implementation process should at all times support at least the level of functionality of the present manual Patent Office operation. This means that in no phase of the process, a division of the CPO or an external user group should be subject to a decline in performance or services levels, even if this might be justified against future improvements.

- The proposed solutions should be open to future advances in technology. This rule implies that design decisions and commitments to hardware or forms of data representation should be avoided which might 'lock in' a solution that would be an irreversible constraint on future advances.

- The CPO will not support any solutions that rely for their success on significant research and development efforts.

- The implementation path should provide opportunities to 'back out' into levels of technology or functionality previously reached. This rule was adopted so as to limit the risk of development in the face of future effects in Canada of new legislation (Bill C-22), and in order to absorb uncertainty arising from the progress of other countries in automation, especially from developments in the standards field.

- The planned process of conversion of paper files to electronic form, and the mode of future storage, should at all times exclude a risk to the integrity of Canadian holdings of applications and patents. The implementation effort as well as the resulting system must provide full security for protected intellectual property against inadvertent or intentional infringement.

- These development principles were all treated as mandatory constraints in the process of comparing alternative solutions, and they are reflected in the proposed implementation plan.

1.4. Purpose of this Document

This document comprises a Feasibility Study, Business Case Analysis and Implementation Plan for the automation of the Canadian Patent Office. It consolidates the findings of previous studies with independent research to determine the feasibility of a match between operational requirements and an automated solution.

The Feasibility Study commences by establishing user requirements. Within the context of these requirements, a logical design of each environmental sub-system is defined together with projected measures of storage and transaction volumes. Design requirements are then established by translating the results of the data analysis into absolute measures using alternative physical data formats. These criteria are then used to assess available technology for the determination of both technical and economic feasibility.

The alternative solutions for automation developed in the feasibility study are then compared using classical economic evaluation methodologies and an Implementation Plan is developed for the preferred solution. The plan incorporates a schedule of major activities, a definition of the resources required for implementation and cost estimates with supporting rationale.

Finally, the cost estimates are summarized and presented together with a summary of the benefits to be derived.

In support of this document, detailed study documents and descriptions of modeling methodology have been attached.

1.5. The Logic of the Plan Development

The general objectives of the automation project were finalized after a series of studies had been completed. These studies were defined and commissioned under a general workplan of the Automated Systems Branch for 1987/88. Each objective is a result of detailed assessments by study teams and Patent Office management of a number of factors, among them the internal needs of the Canadian Patent Office, needs of the user communities in Canada and abroad, realistic limits of technology, plans and results achieved in other jurisdictions, and demonstrations of vendors.

A brief sketch of how other projects were logically linked, and how the present feasibility study progressed will show how the proposed automation solution emerged through an orderly process of examination of options, constraints and assumptions.

First, an analysis of present internal operations of the patent office was performed. Then, the internal automation needs of the CPO were analysed and compared to a logical analysis of its workflow. At the same time, a major market research project sought to secure data on expected external demand for patent-related documents and services. A technology assessment study described current trends in technology such as storage, workstations, and telecommunications. These estimates were used to predict certain future prices and capacities in this study. An analysis of problems to be anticipated in the area of file conversion addressed questions of file integrity and likely costs of data conversion. An assessment of the societal economic benefits of an automated patent office of the future is underway; its results are not reflected in the present document. Concurrent with these projects was an activity of the Patent Office itself to clarify its statement of objectives, update it as needed, and adjust organizational objectives to requirements arising from new legislation.

These projects were planned in terms of content and timing to produce an orderly progression toward the feasibility study document and implementation plan. Despite some overlap in timing, the results of prior projects were generally sufficient to focus crucial questions of feasibility properly, and narrow the options for solution in an orderly manner. The progression of logical steps can be summarized as follows.

Current operation of the CPO
Knowledge of weaknesses
Needs to be met by an automated system
Expressed needs and estimated demand from other users
Needs and preferences converted to preliminary options for solution
Generic design of an automated system
Preliminary costing of options
Derivation of requirements (mandatory/desirable)
Detailed examination of technical feasibility
Comparison to other options
Choice among options, discarding low-priority options
Detailed costing of preferred option
Implementation plan for preferred option
Verification of technical feasibility of preferred option with vendors
Final development of project plan

The key problem in this study has been the need to cut an extremely complex decision problem down to manageable size, without, on the other hand, shutting off opportunities to benefit from expected technical change. This process had to be managed in a rational, accountable manner. For example, the likely design to be chosen must not be specified down to a level that appears to favour a particular vendor of hardware or software, yet it should be detailed enough so that vendors can respond to enquiries with comparable pricing information. The project authority and project teams adopted three precautions to ensure an orderly progress of the feasibility study. These are summarized next, followed by a summary description of methods employed in the study.

- Automation Assumptions

  It was found that the wide choice of possible alternative solutions to the problem of automation of the Patent Office had to be somewhat restricted in the interest of addressing the most promising solutions in sufficient depth. Therefore, a set of 'automation assumptions' was written down early in the process of requirements definition, and approved by management. These are found in report 4 of the study on General Requirements. By way of example, one such assumption set down a rule to the effect that none of the solutions should rely on research and development projects for their success.

- Preliminary Definition of Options

  Another precaution to avoid duplication or circular work was the decision to adopt general options for automated solutions early as a point of reference for further research. Four such options were adopted (General Requirements study). They differ mainly in terms of size of data holdings and likely cost. It was also decided to treat the 'status quo' option, that is the

option of not automating, only as a point of reference for development of costs. It was decided that failure to automate is NOT a serious, viable course of action for the Patent Office. The status quo option in turn has certain add-on costs, which are identified in the Long Range Informatics Plan. These relate to essential investments in restoring the Canadian Classification System to full usefulness, a project that must be taken in hand if full automation were not possible.

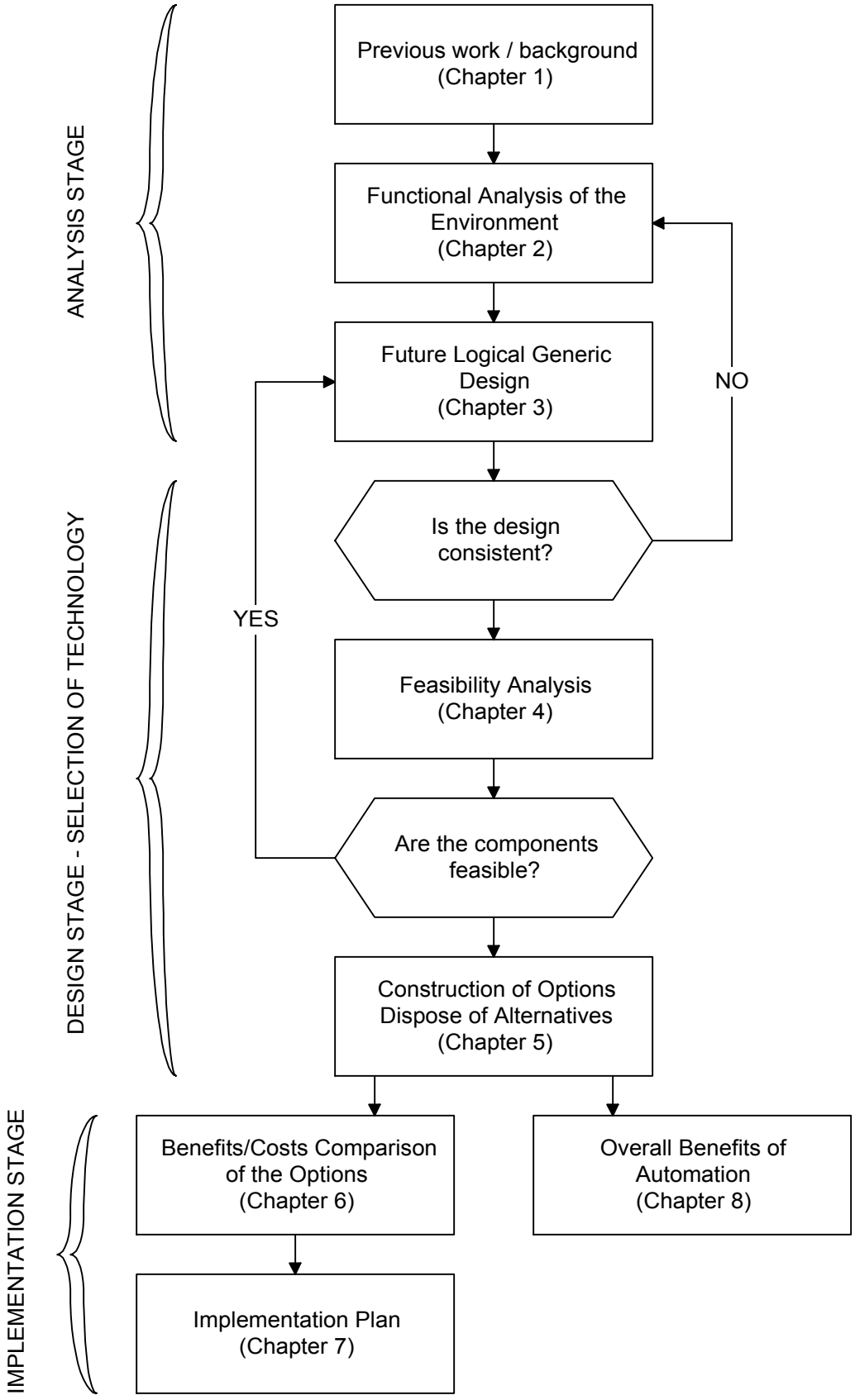
- Rules for Business Case Analysis

  Similar to the general design assumptions, the treatment of cost data was described in a set of rules (Business Case Analysis - Rules). Their purpose was to lay the groundwork in terms of method well before results were visible. For example, one such rule says that costs will be expressed in constant 1987 dollars.

  The way this plan evolved is shown in Figure 1.1. Following the functional analysis and the future logical design, a major decision point was reached. The logical design was analysed for internal consistency as well as consistency with the automation objectives. If it failed to be consistent, the analysis of the environment was repeatedly performed. Another major decision point followed the feasibility analysis. If any of the proposed generic system components proved infeasible, the logical design had to be modified.

The appendices attached include:

- volume and throughput requirement calculations, tables, and methods,
- a detailed plan for a particular implementation alternative, including a cost study, for reference purposes, and
- a study on telecommunications requirements and basic service costs.


ANALYSIS STAGE
DESIGN STAGE - SELECTION OF TECHNOLOGY
IMPLEMENTATION STAGE
Previous work / background (Chapter 1)
Functional Analysis of the Environment (Chapter 2)
Future Logical Generic Design (Chapter 3)
Is the design consistent?
NO
YES
Feasibility Analysis (Chapter 4)
Are the components feasible?
Construction of Options Dispose of Alternatives (Chapter 5)
Benefits/Costs Comparison of the Options (Chapter 6)
Overall Benefits of Automation (Chapter 8)
Implementation Plan (Chapter 7)

1.6. Methods

Methods employed in the feasibility study followed the system life cycle methodology chosen and amended by the project authority.

An informal survey of available technologies was performed from literature studies and interviews. Samples of proposed solutions were obtained from a survey of vendors together with cost estimates. The vendors were provided with design requirements in terms of storage requirements, peak-load estimates of transaction volumes and performance requirements to determine capacity, define specific architectures and quote estimates of costs.

In each area of costs or benefits, a set of working papers was developed as a basis for costing. Major subjects, such as the sizing of the database or estimation of the conversion effort, were developed in the form of costing models which supported sensitivity analysis and assessment of design risks. For example, we know how much more storage is used by Canadian, typewritten patents as compared to U.S. style printed documents. The costing model permits other analysts to insert their assumptions on different document sizes. In general, the greatest effort of costing was devoted to principal cost items - those in excess of ten percent of the overall project cost estimate.

All costs were rolled up into spreadsheets which carry sufficient detail to reflect an implementation strategy. This means that the effect of accelerating and retarding of major components of the overall investment can be checked easily.

The question of technical feasibility was crucial. This means: can we be certain that the proposed design, computer facility, storage, and local network will in fact produce the performance which the CPO has set down as required? Therefore, non-binding quotations were sought and obtained from three vendors in order to test the feasibility of the solution and the adequacy of its description for estimating purposes. The vendors provided independent cost proposals which agree to within 10 percent, demonstrating in the process that:

a. the CPO design criteria and requirements had been stated in terms detailed enough for the vendors to respond with precision;

b. the proposed configuration could be implemented within current technological limitations.

All potential suppliers confirmed that they could meet the technical requirements with existing hardware. The closeness of the estimates, remarkable at this stage of an implementation project, gave comfort to the authors as to the validity of the total cost estimate.

In the area of accounting for benefits, it was decided to be very conservative. While some savings of person years in the later part of the project history are likely, the CPO must adapt to a major change in legislation, and the net effect of concurrent introduction of automation is best left to

future updates of the plan. The major internal benefits come from maintenance and improvement of effectiveness, if not simply the survival of the Patent Office as a viable organization. Societal benefits are demonstrated by the surprisingly high estimates of external user demand (Market Survey). Their dollar value to the Canadian economy has been estimated as being $163 M over five years after project completion (Economic Impact Analysis).

Finally, the proposed design had to be backed by an implementation plan, if only in order to have a rationale for the timing of major investments. Another reason for developing the implementation plan in some detail was the fact that the staffing plan depends on it, and the mix of government staff in the Automated Systems Branch and contractor project personnel could only be rationalized after both these categories were identified in detail.

1.7. Participants and Responsibilities

Ray Taylor
Director, Automated Systems Branch
Canadian Patent Office
Consumer and Corporate Affairs
Project Authority
Henri Denolf, Project Coordinator
Pierre Lalonde, Coordinator, Task 10
Richard Woloschuk, EDP Project Officer
Don McMaster, Project Officer

C. Guruprasad
Executive Vice President
Gamet Consultants Ltd.
Senior Consultant

RES Policy Research Inc.
Study Team: Rolf E. Schliewen, John R. Thomas, Ted Eastman, Viktor Toth

Advisory and Steering Committees

Industry Advisory Committee
Composed of senior CPO management and representatives of the Patent and Trade Mark Institute of Canada (PTIC). Advises on needs of patent agents and the general public.

Roger Gagnon, Chair

CPO Steering Committee

Represents management of the CPO. Approves substantive system requirements, and settles priority issues concerning internal needs of the CPO.

Mart Leesti, Chair

Technical Advisory Committee

Composed of senior EDP managers of six major Federal Government Departments. Advises on the general adequacy of proposed methodologies and technologies.

Ray Taylor, Chair

2. ANALYSIS OF THE ENVIRONMENT

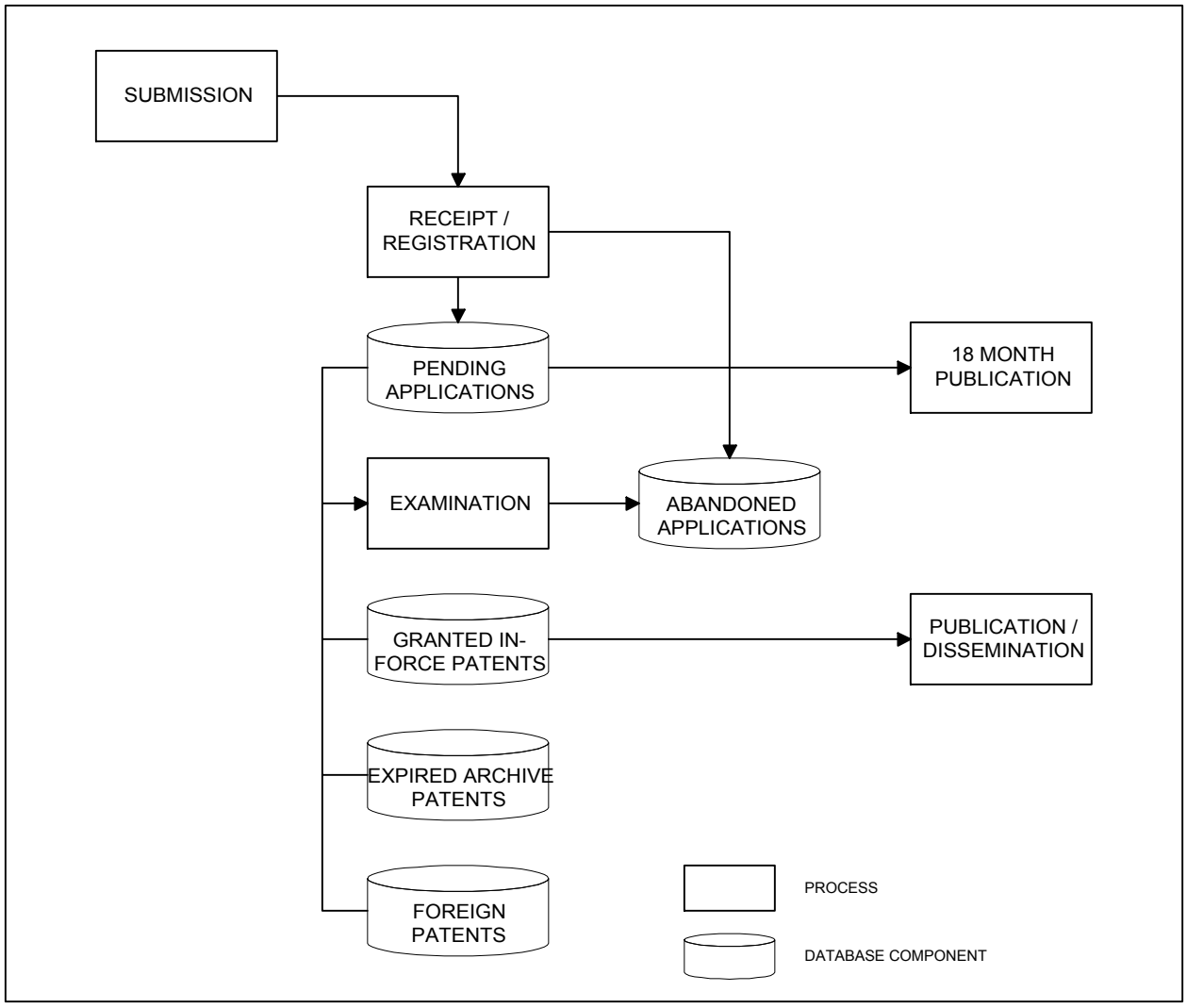
This chapter provides a brief overview of the major processes performed by the CPO and the document volumes involved. It is not a substitute for a more complete analysis but an extract of the one performed in the study on General Requirements.

Figure 2.1. identifies the major stages of the workflow at the CPO.

This chapter is further divided into three sections, representing the three major areas of activity. Section 2.1. identifies the administrative and document management duties associated with the processing of patent applications. Classification, examination, and search are described in Paragraph 2.2. (Patent Search/Examination Process). Publication of patent documents, abstracts, and other means of providing information to the public are analyzed in Paragraph 2.3. (Information Dissemination).

Each following paragraph is divided into two parts: process description and volumetrics. The volumetrics sections describe the size of the databases which the processes operate on, as well as annual volumes, in units of patent/application documents and pages.

This chapter relies extensively on Appendix A containing tables describing database volumes, annual throughputs and peak-load estimations.



2.1. Patent Application Processing

The application file contains the following document portions: the application itself, consisting of disclosure, claims, abstract and drawings; the correspondence portion, including letters, reports, page replacements, etc.; and the front page, created at the CPO, containing title, classification and other indexing data. Many parts of the file, especially correspondence material are created during the application processing and examination.

From the receipt of an application until the grant of a Canadian patent, the patent application workflow process can be divided into four major parts. These are: receipt, 18 month application publication, examination and issue. Descriptions of these processes follow.

Annual and cumulative volumes of applications will also be examined. These results are translated into EDP terms in Appendix A.

2.1.1. Process Description

The receipt of an application, as a process, consists of administrative elements, such as acknowledgement. The application is also examined as to whether it conforms to certain formality and quality requirements, as defined under the Patent Act. If that is not the case, the applicant is notified, and the application is either abandoned, or re-submitted by the applicant.

At the next stage, applications are classified and examined. Classification determines the field(s) of technology the application belongs to, and those likely to contain relevant prior art. The search process, conducted by patent examiners, is to ensure that the application is unique and does not violate rights of prior patents. Because of its importance, the examination/search process is viewed as a major activity, and is examined in the next paragraph. The present paragraph is concerned only with the administrative elements of the process, such as examination requests, search reports, applicant responses and allowance/abandonment.

According to new legislation, applications are to be made public 18 months after receipt. This represents a sizable extra workload for CPO personnel, and creates an additional requirement for any proposed automated system.

Allowed applications enter the final stage of processing: issue. At this stage, the applicant responds to allowance, which is followed by the granting of the patent - the patent will then be published, thus made available to the general public. Also, if renewal procedures are introduced, regular renewal requests and fee payments will be processed.

2.1.2. Volumetrics

The annual volume of applications is expected to grow to 35,000 documents by the end of the automation project. An application file contains, on average, 56 pages; 22 of which are the application text (disclosure, claims, abstract and drawings), and the remainder is correspondence (letters, reports, etc) including all material generated during application processing. Altogether, approximately two million application file pages will be processed annually.

The processing of an application includes approximately 40 document accesses or other transactions, involving about 230 pages. On an annual basis, this represents 1.3 million transactions/accesses and eight million pages. These figures do not include the volumes of search/examination.

At present, there are three years of pending applications, representing a total of six million application pages. One of the objectives of automation is to reduce the examination pendency time to two years, equivalent to 70,000 documents or four million pages.

2.2. Patent Search/Examination Process

The patent examination and search processes are at the core of all CPO activities. The quality of the work of the Patent Office as well as its international reputation depends on the accuracy of this process. It is also one of the major areas of concern, for a number of reasons: growing number of applications, outdated classification system, filing errors, etc.

During search/examination, a patent will first be checked for formalities. It will then be classified and cross-referenced; these functions determine the technological areas where a search will be made.

The purpose of the search is to check the application against earlier documents and hence, to determine its uniqueness.

2.2.1. Description

The first step of examination is a formalities check of the application which is followed by classification.

Classification determines the technological areas to which the patent belongs and hence the area to be searched.

The primary purpose of the search is to determine the uniqueness of the invention in question. In most cases, in-force Canadian patents and prior applications will be searched. Archived (expired) documents may also be involved as well as foreign patents.

The search process itself consists of a number of individual steps. These are index search, abstract/first claim/drawing search (i.e. the contents of the first page of a US patent document) and full document search. The objective of the process is to find references to similar inventions or to find proof that no such invention is in existence.

The first step, an index search, discloses the group of patents in the selected technological areas which may be related to the invention.

The examiner then retrieves the abstracts, first claims or first drawings of a selection of documents, perusal of which permits a further reduction in the number of full patents to be examined.

The full text of the remaining number of documents will be examined.

In difficult cases (for example, when it is hard to determine to which technological area a certain invention belongs), electronic tools, such as full-text search capability on a wide set of documents, would be of great help.

Depending on the search results, an application may be granted, abandoned, or have its claims modified.

2.2.2. Volumetrics

An average Canadian patent consists of 22 pages (1 front page, 1 abstract page, 15 disclosure pages, 3 claims pages and 2 drawing pages). At present, there are 1.2 million patents issued, approximately 400,000 of which are in-force, the rest is archive material. This results in a total of 26 million pages, 9 million of which belong to in-force patents. These patent documents (especially the in-force documents) comprise the primary population to be searched at the CPO.

It is expected that, by the end of the project period, approximately 200,000 new patents will be created. It will not affect the number of in-force patents (approximately the same number of patents presently in-force will qualify for the archives by then). The total number of Canadian patents will be 1.4 million by 1996.

The examination process involves data retrieval from the classification index, retrieval of front and abstract pages, full documents and, in case of an automated system, full-text search on documents. Altogether 7.3 million accesses, resulting in 15 million page retrievals, are predicted for the future system. Appendix A describes in detail how these numbers are derived, and the consequences in context of automation.

2.3. Information Dissemination

In the context of automation, there are two ways to publish information: paper publications and electronic form.

As part of the application processing, paper publications of applications ('18 month publication') and granted patents have been discussed already. In the future, however, the CPO may elect to develop more sophisticated publications intended for the general public. With an electronic system, publications containing listings of newly issued patents, containing front pages, perhaps organized by art, are completely feasible and do not require significant investment. In fact, such future publications may be turned into a revenue-generating operation of the CPO.

Non paper-based forms of information dissemination are of increasing importance. Two forms of electronic information distribution exist in general. For the purposes of this discussion, we have classified them as 'on-line' and 'off-line'.

An off-line publication is, for example, an electronic media (floppy disk or CD-ROM, for example) containing portions of the patent index. These forms, though not presently in existence, may become very popular since most of the potential users will very likely have computer equipment on-site.

On-line information services are typically remote-access data centers. The automation project designers should take into account, that more and more potential users in the general public have the equipment to access such databases. After automation, the CPO could easily provide such services.

With external access provided, security measures should be taken into account. System design and management methods must prevent unauthorized access - applications prior to disclosure (18 month publication or grant) are, for example, classified as 'commercial confidential'.

2.3.1. Description

At present, CPO publications are produced with no significant help of electronic tools. The publication of granted patents is very largely a manual process - actually, they are photocopied. These publication activities together with the proposed '18 month publication' of patent applications are covered in Paragraph 2.1. (application processing). Other publications of the CPO are also created in labour-intensive ways. In fact the CPO is, at present, far behind the patent offices of other developed nations (Japan, for example), in terms of the quality, variety and amount of published material.

An automated system at the CPO offers the possibility of using desktop publishing tools for inexpensive, effective, quality publishing. The technical background option is described in reasonable detail at Paragraph 4.3. (technical solutions). As it was mentioned above, it would permit the CPO to issue periodicals aimed at different sections of the general public, with information about new inventions, CPO activities, etc.

Another major area of information dissemination, as it was mentioned above, is the distribution of electronic media, such as diskettes, tapes, CD-ROM, etc. These media may contain patent information (text or graphics) or, for example, classification system extracts together with retrieval software for commercial microcomputers.

On-line type access is also technically feasible with an automated system. 'Heavy' users, such as patent agents, institutions and large commercial users would be the likely users of such a system. Aside from constraints imposed by the comparatively low-speed of wide area communication facilities, and obvious security restrictions which would apply, these users will enjoy the same level of service as the internal staff of the CPO.

2.3.2. Volumetrics

The volumes involved in paper publishing are included at application processing. They represent 35,000 applications ('18 month publication') and 20,000 new patents, adding up to a total of 55,000 documents or 1.2 million pages.

On-line user requirements were measured by C-L Market Survey. It is estimated, that the patent agent community will represent approximately 20,000 'uses' of an on-line system, resulting in the

electronic equivalent of 600,000 page retrievals. Other users will access the system approximately 1 million times a year, according to the above-mentioned survey study.

There is no information available at present about the demand for other types of information. A survey of the general user community in that respect would help planning such publications. However, the volumes we can anticipate will not be significant, at least for an automated system.

3. GENERIC LOGICAL DESIGN

A general statement of requirements for an automated solution for the CPO was presented in Chapter 1. In Chapter 2, a summary of the results of the detailed study of the CPO environment was presented in order to determine the boundaries and context for defining an automated solution and in order to define the future logical design. In view of the environmental analysis, this chapter translates the automation requirements into a generic system design.

The automated patent system should incorporate the following requirements for four major system components.

A Patent Information Processing (PIP) system should provide the application processing workflow, data storage/retrieval, examination/search and dissemination functions.

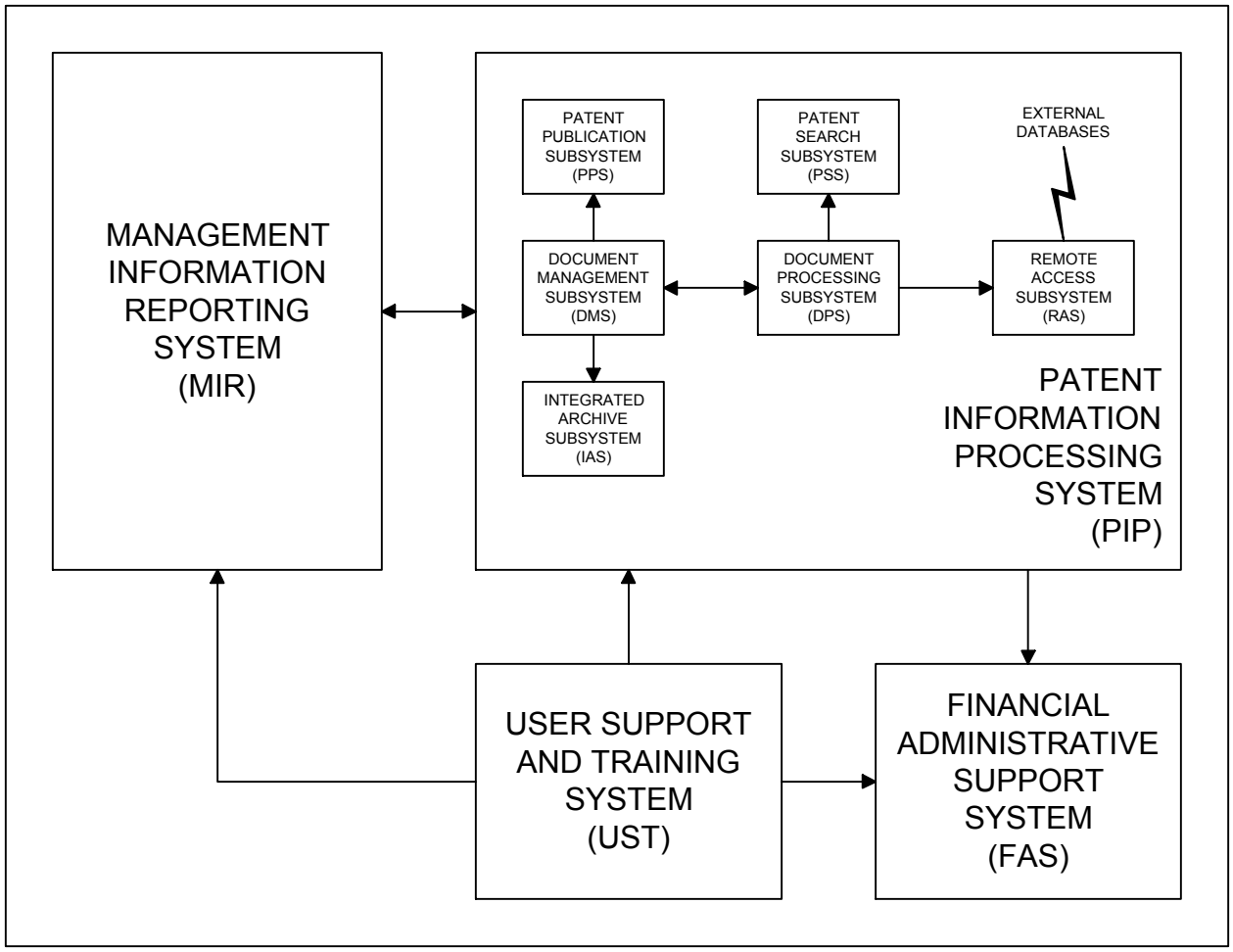
A Management Information Reporting (MIR) system will support project management, systems operations, scheduling/tracking of application processing and systems security.

A Financial and Administrative Support (FAS) system will handle the increased workload occasioned by introduction of Bill C-22 in the area of fees management and, if required, will be capable of supporting other financial systems.

A User Support and Training (UST) system provides training tools for both internal and external users and operational support (e.g. on-line help facilities).

Figure 3.1. is an overview of the major system components and the more important relationships among them.

The following sections detail these components on the abstract level.



3.1. Patent Information Processing (PIP) System

The Patent Information Processing system includes six functional subsystems, which are defined below.

3.1.1. Document Management Subsystem (DMS)

A document management subsystem is required to provide tools for storing, retrieving, locating and controlling the flow of documents. Specifically the subsystem must incorporate features to automate the document conversion, document filing and index editing, document retrieval, and audit trail processes.

The document conversion function must provide the capability of converting documents received at the CPO, such as applications or revisions to pending applications, and existing documents required for entry into the system databases.

The system must be capable of allowing the creation of an index or file folder for associative reference by which means the document may be accessed and retrieved when required.

The system must be capable of delivering any document or any portion of a document to the user, using either direct document access or access through indices. It must also support contextual search tools which are implemented (e.g. full-text search).

3.1.2. Document Processing Subsystem (DPS)

A document processing subsystem is required to store and distribute documents and to support the processing of the information contents of documents with automated tools. Specifically, it is required to support the functions of application processing as described in Chapter 2.

3.1.3. Patent Search Subsystem (PSS)

A patent search subsystem is required to provide both a direct search capability on patent and application index data to locate text and drawing pages of patents and applications, and a full-text search capability on abstracts, claims and disclosures.

3.1.4. Patent Publication Subsystem (PPS)

This subsystem provides all publication and dissemination functions. These are the 18-month publication of pending applications, the publication of newly granted patents and publications intended for the general public such as the Canadian Patent Office Record (CPOR). An automated system may facilitate positive responses to special requests by subscribers, upon-request publications of selected documents and the possible creation of new periodicals of the CPO - such as publication of front pages and abstracts of new patents in selected technological areas. These forms of publications may actually be turned into revenue-generating operations while increasing the quality and range of information dissemination.

3.1.5. Integrated Archive Subsystem (IAS)

An integrated archive system, capable of handling a variety of storage media (e.g. microfilm), is required to provide a reliable means of preserving information and for providing a means of locating prior art automatically and delivering it to the user through the same system as other information is accessed.

Documents archived in the future will preserve their particular storage format. Stored on slower, inexpensive media, the IAS will be responsible for their maintenance.

3.1.6. Remote Access Subsystem (RAS)

This system serves two purposes. It should provide access to the CPO databases for external users as part of the dissemination function, and it should provide reliable means to interconnect to foreign databases and commercial data sources.

3.2. Management Information Reporting (MIR) System

Five major activities are covered by this system, in the fields of application processing, examination and search, project management, dissemination and security.

In the application processing area, this system is responsible for the tracking of all activities, providing audit trails. Scheduling of activities, automatically or manually, is also one of this system's responsibilities.

Scheduling examiner activities becomes extremely important when technical difficulties do not permit ad-hoc document retrieval. Tracking examiner activities may help to ensure examination quality and to produce reliable search records.

The tracking of external user activities is important for accounting purposes. This system may also provide auxiliary functions (i.e. statistical information about fields of interest).

Support to project management is very important in ensuring system consistency. It may provide information about subsystem tests, integration criteria, development efforts, etc.

Access security must prevent unauthorized access by internal or external users. It should also provide reasonable safeguards against accidental or intentional database destruction (such as unauthorized modification of the patent index, for example).

3.3. Financial and Administrative Support (FAS) System

With the implementation of Bill C-22, the CPO will experience an increase in workload with respect to fees management.

As a minimum requirement, this system should be capable of monitoring renewals and corresponding fees. It should also be open-ended, both in respect to future improvements, and to interconnection to other departmental systems.

3.4. User Support and Training (UST) System

An issue of key importance with an automated system is the provision of quality training and support to users. Without it they would not be capable of using the system effectively.

Training programs and facilities are the first step. Trained users will also need support - documentation, both off-line and on-line (interactive help facilities) will serve this purpose.

This system will handle the on-line support facilities, and support training by providing simulated training environment, for example. It will also provide tools for documentation development.

4. THE FEASIBILITY OF AN AUTOMATED SOLUTION

This chapter, is based on the analysis of the environment (Chapter 2) and the definition of generic system design (Chapter 3). First, design criteria will be derived. Subsequent sections will address certain areas of information technology in order to determine the feasibility of these criteria.

Previous work shows that the most critical factor for the success of the automation project is the representation of the data. Technical feasibility and performance as well as costs depend heavily on optimal database design.

Other areas under examination are architectures, local data traffic, workstations, remote access facilities and software.

In sum, this chapter absorbs and assesses technical components of solution to the automation problem, and thus lays the groundwork for a comparison among major alternative solutions in Chapter 5, leading to a definition of the preferred solution.

4.1. General Design Assumptions and Objectives

A successfully implemented automated patent system would be capable of supporting and monitoring application processing activities, delivering any document or any portion of a document to users (patent examiners), and supporting and aiding various electronic search functions. These general objectives define the following framework of the system:

The system must store all relevant documents. This is essential for delivering document data to any user of the system. It will be addressed later, what database size can be considered economically feasible.

For documents which are not included in the database for economic reasons, it is desirable to provide other kind of access. Index information (including abstracts) can be stored in this case, or telecommunications connections can be provided to other (foreign) data holdings.

It is desirable to store the text contents of the documents in machine-readable form. Advanced electronic search tools can not be implemented otherwise. This means, that while a direct search (using the patent index) can be supported by a different system, full-text search is only possible if the text contents of documents are stored as text, rather than images.

In order to support advanced search capabilities, efficient search software should be deployed on this text database. This, in turn, may define a requirement for inverted files to support word search, or other more effective methods.

For effectiveness in the patent examination process, image data is essential. Patent examiners heavily rely on the drawings contained in a patent document. Some documents would become incomprehensible without this information. It is also worth mentioning, that graphic information

often occurs embedded in text (a chemical formula, for example, or as a matter of fact, a Greek letter or a mathematical symbol). A text-only database may exclude this type of information, therefore rendering document files completely useless.

The system would accomodate the standards, recommendations and guidelines of the World Intellectual Property Organization, (WIPO).

A reliable index system must be implemented. A very likely candidate for this purpose is the IPC which, once fully implemented, can support national classification codes as subsets. Patent abstracts or first claims (or both) are also parts of the IPC indexing. This index system would provide a reliable organization of patent data and a powerful search tool as well, permitting search on index data, classification(s), abstracts, etc.

Workstations must be provided for application processing and patent examiner staff. As a minimum requirement, application processing workstations can be standard office automation equipment, while patent examiner workstations must be capable of displaying high resolution graphics information.

Efficiency of the system heavily depends on its speed. Users should be served to their greatest convenience. As a general requirement, it has been established that, whatever transaction a user initiates, a 10 second initial and 1 second subsequent system response time should be provided. (This rule may not apply to certain transactions - for example, a simple index query should be serviced in much less than 10 seconds, while a data conversion transaction may take up to several minutes using presently accessible OCR technology.)

Local communications links must be provided for data traffic between workstations and data storage. The link must be fast enough to support the design criteria of delivering a document page in less than a second. The large number of users and the predicted frequency of transactions represent a severe criterion on local data communications network design. If external users are provided with on-line telecommunications access, the thus increased demand for information must be taken into account.

To improve information dissemination, (telecommunications) access for external users can be provided. Also, the system should be capable of supporting improved publication activities. (As a matter of fact, certain publication services can be provided with virtually no additional costs, such as the creation of an improved CPOR, while other services can generate enough revenue to cover any additional investment made.)

As for the index system, existing or evolving international standards must be observed when designing the document database. For example, the USPTO standard of image representation should be adopted. Other standards for representing patent information in a more advanced form should also be observed. The principal proposed standard, best suited for patent documents, is DATIMTEX.

Certain standards should be adopted for the system software used. First, as a Government policy requirement, all user interfaces of the system must be fully bilingual. The system should also be capable of handling documents in both official languages. (Handling documents in other languages is not required, but may be fully feasible at no additional cost with certain systems.)

Finally, security considerations must be taken into consideration, to ensure data integrity, prohibit unauthorized access to commercial confidential documents, and prohibit unauthorized modification of files.

Future advances of information technology must be anticipated. Generally, an open-ended design philosophy should be adopted, which does not prohibit the incorporation of new software (e.g. advanced contextual search tools, such as search on chemical formulas); new storage methods (improved methods for storing documents incorporating both text and image); and other major improvements (e.g. machine translation). Failure to employ such a design strategy would result in the need for a complete system reconfiguration if a new development is to be incorporated.

The following sections address individual fields of technology in respect to the above objectives.

4.2. Technical Solutions and Economic Feasibility

4.2.1. Data Representation and Data Conversion

This section first examines different data representations in terms of functionality, then identifies data storage technologies which can be used to accomodate these databases.

The data handled by the automated system consists mostly of patent or application documents. As it was stated, a text-only representation of these documents would not suffice because it does not include drawings and embedded images which are of vital importance for efficient patent examination.

Image storage, although widely used in a number of areas, presents some severe difficulties. First of all, the size of such data files is enormous; extremely large storage facilities would be required to store a sufficient number of patent documents. Appendix A provides some volumetrics calculations on such databases. Although data compression methods can certainly be improved (a conservative assumption was used for these calculations), the size of such a database is still around one Terabyte (one trillion bytes).

Functionally, the system should incorporate both text (in machine readable form) and images, in order to achieve the best performance. Different levels of information technology can be identified in this respect, raising different feasibility questions:

Separate holdings of text and image information can be constructed without any foreseeable technical difficulties. The only question is that of data conversion. Manual key-entry of the text of a large number of documents can be prohibitively costly. Also, the quality of the data acquired in this manner is questionable. The ITP study on data conversion shows that, although the keying

error rate is relatively low, the errors can be severe, often changing the meaning of the entire document.

The same study examined the accuracy of electronic character recognition (OCR). It concluded that the error rate on documents of good quality (well-preserved, typewritten or printed paper files) is still within acceptable levels and, more importantly, these errors are easy to detect (e.g., a 'G' interpreted as the digit '6' in a word). These results apply for English text only. There are difficulties in the electronic recognition of accented characters. However, the proportion of non-English (French) language documents is low, therefore manual keying is affordable, and future advances in OCR technology can be anticipated.

These text and image databases can easily be integrated on a functional level. Users do not have to know, which database a page is coming from, when they place they requests. From these data holdings, pages containing both text and image information can also be constructed.

In the ideal case, a full integration of text and image data is implemented. As in some state-of-the-art desktop publishing systems, the origin of text and image 'objects' on a page is indistinguishable. The reproduction of a page is identical to the original. Once that level of integration is achieved, the image-scanned versions of documents can be disposed of.

Presently available data storage technology can be classified as follows. Sequential media (such as a magnetic or optical tape) permit the retrieval of data in the order it was recorded. This technology is used usually for backup purposes. The speed and nature of sequential data retrieval does not support effective on-line systems.

Random access storage (where the availability of data does not depend on its relative position on the media) can be further classified. Off-line systems require manual intervention. An example is the floppy disk drive of a desktop computer - whenever new information is to be accessed, the user has to replace the floppy disk. Average data access time can be up to several ten seconds.

An example for an on-line system is the hard-disk drive of the same equipment, where information is always readily available. The media in this system is always mounted, in contrast to off-line systems. Data access time is usually measured in milliseconds.

High-volume optical jukeboxes represent a third category of random access systems. Although no manual intervention is required, the media is not mounted, and the data access time is comparable to off-line systems. Storage devices of this kind are sometimes referred to as 'near-line'.

Sequential data storage is by far the least expensive. Near-line systems are economically effective for very large data holdings. On-line systems are the most expensive, but they provide the most benefits as well.

A database of 400,000 in-force Canadian documents represent 1.6 Terabytes, when stored as image. A database of such size can only be stored on near-line or off-line equipment; the costs associated with any other solution would be unacceptable.

A database consisting of text and image data can be far less expensive. Preliminary calculations in Appendix A show that the size of such a database will be approximately 75 Gigabytes (75 billion bytes). A database of this size can be constructed without employing exotic technology. On-line optical storage is a very likely alternative. It is cheaper than magnetic storage (although considerably more expensive than optical near-line equipment), and offers a speed comparable to on-line magnetic systems, therefore not rendering the construction of an on-line system impossible. Optical devices of write-once nature (WORMs) may also help to implement security measures against unauthorized modification of files.

As a final consideration it should be noted, that whatever system is selected, the 'collision' of simultaneous requests from a large number of users may increase data access times by orders of magnitude. Appendix A offers some calculations in this respect.

4.2.2. Full-text Search Tools and Methods

As stated, it is highly desirable that the automated system would be capable to perform sophisticated word search on documents. Tools performing word search on a database of this size exist based on various principles. In the present case, employing dictionary (Thesaurus) dependent word search tools is not desirable. As stated in the Automation Study of the GPO, the creation of patent-related dictionaries is a formidable task. In the Canadian case the biligual nature of document holdings results in increased difficulties.

The following section focuses on the difference between centralized and distributed architectures. Non-conventional, distributed architectures are often employed for increased search efficiency. (Several processors perform the same search functions on relatively small portions of the database.) Other methods include the automatic creation of index files. When selecting a systems architecture, the existence of high performance search tools for that particular system should be verified.

4.2.3. Alternative Architectures

Earlier studies as well as foreign documents examine the benefits and drawbacks of centralized vs. distributed architectures in the context of the automation of a patent office. Two system components are in question: the processor and the database. Since most of the processor's work is related directly to data storage and retrieval, the distribution of processing power should be considered together with the distribution of the database.

Centralized architectures are conventional. Software as well as operational methods are well established. In these systems, however, the processor is the main bottleneck - system expansions

are limited by the processor's throughput capability. For these systems, further expansions require complete reconfiguration, or the purchase of new equipment.

Distributed, or decentral, architectures are defined for the present project as follows. They include all solutions where the path of data, for example, from a storage device to a workstation is not uniquely determined with respect to major system components. In the case of the CPO system all processors and the data bases will be located at the main site, regardless of the solution chosen.

Distributed systems hold great promise. Each element of such a system is fully functional, loosely connected to the rest of the system. There is theoretically no limit to add new elements to such a system. The individual elements do not have to handle a large workload, therefore they are technically simple. Considering a (rather primitive) brute force full-text word search on in-force Canadian patents as an example (without suggesting that such a method should be employed at the CPO), a centralized system's processor will have to search through 400,000 patents and provide a result in a short time, which is not possible even with state-of-the-art supercomputers. In a distributed system of a thousand elements, one processor would have to handle 400 documents only - this task is not beyond the capabilities of a high-end personal computer.

Naturally, state-of-the-art search methods do not employ the primitive scenario described above. It is an important fact, however, that in order to avoid the rather significant overhead created by document indices or inverted files, distributed architectures are proposed. If a requirement exists at the CPO for such an architecture, it is the full-text search capability.

Distributed systems are sometimes described as providing inadequate security in respect to database integrity and prevention of unauthorized access. Another problem is, that software development for these systems may require significant R&D efforts. However, where these difficulties are overcome, the distributed system may prove equal to, or superior to, the centralized system. The likely solution will probably involve a moderately distributed system. In any case, the boundary between these classes of systems is becoming less distinct.

The CPO automation requirements do not contain anything in favour of a centralized or a distributed system. Either can be used, if the existing criteria (e.g., full-text search capability) are met.

4.2.4. Workstations and LANs

As was already stated, two groups of workstations must be provided. Patent examiner stations (as well as the workstations provided in the public search room) must be capable of displaying graphic pages in sufficient detail. Application processing workstation can be less advanced, although limited graphic capabilities are desirable.

A more severe requirement is that the workstations must a.) support the speed of local communications facilities, and b.) support the composition/display of document pages in less

than one second in order to achiveve the performance required. In order to support efficient work on behalf of patent examiners, capability of displaying two full-size pages simultaneously is highly desirable (e.g., for providing the application under examination and the retrieved patent document at the same time).

All workstations must support bilingual communications and data. Desirably, workstations would support individual printing and data entry devices.

A text and image database, as shown in Appendix A, does not impose severe criteria on local communications facilities. The data traffic volume is not unacceptably high, even under unfavourable (peak-load) conditions.

An image database requires a network of extremely high speed. As Appendix A shows, throughput requirements are measured in several Megabytes a second. More importantly, workstations must be capable of communicating effectively at high speed - this requirement alone may represent an enourmous cost increase. Even in this case, rigid operational methods will likely have to be adopted for effective usage of the system. High-volume data storage at the workstations may be required, as well as pre-scheduling of data retrieval activities. In simle terms, the throughput requirements will define a system not permitting ad-hoc access to the database: a batch processing system will be the likely result.

4.2.5. Remote Access Facilities

Remote access facilities provide means for data access for external users, and interconnection to foreign patent offices and commercial databases.

The main technical question is the feasibility of transmitting patent documents over wide area networks. At present these networks provide limited performance in terms of speed within affordable cost frames.

A simple calculation shows that transmitting a document in compressed image format at the highest available speed (19.2 Kilobaud - 19,200 bits per second) takes 1/2 to 1 hour, depending on the system's overhead. This is clearly not a preferred solution.

Integrated text and image documents can be transferred in a matter of minutes, their text portion in a matter of seconds. Lower line speeds can be used, with lower communications costs.

Conclusively, providing compressed image patents to external users is not practicable, although feasible. If patents are stored in image format, index and abstract information should be provided only. There is no similar problem when integrated text and image databases are in use.

To access graphics or formatted text data, workstations at external user sites must be capable to decompose compressed image or integrated text and image files. Decomposition at the CPO (and transmission of the decomposed document) would increase the transmission time by orders of

magnitude. The decomposition can be solved by software or hardware means. For text access only (e.g. contextual search or retrieval of abstracts), any remote workstation would suffice.

The main question for successful interconnection to foreign offices/external databases is the compatibility of data representation. There is no technical problem for the access itself.

4.2.6. Software

There are some areas where software implementation is critical. Most important is the text search capability. Full-text search in the literal sense (that is, retrieval of all documents in question and comparing all words against keywords) is definitely NOT feasible. Sophisticated search tools exist, but they must be adapted to use on integrated text and image databases.

The complexity of the PIP system as a whole may lead to unexpected difficulties during the course of software development and integration. Usage of state-of-the-art software engineering tools (CASE tools) will help to reduce the risk in this area.

The composition/decomposition of integrated text and image files is also a critical software question. In particular, workstations must be able to complete the page composition in a few hundred milliseconds, to construct and display the composed image on the screen in time.

Similarly, if an image database is used, compression/ decompression software, which meets the criteria, must be provided. It must be noted that workstations will have to perform the much simpler task of decompression only.

If an image database is used, the physical length of document files (and consequently the time required to transmit them on local area network facilities) prevents the creation of an efficient on-line system where ad-hoc data retrieval is possible. In order to achieve a reasonable system performance, advanced work scheduling is necessary. Such a batch system will require significant software development efforts, because the efficiency of the entire system will depend on its performance.

Finally, if electronic data interchange is being done with external databases, data conversion must be done. In special cases (different image resolution, for example) this may represent a serious software development activity.

5. CHOICE AMONG ALTERNATIVES

Based on the findings of Chapter 4, it is now possible to translate generic system functions (Chapter 3) into technical solutions. For certain subsystems the solution is obvious and does not require further consideration. Other subsystems represent technical or economic difficulties (data conversion is an example), therefore much care should be taken when selecting a solution.

The first part of this chapter defines a preferred path for automation. Then this solution will be compared to other major alternatives which were considered in earlier stages of this project. Finally, four options within the preferred solution will be identified and described. Chapter 6 is devoted to a detailed cost analysis of the preferred solution, a comparison of the options in terms of costs and benefits, leading to a final recommendation.

5.1. Selection Among Solutions

5.1.1. The PIP (Patent Information Processing) System

For this system to perform its function, a database solution must be selected. All subsystems are directly related to the patent information stored. Questions to be answered here are about database contents (document portions to be included), database size (selection among documents), and data representation (electronic format).

5.1.1.1. Database Contents

Systems which are not capable of storing/retrieving patent documents should be considered only as fall-back options. Such systems would not be capable of performing a number of functions, such as application processing, or fast retrieval of document contents for examination purposes.

Economic reasons may prevent the storage of all portions of every available document. Therefore, the following order of importance among document portions was established:

1. Index Data
2. Abstracts/First Claim
3. First Drawing
4. Full-text with Drawings
5. Correspondence relating to Applications

The electronic equivalent of a patent or application file including some or all of the above portions, together with the necessary software to implement it, is often referred to as 'Electronic File Wrapper' (EFW).

A fully implemented WIPO indexing schema would include text abstracts or first claims. If it is extended with at least one of the drawings, an equivalent of the title pages of US patent documents is the result.

5.1.1.2. Proposed Data Holdings

This section extends the conclusions on technical feasibility by addressing questions relating to the size of the database. The question to be answered here is, that given the best technical solution, what range of documents and which elements of these documents should be included in the database to provide an effective and affordable solution?

Systems with very large data holdings are extremely costly. The USPTO automated system can be taken as an example, holding 4,500,000 American patents. Even if R&D expenditures can be avoided (based upon the experience of the USPTO), the implementation costs would be as high as 200 million dollars or more.

An automated system implementing an on-line database of Canadian in-force patents has been established as a minimum requirement, and preliminary cost estimates provided in Chapter 6 demonstrate that the implementation of such a system is within acceptable economic limits.

Table IV/B in Appendix A shows that the most frequently accessed patents at the CPO are Canadian in-force patents; US in-force patents, Canadian archive patents, foreign (non-US) in-force patents, and foreign archive patents follow approximately in that order. Moreover, the same table shows that the majority of the accesses to U.S. data, foreign data and other prior art do not involve the entire document (e.g., item 19 in the table indicates that 1.8 pages/document on the average are retrieved during foreign search); in most cases it is on the level of abstracts only. It has been determined that the addition of the title page of U.S patents to the Canadian database of in-force patents would lead to no significant increase in costs, nor would it present technical difficulties.

It is doubtful whether the benefits offered by a database extended beyond this limit would justify an expenditure of the magnitude indicated above. Instead, other alternatives should be taken into consideration, such as on-line access to the automated systems in other jurisdictions, or the implementation of a database of limited size (holding of abstracts, for example).

Although it is not a requirement to include U.S. and foreign patents and other prior art in the on-line database, it is a requirement to access them through an automated system. Such a retrieval capability can be implemented in two ways.

First, older Canadian patents and in-house collections of foreign patents can be filmed, if necessary, and stored on a microform or roll-film based computer aided retrieval subsystem. This system would operate in a batch mode for processing document requests, and could in fact be integrated with the data conversion function into one subsystem. This system can also incorporate existing microfilm holdings. However, some of these films have seriously deteriorated and therefore re-filming of some documents may be unavoidable.

Frequently accessed documents could be digitized and included in the core database where Canadian in-force patents are found. If documents were converted by such a system on an "as and when required" basis, only patents of significant economic value would be converted, and no significant cost increase for this level of integration would be required over the estimates established for data conversion as a whole.

Second, telecommunications access to the USPTO database is feasible. The full-text database can be searched using standard text oriented terminals on public packet switched services. Assuming that the CPO were to establish an agreement with the USPTO for such access arrangements, the communications software in the CPO could be configured to accept a request from any user, and place a call from the CPO computer to the USPTO host machine to provide a pathway to the US search system. (Different user groups, such as internal and external users in general, may of course have different access privileges.) Such access arrangements could be provided with no significant increase above the base costs, since the Canadian design includes the costs of connections to a public packet switched network. The USPTO, in turn, has an installed gateway to a US packet switched network and the two networks are connected through international gateways.

There are about sixty countries which have implemented public packet switched data communications networks to the same standard and have established interconnect agreements with the Canadian carriers. In the short term, this network will provide the only cost effective means for establishing universal access to the world's stock of patents which now number some 20,000,000. There is a requirement for low-cost data network services which provide much broader bandwidth than currently available. Such services will evolve only in the longer term as the necessary economies of scale are achieved through the integration of voice and data networks.

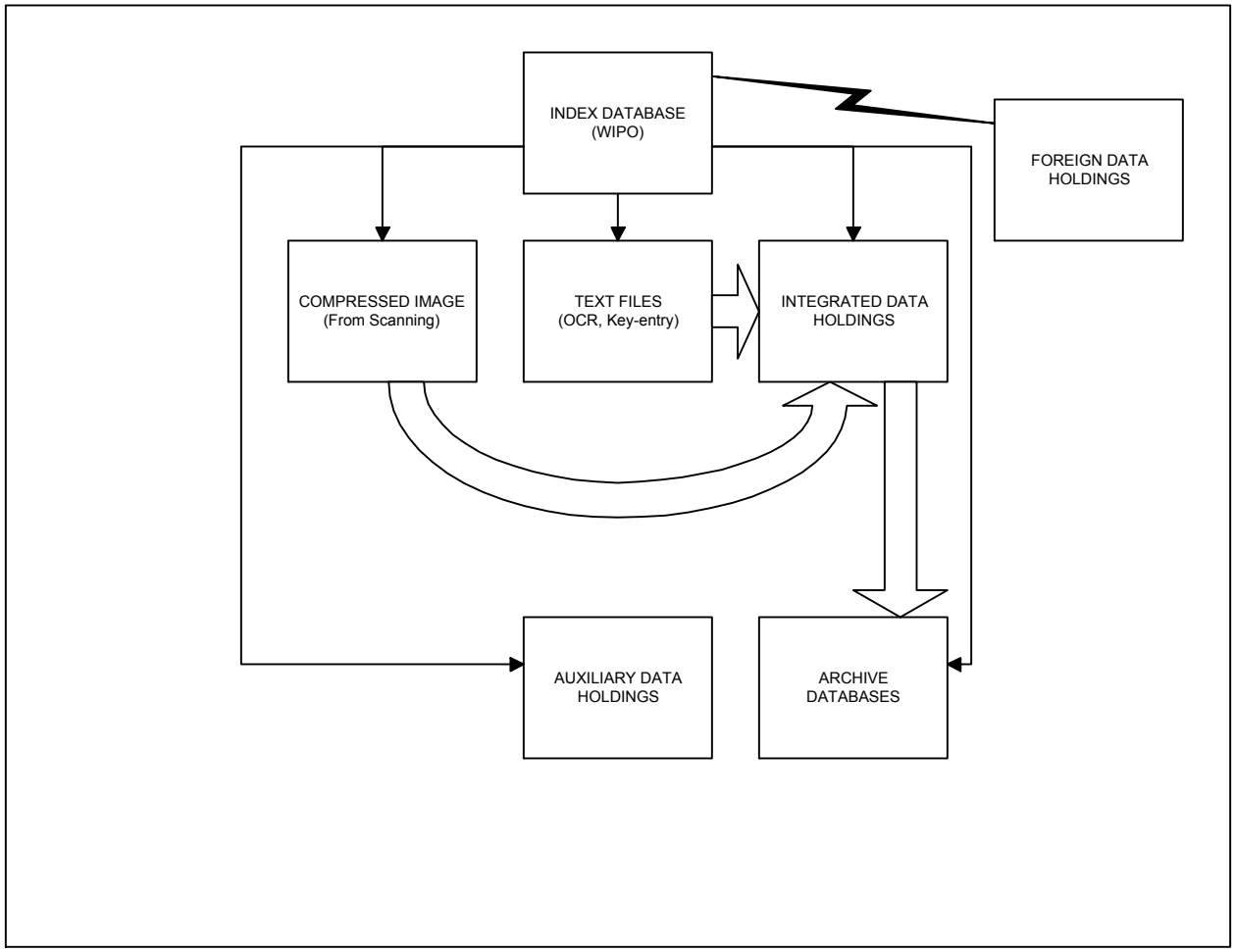
5.1.1.3. Data Representation

According to the findings summarized in Chapter 4, once a decision is made about the documents to be included in the database, the most critical feasibility question is that of data representation. The CPO has determined that a database containing the text of documents only would not suffice. Key portions of a patent document would be lost, such as drawings. The other two options are image storage, and an integrated text and image database. Although the latter may represent technical difficulties, none of these seem critical for the successful completion of the project, and overwhelming benefits are provided. Among other such benefits, the costs of implementing an integrated text and image system would be less compared to a system holding images (in particular, because of the enormously reduced requirements for data storage). Some functional benefits which are unavailable with an image system, are also offered (e.g., full-text search capability).

The actual level of integration in a text and image database will depend on feasibility factors, economic benefits provided, and the availability of de-facto or de-jure international standards. The quality of data conversion plays perhaps the most significant role among technical feasibility

factors. Although the automatic text conversion of paper documents using OCR technology is generally feasible, it has not yet been proven that the resulting data are of the required quality for successful integration. This fact is one of the most important reasons for the scenario proposed later in this chapter.

Figure 5.1. is a schematic diagram of the proposed databases. Canadian pending applications and in-force patents, and the title pages of US in-force patents, eventually in integrated text and image format, are stored in an on-line system. This document files are created by scanning, OCR, and manual key-entry operations. An archive includes older Canadian patents as well as applications correspondence. Auxiliary data holdings include other foreign material in the CPO's posession, for example. All documents are controlled by, and retrieved through, an index database which is implemented in accordance to WIPO guidelines. Telecommunications access is provided to the electronic data holdings of foreign jurisdictions.



5.1.2. MIR (Management Information Reporting) System

Solutions exist for the traditional MIS functions included in this proposed system component. Activities tracking/scheduling, user accounting, and access security are software problems well researched.

Support to project management is of an entirely different nature. This project is very complex, and involves a considerable number of risk factors. Project management must be capable of responding to unforeseen situations quickly and firmly. Thus there is a need for sophisticated management support tools to be employed during the project period.

More and more computerized information engineering tools exist, providing support to all levels of system development. They help management to notice and resolve ambiguities during systems and software design. They also provide tools for strategy decisions, prototyping, modular development, end-user involvement, system integration and other key areas. The most important benefit of employing such tools is that they ensure a consistent system in every respect.

5.1.3. FAS (Financial and Administrative Support) System

The functions of this component, as described in Chapter 3, do not require special considerations. Emphasis should be taken to develop information interchange between this and the appropriate portions of the MIR system. The MIR system is capable of providing the necessary information for this component.

5.1.4. UST (User Support and Training) System

In a system of this complexity, uniformity of user interfaces and supporting tools should be rigidly enforced. The automated design tools, mentioned in Section 5.1.2., would be of great value in this respect. A training environment can be created using the prototype system(s) constructed to support design. These arguments support the recommendation for employing the state-of-the-art information and software engineering tools in the development.

5.2. The Preferred Solution - Alternatives Set Aside

In brief, the preferred automation solution is based on text and image databases which are integrated to a level as defined by feasibility factors. The database includes Canadian in-force patents and the title pages of US in-force documents in the core on-line data holdings. Convenient access to other documents is supported by off-line or near-line systems. These documents may be stored on analog media (incorporating presently existing roll-films, if possible), or digital media (for documents archived from the automated system, in digital form). Two internal user groups are identified, examiners and administrative users. Examiners are provided with state-of-the-art high-resolution graphic workstations, administrative users require full-featured office automation equipment. External users can connect to the system by the

provided telecommunications link, probably controlled by a front-end computer to prevent unauthorized access. Telecommunications connection to other jurisdictions and regional offices is also provided. The system will also support paper-based dissemination functions, in the form of quality publications.

Significantly different alternatives were considered in earlier planning studies. Some of them, for example, incorporated large digital image based document files, or various different document holdings. A range of these alternatives were eliminated based on the reasons described above.

Typically, feasibility studies develop a number of mutually exclusive alternatives in parallel. A cost and benefits analysis is then performed to derive a preferred design. In studies preceding the development of the present plan, four general alternative solutions were considered. These alternatives differed in the amount of documents included in the database, and the telecommunications access provided.

A generalized illustration of significantly different alternatives involving two different technologies to implement two different databases is presented in the following table.

Technology	Definition of Database	
	Canadian In-force + US Title Pages	Can. In-force + US In-force
Integrated Text and Image	$65M	$165M
Compressed Image	$100M	$200M

Table 5.1. Automation Alternatives - Cost Comparison

Integrated text and image data holdings are envisioned on on-line equipment (magnetic or optical), while slower equipment was considered (e.g., optical jukeboxes) for the compressed image solutions.

The estimate of total project costs for the integrated text and image database of Canadian patents ($65M) is developed in detail in the following chapter.

The estimate for the compressed image solution involving Canadian patents and the title pages of U.S. in-force patents is based on the estimate introduced above, which was increased to support a) the necessary storage and increased CPU throughput capacity (partially compensated by the lower price of 'near-line' equipment: $20 million), b) the need for a high speed local area network solution, probably involving fibre-optics technology ($10 million increase), and c) the need for workstations with high storage and networking capacity ($5 million increase).

The estimate for the integrated text and image solution involving all Canadian and U.S. in-force patents ($165M) reflects the increased database sizes which result in increased costs for storage capacity ($50M), data conversion workload ($35M), and database updates ($15M).

Finally, a design alternative encompassing all Canadian and U.S. in-force patents in compressed image format was considered.Its estimated cost is based on further proportional increases for a) storage capacity (in part 'near-line' equipment) and hardware for additional system throughput ($70M), b) backfile conversion workload ($35M), c) database updates ($15M), d) workstation capabilites ($5M) and e) local area network speed ($10M).

The integrated text and image solution ($65M), provides maximum benefits at minimum cost.

5.3. Definition of the Preferred Solution

5.3.1. The Preferred Automated System

The considerations presented above clearly define the preferred path for automation. Major system components are defined, solutions for implementing them are identified. Figure 5.2. provides an overview of the system, identifying major system components and data holdings.

The preferred system would incorporate in-force Canadian patents and the title pages of in-force US patents in its core database. The data representation would be in integrated text and image form. Preferably this representation would suffice for the identical reproduction of the original documents.

It is recommended to include in this database the collection of Canadian patents from patent number 1,036,300 (granted August 15th, 1978) to the most recently issued patent and all future patents while they are in-force. By the end of the project period, this will represent a full database of in-force Canadian patents. The selection of this starting date is also fortunate, since the CPO started to apply the IPC system at that time - therefore no significant reclassification efforts will be necessary.

It should be noted, that the planned telecommunications facilities provide the technical means for accessing foreign data holdings; in case an agreement is achieved between the CPO and these foreign offices, those data holdings may also become available.

Application files with full contents (all correspondence) will also be stored in the core data holdings. Because of its nature, these files must be stored on erasable (magnetic) media, to permit changes and additions. Patent files can be stored on write-once media; the rare event of a modification to a patent will not present a significant storage overhead (the new - modified - file will also be stored and the old one invalidated in the indices but still occupying space).

The volume of Canadian in-force patent files will be approximately 75 Gbytes in integrated text and image format (Appendix A). The number of pending applications (presently approximately 90,000, not expected to grow over 100,000) represents a requirement for 34 Gbytes.

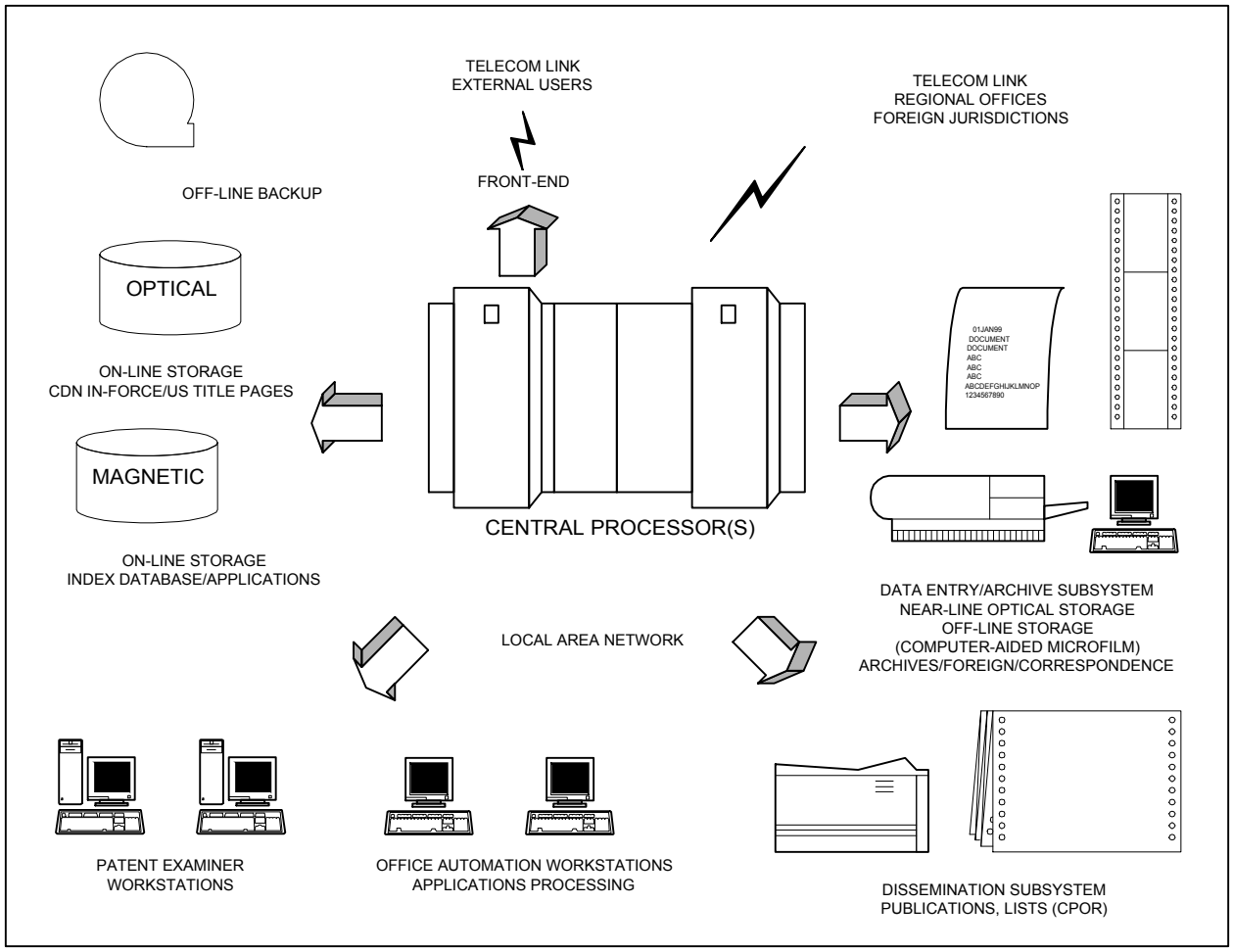


Figure 5.2.

An index, preferably identical to the IPC, would be employed for two purposes - as a primary way to organize the database and as a search tool. The efficiency of the index as a search method would depend on the quality and up-to-date nature of the index, but its function as a database organization tool would not suffer, even if errors (e.g., misclassification) are present. Inconsistencies in the index data will be pointed out by the system when the database is loaded, so that existing ambiguities can easily be resolved.

Since the index file can represent a much greater number of documents than those actually stored, a storage capacity of 10 Gbytes is assumed to be required.

Based on the above, the core database will be stored on 45 Gbytes of magnetic and 75 Gbytes of on-line (mounted) optical media. No storage overhead is included for indices required by the full-text search system. If that represents a 100% overhead on the text portions of the database, it is approximately 16 Gbytes for the 500,000 documents. This overhead is compensated for by a.) the conservative estimating method employed in the volumetrics assumptions, and b.) a conservative price estimate in the costing exercise.

Archive patents and other documents (correspondence material of in-force patents, foreign documents, abandoned applications, etc.) will be stored in a separate subsystem. The speed requirements for this subsystem are much lower, permitting the use of near-line or off-line media. Existing data holdings should be incorporated, whenever possible. For example, a computer-aided microfilm retrieval device could incorporate good-quality microfilm holdings. This subsystem would therefore consist of analog and digital storage devices - patents archived from the electronic database in the future should be stored on the latter. Regardless of the nature of the media, information would be sent to the main system in digital form, so it can be viewed on electronic workstations.

This archive database may be integrated with the data-entry subsystem. Some processes are common to both, such as the digitization of a paper or film based document image. Maintenance of the archive data holdings would likely require similar equipment to that used for data entry. This integration is feasible (known products exist) and may prove to be a method for cost reduction.

Documents entered in the system will be scanned and digitized first; OCR processing will follow. Different architectures can be envisioned to implement this stage. In one solution, the OCR process can be incorporated in the scanner equipment. Known products exist with excellent performance with respect to the accuracy of character recognition. It is questionable, however, whether these products can incorporate the results of rapid development in this field of technology. Therefore an alternative solution should also be considered, in which the OCR processing is implemented by software, running either on the central processor(s) or on specialized equipment. In this alternative, scanned images of documents are temporarily stored and OCR processed later in batch, according to system utilization schedules. The selection of an alternative should be performed before specifying systems equipment requirements, because the second alternative may represent a significant additional workload on both the processor(s) (OCR software), and the network equipment (optional traffic of image-form data).

To ensure the quality of the resulting text files, format requirements may be implemented for the paper files submitted by the applicants. Such requirements would result in virtually 100% accurate text conversion.

The central processor(s) would perform the task of database management and maintenance. Other system components (MIR, FAS) will also be implemented as software running on the central processor(s). The task which requires most consideration is full-text search on the text portion of the database. Although tools of excellent performance are either in existence or rapidly evolving, all known solutions present additional requirements. On the one hand, software products of excellent performance are known. These products, however, generate a significant storage overhead (15-100%) for the generated indices or inverted databases. On the other hand, hardware solutions usually require the distribution of the database which, in turn, may represent additional problems of database management and security.

Information dissemination will be supported in two ways. First, telecommunications access will be provided for external users. They will be able to retrieve any (non-confidential) document and

perform all implemented search functions. Since the physical separation of confidential (application) and other (patent) data bases would seriously affect the functionality of the system, a front-end computer may be necessary for preventing unauthorized access to confidential material. Telecommunications costs with respect to large file transfers prevents the distribution of graphics data.

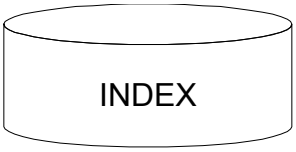
Second, paper-based dissemination methods will also be supported. Some publications (e.g., CPOR) can be generated in a semi-automatic method.

5.3.2. Automation Options - Implementation Strategy

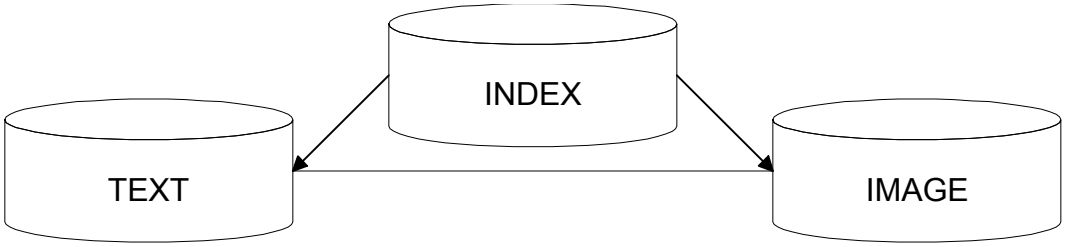
As a refinement of the traditional approach, specific criteria are introduced to develop four new and specific options all of which are related to the preferred general design. As a departure from the traditional approach, we recommend the implementation of all four options since they form a progressive sequence of fall-back systems. The definition of the four automation options follows.

Figure 5.3. is a symbolic comparison among these options. Option 1 is the most simple, implementing only a sophisticated index database serving as an index and classification search tool, and a tool for paper-based document management. Option 2 implements text and image databases controlled by this index database. In this case, text and image holdings are handled separately. A functional level of integration is achieved in Option 3. With this system document pages closely resembling the original can be re-created on workstation screens. 'Absolute Appearance Fidelity' is achieved in Option 4; in this case strict boundaries between text and image holdings do not exist any more. In any respect, text and image portions of document files are fully integrated.

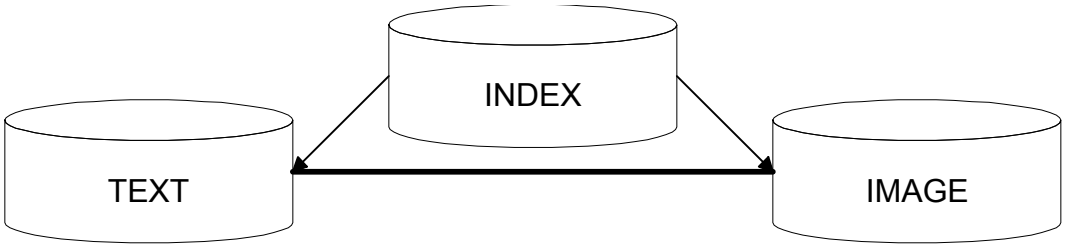
OPTION 1: INDEX DATABASE



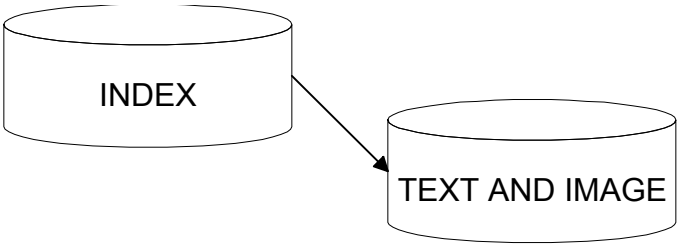
OPTION 2: SEPARATE TEXT AND IMAGE



OPTION 3: FUNCTIONAL INTEGRATION



OPTION 4: FULL INTEGRATION



5.3.2.1. Option One: The Index Database

The base design will consist of the index database and common access and control facility of the above options. The database will be based upon international standards (WIPO) including traditional (Canadian) indexing and classification information as a subset.

This design will provide essential improvements in document management and workflow processing.

Search capabilities on indices, and classification search can be implemented on this system.

The index database will be essential for the other three alternatives as well, providing a central database management, workflow management, and data access facility.

The design of the index database would take place in the first phase of development, since all other databases would be indexed and accessed through it. It would require the minimum of resources and would be capable of operating as a stand-alone system during the development and testing of higher level components. The development and implementation of this system is considered feasible by reference to the system now in operation in the Spanish Patent Office. This system does not meet the mandatory requirements of the CPO.

5.3.2.2. Option Two: Separate Text and Image Files

This option is a system with separate text and image files accessible through the common index database. The database of the full text and drawings of Canadian in-force patents and pending applications, augmented for search purposes with the abstracts and first drawings (title pages) of in-force US patents will be stored on-line. Scanned image files of text pages are maintained off-line for reference as are backfile documents. Scanned images of text pages are retrieved through the automated system using batch scheduling, in case they are needed (e.g., a page rendered incomprehensible by the loss of embedded images or formulae).

On the text data holdings, an implementation of full-text search capabilities is feasible.

This system satisfies the mandatory requirements but it provides limited benefits. It will permit further development towards a level of integration of text and image files.

5.3.2.3. Option Three: Functionally Integrated Text and Image System

This option consists of an integrated text and image system with an on-line database of Canadian in-force patents and pending applications, and augmented for search purposes with the title pages of in-force US patents. In this system, identical reproduction of original (Canadian) documents is not supported. Scanned image files are maintained off-line for reference as are backfile

documents. Scanned images of on-line files and backfile documents are retrieved through the automated system using batch scheduling.

As in Option Two, full-text search capabilities will be implemented on text data holdings.

Should the opportunity present itself, the maintenance of scanned image files will permit the construction of a fully integrated text and image database at a later stage.

The results of the technical research summarized in Chapter 4, conclude that there are no major questions of feasibility with respect to the functional integration of text and drawing pages. It remains to be determined whether or not embedded images within text pages would significantly improve the effectiveness of the system and whether the functional integration of embedded images within text pages is feasible. The degree of functional integration, and therefore the commitment to the level of implementation, will be determined by a pilot study and by the level of implementation in foreign offices.

5.3.2.4. Option Four: The Preferred General Design (A Fully Integrated Text and Image System)

The preferred general design consists of an integrated text and image system with an on-line database of Canadian in-force patents and pending applications, and augmented for search purposes with the title pages of in-force US patents. The backfile documents are maintained on an off-line system and retrieved through the automated system using batch scheduling. In this system, integrated text and image files are 'perfect': identical reproduction of the original documents is feasible.

Among other issues, critical technical factors of the greatest importance are: a.) the success and quality of data conversion, and b.) the timely evolution of successful international standards.

The full integration of text and image holdings becomes possible, when a.) standards have fully evolved, and b.) OCR technology for this purpose has matured. Both events are expected to happen during the project period (1990-92). The implementation of full integration therefore does not depend on in-house R&D efforts. The cost increment over Option Three also seems to be minimal, once the technology and standards are given. These reasons call for a plan targeted for Option Four, keeping the other options as fall-backs.

5.4. Conclusion

In summary, the proposed automated system would incorporate databases of Canadian in-force patents, pending applications, and title pages of US in-force patents. A slow archive database would handle other material, possibly in its present, microfilm form. The system would provide search capabilities on indices, classification search and full-text search where applicable. The development effort is targeted for a fully integrated text and image data representation.

Automation alternatives incorporating significantly more documents or different data representation technologies were discarded for economic reasons, and due to the technical risk associated with distributed databases and networks with very high throughput.

Within the preferred solution, four automation options were identified. Except for Option One, which does not include document databases, they differ in the level of integration of text and image files. Each option is included in the next, therefore an incremental approach is proposed to minimize risk factors. In this approach, each successfully implemented system is a fall-back for the next level - should the development effort fail, the system of the previous level will remain in operation.

This approach ensures the full protection of investment. At any point, the equipment purchased is fully utilized by the system in operation. For example, the storage requirements for Options Two, Three, and Four are roughly the same, therefore the purchase of storage devices does not imply a premature commitment to the preferred option. In case of a failure, only a limited development effort is wasted.

Pilot studies to isolate singular questions of technical feasibility always precede the commitment to full system development activity, which in turn involves an extensive testbed quality assurance evaluation prior to commitment to and deployment of resources for the operations environment. In summary, the recommended solution is accompanied by low, and easily controlled, financial and technical risks. At the same time it is a very advanced solution in terms of its utilization of expected advances in standards for electronic document interchange and search technology. The solution satisfies the mandatory principles of system development and implementation which were set out in Chapter 1 of this document.

6. PREFERRED SOLUTION - COSTS AND OPTIONS

In the previous chapter, the crucial choice among major alternative solutions to the problem of Patent Office automation has been set out. A benefit-cost rationale was presented to justify the decision to set aside three alternatives and develop the preferred solution in depth. Essential technical features of this solution were described. This and subsequent chapters of this document are devoted to cost analysis, a technical description of the recommended solution, and a summary of its benefits. The technical concept of the preferred solution is incremental, in effect providing four options. In this chapter, the estimated project costs are introduced, beginning with option 4, which comprises the others. Then the other options are described in terms of cost and benefits, and compared.

6.1. Project Cost Estimates

The attached financial table contains cost estimates for the preferred solution. The format of presentation is that chosen for presentations to central agencies. With minor exceptions, it is also the format in which the study team for general design and requirements had previously assembled preliminary estimates. The table is accompanied by two summaries. A number of more detailed cost studies were performed. One of these is included in Appendix C.4. That costing study shows all eight years of the project history, and it is based on a particular path of implementation. The annual totals and major line items agree with the estimates discussed here. All major line items were verified with vendors. Others were estimated by project teams in preparation for the present report. This section provides a brief summary of the rationale for each cost item. In order to support a comparison of options (section 6.2) without repetitive discussion, each paragraph describing a cost item is followed by a brief explanation of how the other options differ in terms of cost.

CPO Automation Project Cost Estimates January 22, 1988

	1988/89	1989/90	1990/91	1991/96	Total
1. Salaries - ASB Staff	600	900	1400	7000	9900
2. Project Management	550	890	600	1000	3040
3. Hardware	150	8500	5000	3000	16650
4. System Software		500			500
5. Appl. SW Development	100	800	710	4475	6085
6. Implementation		50	50	250	350
7. Facilities/Preparation	180				180
8. Backfile Conversion	500	2600	2000		5100
9. Extern Database Access Setup	10	50			60
10. Marketing & Promotion				750	750
11. File Conversion	10				10
12. Operations Management	110	100	370	2183	2763
13. Computer Operations		200	600	3000	3800
14. Hardware Maintenance			510	4122	4632
15. Communications	10	10	10	1500	1530
16. Software Licenses			100	500	600
17. Extern Database Usage				300	300
18. Appl. SW Maintenance		150	300	1500	1950
19. Training & User Support		25	25	300	350
20. Facilities		250	250	1250	1750
21. On-going Data Capture		970	510	2550	4030
22. Publishing					0
23. Translation			100	500	600
Totals:	2220	15995	12535	34180	64930

Summary

Equipment & Maintenance	150	8500	5510	7122	21282
Backfile Conversion	500	2600	2000	0	5100
On-going Data Capture	0	970	510	2550	4030
Software	100	1500	1160	6725	9485
Facilities & Training	180	275	275	1550	2280
Operations & Communications	10	210	610	4500	5330
Project Consulting Support	550	890	600	1000	3040
Salaries	600	900	1400	7000	9900
Operations Management	110	100	370	2183	2763
Other	20	50	100	1550	1720
Total:	2220	15995	12535	34180	64930

Summary 2

Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	23680	37100
Total:	2220	15995	12535	34180	64930

1. Salaries - ASB Staff

   The first item covers salaries for ASB staff. The proposed staffing plan contains a strength of twelve person years for 1988/89, followed by eighteen and twenty-four, and twenty-eight in 1991/92, which is the full strength of the permanent ASB organization. Estimates are based on an average salary cost of $50,000 per person year. An extra $200,000 was allocated for 1990/91 to cover temporary needs in the transition from major development and installation of the hardware to operation of the automated patent system.

   These costs apply to options 2-4; person year requirements are reduced to eighteen in option 1.

2. Project Management

   This item contains contracts for consultants to assist the growing ASB organization in project management. Figures are based on present levels and estimated future needs. The average consultant contract is assumed to be $100,000 per person year.

   These costs are reduced by 25% and 50% for options 3 and 2 respectively, in the last five years of the project. They are reduced by 50% in year three, and are eliminated completely in the remaining years for option 1.

3. Hardware

   Cost estimates are based on non-binding vendor information available in January, 1988. The costs include all items, from the central processors to storage devices and workstations. There is no assumption that commits the project to a particular hardware solution. An assumption was used on expected price drops of storage media and workstations. These assumptions are considered safe as they were based on several sources.

   These costs apply to options 2-4; costs were reduced in accordance to the limited requirements for option 1.

4. System Software

   The estimate represents the initial licence fee for operating systems and all essential utilities for managing the EDP resources; it is based on a vendor quotation.

   This cost applies to all options.

5. Application Software Development

Based on estimates for systems development contracts and some purchases of software, this line item reflects the plan for software development and a preliminary staffing plan. Each major component of software is assumed to require a team of up to four persons.

Year four to year eight activities are reduced by 25% and 50% for options 3 and 2 respectively; for option 1, year two activities are reduced to a half, the rest is eliminated.

6. Implementation

This item reflects cost estimates for system installation, installation of the databases, and system integration and test.

These costs apply to options 2-4; in case of option 1, activities after year two are eliminated.

7. Facilities - Preparation

This item contains costs for building a computer facility of 5,000 square feet and additional preparation costs for 7,000 square feet of space for staff, contractors, and training needs.

These costs apply to all options.

8. Backfile Conversion

Costs are based on estimates prepared by vendors, followed by tests, for the supply of OCR contract services for conversion of 250,000 Canadian patents.

These costs apply to options 2-4; option 1 includes costs for extracting data for the index database, and building a database of abstracts.

9. External Database Access - Setup

Software purchase and installation for accessing commercial databases.

These costs are common to all options.

10. Marketing and Promotion

Marketing and user education is an essential part of a more active dissemination strategy. It was assumed that five dissemination 'products' will be defined, launched, and publicized, and that the project contains contracts for each of these in the amount of $150,000.

For options 3, 2 and 1, the number of such products was assumed to be four, three and one, respectively.

11. File Conversion

This line item includes costs for acquiring and converting patent files from other jurisdictions, for example, the USPTO. Due to lack of supporting information, only a pilot project was planned.

This cost is common to options 2-4; for option 1, it was eliminated.

12. Operations Management

The item covers contract resources for management of computer operations and software maintenance, that is, all tasks in support of the continuing operation that are not in the hands of ASB staff, and not covered by staff of a facilities management firm. Strengths are shown in the staffing plan.

These costs are common to options 2-4; activities are reduced to one half in year three and one fourth in remaining years for option 1.

13. Computer Operations

Cost estimates are based on planned requirements for operators and consumables for running the data centre, for example, under a facilities management contract.

These costs apply to options 2-4; activities (due to a smaller data centre) are reduced to one third in year three and after for option 1.

14. Hardware Maintenance

Estimates are based on vendor quotations. Where these were not provided, a rate of six percent per year of the hardware price was assumed. Maintenance charges begin in the year after the hardware is installed.

Costs are common to options 2-4; reduction has taken place (due to less hardware) for option 1.

15. Communications

This line item is based on a detailed study of telecommunications costs for delivery of patents to users across the country. Volumes are based on the market survey performed in 1987. For the first three years, a pilot project was assumed, and estimated at $10,000 per year.

These costs apply to options 2-4; operational costs (year 4 and onwards) are reduced to one tenth for option 1.

16. Software Licences

Continuing fees for commercial software acquired under licence are covered here. Most of the applications software will be custom written, and will not attract licence fees. However, such software must be maintained (item 18 below).

These costs apply to all options.

17. External Database Usage

This item represents an estimate of a moderate increase in usage of external database services by examiners in the period after major parts of the system are installed.

These costs apply to all options.

18. Application Software Maintenance

Major in-house development of software must be followed by continued effort for maintenance and update. It is assumed that a staff of up to three maintenance personnel is hired under contract.

For options 3 and 2, activities in the last five years are reduced by 25% and 50% respectively; for option 1, a significant reduction was applied.

19. Training and User Support

The estimates reflect assumptions in the staffing plan on needs for training of CPO staff and continuing support for users in the public search room as well as in order fulfillment. Estimates are based on contract resources and materials.

These costs apply to options 2-4; efforts in the last five years are reduced to one third for option 1.

20. Facilities

Ongoing costs in this area are based on the gross rental for eighty percent of a floor for use by ASB.

Costs apply to options 2-4; floor requirements are reduced (60%) for option 1.

21. Ongoing Data Capture

The estimate reflects quotations and tests for OCR conversion of 35,000 applications per year, using contract resources.

These costs apply to options 2-4 and are eliminated for option 1.

22. Publishing

No costs for publishing are assumed. The computer system will produce a form of revised CPOR as a benefit. It is assumed that variable costs of printing and distribution will be recovered through sales.

These assumptions apply to all options.

23. Translation

Translation of manuals and user documentation is included. The estimate assumes a price of $0.20 a word and 500,000 words annually.

Costs apply to options 2-4; translation costs after year two are eliminated for option 1.

The financial table consists of three parts. Part one gives a breakdown of principal cost items. The first item shows salaries for ASB staff. Items two to eleven are one-time costs during the

project period, and items twelve to twenty-three contain operational costs which are expected to continue after the project period. The breakdown in part one was used for systematic development of costs in the feasibility study. Part two of the table shows another breakdown of the estimated project costs by major categories for control by management of project costs. Part three of the table contains a further breakdown into three categories for use in the submission to the Treasury Board, showing salaries of ASB staff, capital costs and O&M costs. Salaries are as shown in the more detailed part one, capital costs are composed of hardware (item 3), system software (4), facilities/preparation (7), and software licences (16); all other costs are grouped in the O&M category.

6.2. Options - Comparison of Cost and Benefits

As was explained in Chapter 5, the preferred solution to the automation task of the CPO consists of four options. These are incremental in the sense that each higher option includes all functionality of the next lower option, and adds one or more functions of value to the CPO and the user community. Since there is a degree of technical risk associated with the higher options, the options have been distinguished. The project management can thus assess when and under which conditions the development should best be advanced to a higher level option. This section compares expected costs of the options, and provides a comparison of advantages and limitations for each. The following financial table gives a summary cost comparison. More detailed spreadsheets in the format introduced in section 6.1 are in Appendix B.

CPO Automation Project Cost Estimates January 22, 1988

	1988/89	1989/90	1990/91	1991/96	Total
Option 1: The Index Database					
Salaries	600	900	900	4500	6900
Capital	330	4500	1100	500	6430
O&M	1280	2425	1260	4747	9712
Total:	2210	7825	3260	9747	23042
Option 2: Separate Text and Image Files					
Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	19742	33162
Total:	2220	15995	12535	30242	60992
Option 1: Functionally Integrated Text and Image					
Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	21786	35206
Total:	2220	15995	12535	32286	63036

Discussion

Option 1 - Index Database

This option consists of the development of an index database for all Canadian patents with appropriate classification information, and including the abstracts. U.S. abstracts are included as well. There will be on-line access to this database for CPO staff and access for external users. The other major component will a document tracking system that is complete except for applications. The system does not support full text of patents or drawings. Therefore, there can be no delivery by other than manual means of full documents to external users, and, of course, no full text search on patents. The system does provide text search on abstracts.

Essentially, this system provides a basic functionality which would improve the efficiency of patent office operations and support some additional service to outside users. However, examiners and external users would require access to paper documents as before. This option would not meet certain requirements accepted as mandatory, mainly in the areas of full text search with delivery of images to the workstation and document delivery through telecommunications. The option would not support document exchange with other jurisdictions.

The estimated cost of option 1 is $ 23 M., and savings would occur in eighteen of the twenty-three cost items discussed above. The size of ASB staff could be restricted to eighteen person years, as a much smaller data centre and software development effort is required. This option is provided in case of delays in major data conversion efforts, or in order to benefit from technical opportunities not presently seen. However, in that case there will be inefficiencies in investment, for example, in that a less than optimal hardware configuration might be purchased.

Option 2 - Separate Text and Image Databases

This option meets all mandatory requirements for processing, data holdings and storage, and telecommunications. It is based on the concept of separate files for text data and images. The text presented at the workstation would have no relation in its layout to the format of the typed page. Text portions of documents cannot be presented at the screen with codes tying it to images, although related images could be presented. This option was formulated against the assumption that an attempt to proceed to option 3 has been made and did not succeed. This is the least technically ambitious option that meets mandatory requirements.

The cost of this option is $61 M. The estimate includes an allowance for software development effort towards the ITI solutions which has to be set aside. The option can be upgraded without loss of investment.

Option 3 - Functionally Integrated Text and Image

This option meets all mandatory and most desirable features. It does not support full restorability of documents to the form in which they were scanned initially. Thus, there are restrictions in the capability of this system to support dissemination functions which involve document delivery to outside users, and the examination step may require access to paper documents. The identity of the electronic copy with the paper copy cannot be assured - even if the legal question of equivalence had been settled.

The cost of this option is $63 M. Differences in cost are in a lower cost of software development and some derived costs for software maintenance and dissemination. The option can be upgraded to option 4 without major incremental costs. This option is satisfactory in case the event that the fully integrated text and image solution might not be attainable, or is delayed due to slippage of adoption of standards.

Option 4 - Fully Integrated Text and Image

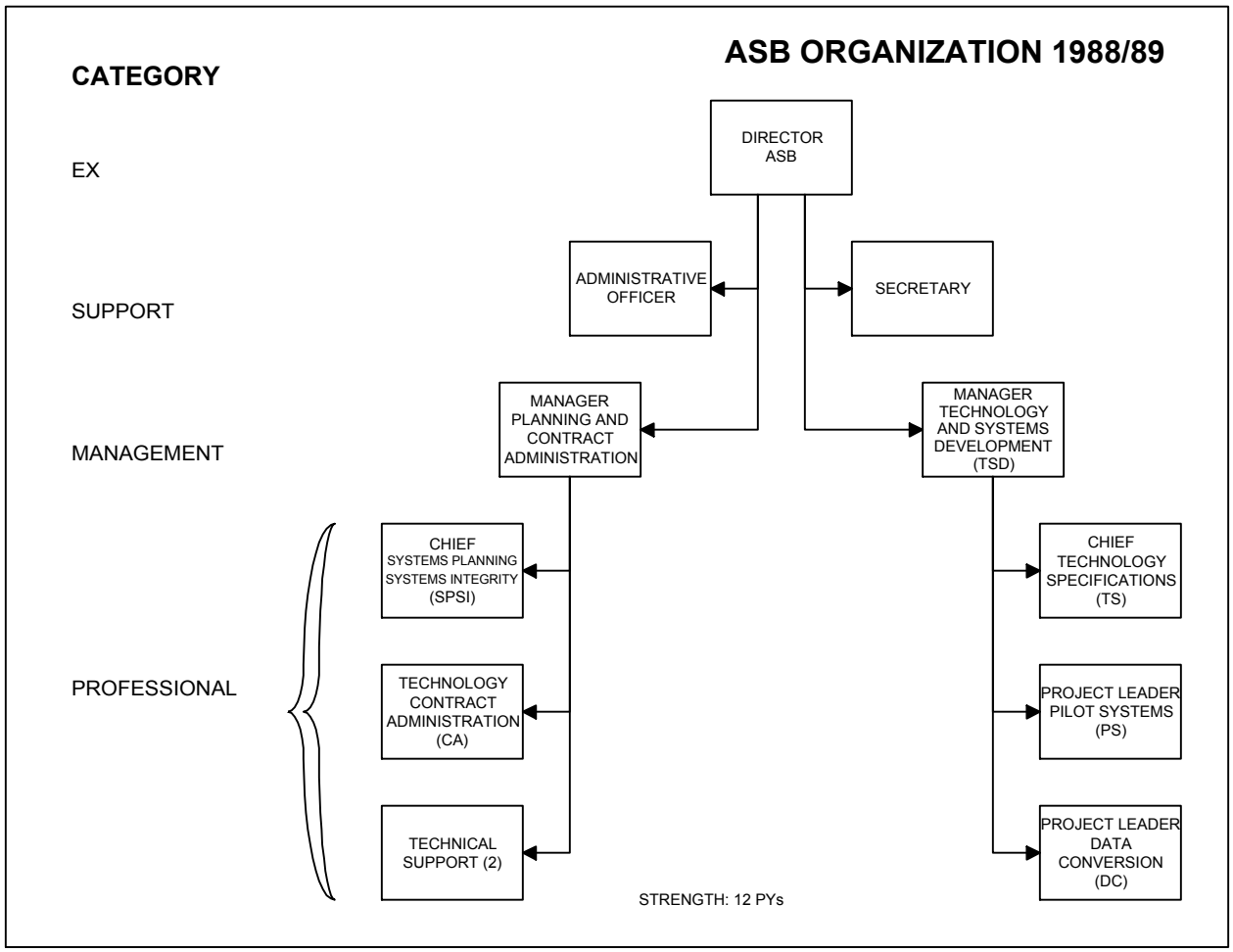
This option meets all requirements described in this document as mandatory and desirable. Its estimated cost has been described. It is technically feasible with present technology, that is, even without awaiting generally accepted standards specific to the patent information systems environment. However, the implementation of this option may be delayed at the discretion of ASB if attractive standards are delayed, or if foreign data holdings offered to Canada should not meet the high standards proposed for this option, or finally, if unforeseen problems should arise, for example, in the area of regulating the format of submission of applications. This option is compatible with holding portions of the CPO database in a format at the level of option 3.

We conclude that option 4 should be developed as the target design, since the other options either provide unacceptable compromises in functionality (option 1), or provide less than full functionality at a price which is not sufficiently lower to warrant their choice as the target system.

7. IMPLEMENTATION

7.1. Preliminary Action Plan for 1988/89

This section presents a preliminary plan for year one of the project. Its purpose is to demonstrate that ASB has a realistic prospect of successfully completing its proposed first year activities, given its present strength, the proposed contract resources, and a timely approval of the Automation Plan. The first-year plan consists of a chart showing a proposed organization of ASB, followed by a brief summary of principal responsibilities, and a GANTT chart of activities with an explanation of critical project control points and their linkages.



7.1.1. ASB Organization 1988/89 - Principal Responsibilities

Director, ASB: Overall management and project responsibility. Supervision of activities 1-5 in Table 7.1. Direct responsibility of activities shown under 6 in Table 7.1. This includes securing management approvals for the action plan, and maintenance of the Automation Plan, organization and staffing, staff development and training, and negotiation with foreign

jurisdictions, as well as relations with user and advisory committees. Final recommendations to management on all contracting decisions.

Manager, Planning and Contract Administration (PCA): Maintenance of yearly systems plans, and update of the Automation Plan. Supervision of RFP development, contract administration and performance control of contractors. Liaison with DSS. Liaison with user community. Responsible for integrity of databases and systems, as well as physical and EDP security. Supervision of related staff.

Manager, Technology and Systems Development (TSD): Responsible for all technical aspects of the systems development process. Tracking of technology and updates of technology portions of the system plan and Automation Plan. Direct responsibility for quality assurance, the execution of pilot projects, and RFP development. Liaison with vendors before and during proposal evaluation, including benchmarks. Technical performance control of contractors, and acceptance of goods and services. Development and update of specifications. Supervision of related staff.

Administrative Officer: Tracking of expenditure against plan and contracts in place. Assistance to the Director on all reporting duties and control functions.

PROJECT PLAN 1988/89

#	ACTIVITY	RESPONSIBLE	1988 A M J J A S O N D	1989 J F M
1	Detailed Requirements	PCA	========2	
2	Detailed Systems Plan	TSD	========9	
	Quality Assurance		o=o o=o	o=o o=o
	Facilities		o========o	
	Security		o===o	
3	Pilot: Backfile Conv.	TSD/DC	spec RFP eval o====o==o====o====4	o...o=5
4	Pilot: Prototype	TSD/PS	spec RFP eval o====o==o=======6=====	===o
5	RFP Development	TSD	RFP prep (prop) o========o.....	8=======o=====8
	Hardware Facilities Conversion Software Other		prop eval 7== ===o=====7	
6	Management Action Plan	ALL	=1	o=o
	Organization		o=o o=o	o=o
	Training		o=o o=	o o=o
	Negotiations /Dt Xchng		o==========9	

Legend:

=== : task
... : task suspended
=o= : milestone
=3= : major milestone

Table 7.1.

7.1.2. Project Plan 1988/89

Summary of critical project control points:

1. May 1: Action Plan for the fiscal year finalized and approved by management. Reason: commitment to pilot projects is required if RFP for major items (5) is to be issued, and proposals received, during the fiscal year.

2. and 3. June 30: Detailed requirements and detailed systems plan must both precede the security plan, RFP preparation, and international negotiations for terms of data exchange.

4. October 31: The pilot project on backfile conversion should have all essential conclusions available before the major RFP can be released.

5. March 31, 1989, is proposed as the latest date for a decision on how to handle the conversion of current applications. This takes into account the presently expected date of start-up of Bill C-22 implementation (January, 1989), and allows for some slippage in that date. The decision to be taken involves adoption of a proposed standard format of submission of applications, and possible start-up of a secure component of the data conversion facility. Critical preceding activities are the prototype of the application processing system (activity 4), agreement of the patent agent community, or the prior release of US standards for submission.

6. October 31: Pilot project for application processing ready for evaluation. This point can be advanced by two months if an RFP for the pilot can be avoided. In that case, the major RFP under activity 5 can be issued on September 30, as opposed to January 1, 1989.

7. November 30: Release of RFPs for major acquisitions in 1989/90. On the assumption that the proposal period is two months, and evaluation takes two months, that date assures a decision by fiscal year end.

8. January 1: Release of RFPs if preceding projects, mainly activities 3 and 4, are delayed. In that case, a decision on major RFPs can be expected by May 31, 1989.

9. November 30: Conclusion of negotiations with other jurisdictions over terms of data exchange. That date is critical as the results of such negotiation might affect the terms of reference in major RFPs. At that point, at least an agreement in principle should be reached on the supply of US title pages, on supply of Canadian documents in electronic form, and on other matters which might affect the major purchasing decisions in 1989/90.

7.2. Project Implementation Schedule

This section describes the major activities and their scheduling for the implementation phase. It is assumed that the implementation phase of the project will formally commence in 1988 and is targeted for completion on March 31st, 1996. Following completion of each project year an annual review and update of the project plan should take place. The following table summarizes major activities which are explained in deeper detail in the subsequent sections.

PROJECT PLAN 1988/96

#	ACTIVITY	88 89 90 91 92 93 94 95 96
1	Organization	==o==o..1..o..o..o..o..o.
2	Planning	==2..o.=o..o..o..o..o.=o.
3	Contracting	==o==3..................
4	Backfile Conversion	==o======4
5	Installation	==5.........=o....=o
6	Document Management and Processing	==o==o==6......==o
7	Search Systems	==o==o==o==7...==o
8	Publication Systems	==o==o==8.....o
9	User Support and Training	==o==o...==o==o==o
10	Remote Access Systems	==o==o==o..9
11	Management and Financial Systems	..o..o..o..o..o==o==o..o

Legend:

```
=== : task
... : task suspended
=o= : milestone
=3= : major milestone
```

Table 7.2.

7.2.1. Organizational Activities

The project starts with the organization of the project management team in year one. Other organizational activities include the development of system operational practices and procedures (year two), and the recruitment and training of systems operating staff (year two). Recruiting, staffing and training of essential operations personnel should commence no later than two months prior to installation of the computer equipment.

By the end of year three, staffing of the project management organization will be completed (Milestone 1).

7.2.2. Planning

The present preparatory planning activities will be followed by the development of a detailed implementation plan (year one: Milestone 2) which includes:

- detailed design requirements,
- the definition of the quality assurance program including the specification of all pilot projects to be initiated,
- detailed security plan,
- detailed facilities plan.

This detailed plan will be subject to yearly review.

As much of the detailed design and prototyping of the software as possible should be done using microcomputer equipment and computer aided software engineering tools. This approach will reduce the dependancy of systems development schedules on the installation schedule for the main computers. In fact, prototype databases if developed prior to installation of the mainframe(s) can be used with prototyped applications software for testing of performance and load capacity as part of mainframe acceptance testing activities.

Further planning activities include the review of the results of the backfile conversion pilot projects (year one); the development of an operations integration plan (year three); and, once information becomes available, the evaluation of the usage of the patent system by external users (year eight).

The development of the operations plan and training programs precedes the full operational deployment of the applications processing and search systems to take place in the following year. The decision that will require critical evaluation commencing from this point onwards will be that dealing with the question of archival of converted paper documents.

The project is completed with a final project review.

7.2.3. Contracting Activities

These activities include the development and issue of RFP-s, the evaluation of proposals, and the awarding and supervision of contracts. RFP-s for the backfile conversion pilot project, facilities, hardware, and software development will be prepared and issued in year one. Contracts for the backfile conversion pilot will be awarded in the first year. Responds to the other RFP-s will be evaluated and major contracts will be issued in year two (Milestone 3).

7.2.4. Backfile Conversion

Backfile conversion activities include the image scanning of in-force Canadian patents and US title pages, the conversion of text portions, and eventually the creation of integrated text and image document files.

Backfile conversion activities are preceeded by pilot projects. Upon evaluation of these pilots, backfile conversion begins in year one. Based on presently available information, this task will have a duration of approximately two and a half years. Completion is expected in year three (Milestone 4).

7.2.5. Installation

The acquisition of major hardware and operating system software components should take place in the second year. The contract for the space conditioning of the facility should be awarded and the contract completed at least one month prior to the installation of hardware. The hardware installation will be followed by a hardware and systems software acceptance test (Milestone 5). Facilities for providing access for external users will be installed in year six, when the software for external user access becomes ready for field trial. Test workstations for external users participating in the field trial will also be installed. Finally, these communications facilities will be extended to provide access for the general user community in the final year of the project.

7.2.6. Document Management and Processing Systems

The development of these software systems begins with the development of the application processing software and the construction of a testbed database (year two). The index database will be constructed, and the application processing software will undergo unit and integration tests. The application processing software will be tested on the functionally integrated text and image database.

In the third year, these tests will be completed and evaluated.

The application processing system will be deployed on the production database of separate and functionally integrated text and image holdings in year four (Milestone 6).

These software will become operational on the fully integrated databases in year seven.

7.2.7. Search Systems

Following the construction of the index database in year two, index and classification search software will undergo unit and integration tests. Full-text search software will be tested on the index database and separate text data holdings.

In year three, the full-text search software will be extended to operate on the functionally integrated text and image database. These tests will then be evaluated.

Search software on separate integrated text and image databases will be deployed in the fourth year. At the same time these software will undergo tests on the testbed database of functionally integrated text and image data holdings. Search systems on the production database of functionally integrated text and image holdings will be deployed in year five (Milestone 7).

Deployment of these search software on the fully integrated text and image database will be completed in year seven.

7.2.8. Publication Systems

The development of publication/dissemination systems will begin in year three. Following test and evaluation on the testbed database in year four, these systems will be deployed and become available for internal users in the fifth year (Milestone 8). Following the development of remote access facilities, these systems will become accessible for external users.

7.2.9. User Support and Training

The development and implementation of user support and training programmes for the application processing systems will happen in year three. In year four, development and implementation of such programmes will follow for dissemination systems.

The development of training programmes for external users will commence in year six. These programmes will first be delivered to those users who participate in the technical field trial of remote access systems, in year seven.

Training programmes will be developed for the general user community in year seven. These programmes will be implemented in the final year of the project.

7.2.10. Remote Access Systems

Year five represents the year in which all systems are deployed in the operatational environment. While internal users become more accustomed and experienced in the use of the automated system, development proceeds to establishing the external user interfaces for access to the system by the public. These interfaces are of both technical and operational nature, including such tasks as the installation of wide area network access facilities, or the development of procedures for trouble-shooting customer calls.

External user interfaces to the automated system will be defined and the technical field trial will be planned in year five. In year six, development proceeds to support requirements for external users. The major focus here is in the configuration of a front-end processor to act as an access control facility between the external users and the main systems. This system must address the broad security requirements of ensuring the confidentiality of non-disclosed applications, and the integrity and availability of the patent data for search. Development also of other support systems (e.g., user accounting) also proceeds, in parallel with the planning of a controlled technical field trial. This trial should be followed later by a full market field trial.

These systems development activities, providing access to the functionally integrated text and image database, will be completed in the sixth year. Interfaces to the full-text search systems will be tested.

These systems will be extended to provide access to the fully integrated text and image databases, and technical field trial will be performed in year seven. Upon evaluation of these trials, software to provide access for general users will be deployed.

In year eight, communications facilities for general users will be marketed and promoted (Milestone 9).

7.2.11. Management and Financial Systems

Project management will be supported by the employment of state-of-the-art information engineering systems during the entire project life.

User accounting and management information systems will be developed in year six. Unit and integration testing of these system components will commence in the seventh year. User accounting systems will undergo further testing with the participation of field trial users.

Management information and user accounting systems will finally be deployed in year seven.

7.3. Organization and Staffing Plan

This section addresses the requirements for an organization and staffing plan that must be defined as a prerequisite to implementation. The organization of the project management team to the level of the Project Director and his current core group of project management staff is assumed. In particular, it is also assumed that the various advisory and steering committees have been established and that the organizational relationships of the core group to these committees is in place and that the mission of this group is clear: to manage the implementation of the automation project.

Prior to the start of the implementation phase, it will be essential for the ASB to define the additional organizational and human resources required to initiate the implementation phase, to acquire these key resources and to provide the necessary facilities for them.

7.3.1. Project Management Functions

The following functions have been included as part of the over-all project management function: project planning and administration, contract administration, facilities and equipment planning and provisioning and user support services.

7.3.1.1. Planning and Administration Functions

This group would be responsible for the overall planning and administration of the automation project. It would define and maintain the project plan and schedule of major development

activities, establish and implement the project standards and methods to be used and, in particular, the definition of the quality assurance program and the development of the security program. It would manage the recruiting and staffing requirements that will be necessary throughout the life of the project. It would support the conversion to the automated system by directing and administering the necessary human resource training programs that will be required by CPO staff. Finally, the planning and administration function of the project management group would include the planning and maintenance of the project budget and other finance and accounting functions directly related to project resource allocations and expenditures.

7.3.1.2. Contract Administration

Resources will be required within the project management group to ensure the accountability of contracted resources. This function would, therefore, cover all the activities inherent in managing contract operations and the quality control function.

7.3.1.3. Facilities and Equipment Planning and Provisioning

The project management group would be responsible for the definition of all facilities and equipment requirements and ensure their acquisition commensurate with project schedules. Initially, the core staff would plan and provide for the facilities required by the additional project personnel required to initiate the implementation phase. These facilities include furniture, floor space, and project support resources, such as clerical support and project management tools.

7.3.1.4. User Support Services

A group such as this could interface between the various CPO operational groups and the software development group. As such, its role would be to develop detailed definitions of the user requirements for software development, on the one hand, and to support the user in ensuring that the end product developed meets those stated requirements.

User Support Services would also be responsible for co-ordinating the planning, design and delivery of CPO training programs to support the conversion to the the automated system.

7.3.2. Development and Systems Support Functions

In addition to the essential project management functions identified resources will be required for systems development, systems management and computer operations. Computer operations would include the data entry operation although additional resources would be required for conversion of the backfile.

7.3.2.1. Systems Development

This group would be responsible for designing and developing the applications software systems according to the user requirements. Given the size of the project, this group should be equipped with computer aided software engineering tools for prototyping and code construction. This group would also be responsible for the development and maintenance of systems documentation.

7.3.2.2. Systems Operations and Systems Management

These groups would be responsible for the day to day operations involved in running the main computer and data conversion systems. The systems management function would be responsible for allocation and maintenance of systems resources. This encompasses such tasks as database maintenance, systems software maintenance, management of user acces control facilities and communications.

In order to develop preliminary cost estimates for project staff resource requirements, an organizational structure was defined around these generic functions. The number of positions in the structure was based on assumptions concerning the relative skills sets required to fully support the function. For on-going activities and critical project management functions, it was assumed that the positions would be staffed with government employees. Activities such as software development, were assumed to be contracted out. The schedule of major activities was then reviewed to determine the man-months and person-years of support required in each year of the project. These results are summarized in Appendix C, and the more detailed implementation plan and cost estimates, also found in that appendix, were based on the definition of these generic functions.

7.4. Security Plan

This section identifies, reviews and summarizes the administrative, technical and procedural practices that must be formally defined by the project management staff and implemented as an aspect of working methods in the automated system environment to comply with Federal Government policy on security.

7.4.1. Familiarization

As a first step towards the implementation of a security plan, it will be necessary for project staff to become completely familiar with current policy on security within the context of an automated environment, and the prevailing practices and guidelines for the implementation of that policy. Present security policy would be defined in Treasury Board Circular 1986-26, dated June 18, 1986, entitled "Government Security Policy". The general practices for implementing that policy are defined in the Government of Canada Information Technology publication: "Electronic Data

Processing Security Standards and Practices for Departments and Agencies of the Government of Canada", GES-14, dated July 1987 and subsequent revisions.

7.4.2. Definition of Security

In this context, security addresses the need to protect the operational systems, data and facilities against accidental or deliberate attempts which would compromise the confidentiality, integrity and availability of information processed and stored. Within the CPO, there are some documents that are protected under the Official Secrets Act. Such documents should not be included in the automated system. Other documents (non-disclosed application files) are commercial confidential. Confidential documents must be handled in such a way that access is available only to those who have a "recognized need to know". The vast majority of patent data is, however, in the public domain by definition and purpose of the patent system. Given that, on the one hand, the information content of a patent is the definition and description of the intellectual property of some individual or organization, while on the other, it is intended for access by the public, it is of paramount importance that the integrity of the information be absolutely protected while the availability of such information be universally guaranteed. We might say that the CPO patent search data base will of necessity be a "read-only" system.

7.4.3. Criteria for Establishing Secure Systems

7.4.3.1. Access Privileges and Processing Capabilities

It is necessary to define criteria in the interests of developing a secure system. User groups and the data they have access to must be defined in order to establish a matrix of access privileges and capabilities to be granted to users.

The commercial confidential portions of the database (application files prior to disclosure) must be separated from the publicly accessible databases by physical or rigid operational methods. The potential loss caused by premature exposure of the contents of a patent application prior to disclosure would be unacceptable to the inventor and the CPO.

The definition of access privileges and capabilities must be established for all user groups with respect to hardware, systems software, applications software, database management systems and data. In particular, the highest level of security clearance must be reserved for systems management and systems operations personnel and others with immediate access to the internals of the facility.

7.4.3.2. Classified Information

It has been adopted as a design criterion that patents protected under the Official Secrets Act will not be held in the automated system. They will continue to be processed manually.

7.4.3.3. Detailed Security Requirements

Definition of detailed security requirements in terms of mandatory practices and guidelines covering

a) organizational and administrative security
b) personnel security
c) physical and environmental security
d) systems security
e) hardware security
f) software security
g) operations security
h) communications security
i) database security

will be an essential activity during implementation.

a) Organizational and Administrative Security

In the execution of this function, departmental security personnel, responsible to the Deputy Minister, will be required to ensure the development, implementation, co-ordination and audit of security policies, standards and procedures. These responsibilities include all areas of personnel security, physical and environmental security, systems security, hardware security, software security, operations security, and communications security.

b) Personnel Security

Security clearance levels for both public and private sector personnel should be obtained for access to buildings, offices, sensitive documentation, system usage, and system information. All personnel should be given an orientation on all aspects of security, including the filing of pertinent security forms, such as the Oath of Allegiance, and the Oath of Office and Secrecy.

c) Physical and Environmental Security

For the CPO automation project, all information and facilities must be protected from natural disasters, industrial hazards, electromagnetic interference and access to facilities by unauthorized personnel. This security encompasses any central processing facilities that might be installed, but also extends to libraries, back-up storage and archives, report

distribution, off-site terminal or nodal distribution points and all information disposal sites.

d) Systems Security

This requirement addresses the need to protect against unauthorized access to hardware and software systems and data. Security practices in this area will be required and will address such issues as the need for terminal identifiers, terminal connection authentication, privilege denial capabilities, audit trails and transaction logs, and multi-level data/software privileges for each user group.

e) Hardware Security

All hardware and respective communication capabilities should be monitored and regulated with supporting schematics and descriptions of all components within the configuration. Where necessary, terminals should have restricted access and all terminals should be identifiable when logged on. This would allow for identification and authentication capabilities for systems and operational processing. Whenever hardware maintenance is required, monitoring of the procedure by qualified personnel should be done to inhibit any unauthorized access to data, application processing software, systems software or hardware facilities.

f) Software Security

Software, in this context, includes applications and systems software, procedural command files, utilities and database management systems. All software in both development and operational areas of development environments must be qualified with certain security attributes to prevent unauthorized access, usage, modification, deletion or removal as well as the unauthorized installation of software. This includes the requirement for quality assurance procedures for analysis, design and acquisition, development and implementation of software for new applications or the changing of existing applications.

g) Operations Security

Centralized EDP operations should be restricted to operations personnel. Similiarly, distributed or microcomputer based processing should also be restricted, although in this case, the operator and the user are often identical. The restrictions might include assigned physical access to computer equipment, access to documentation, physical system locks and keys, and various levels of password protection. Operations personnel will, of consequence, be responsible for the control of input, output, software back-up and

recovery, software updates, data transfers, and the external distribution of software, data, and information media.

h) Communications Security

All communications lines and connections should be charted to identify type and location of the lines, terminators and the expansion/recovery capabilities. External lines should terrminate on a front-end processor to control access privileges, especially where the access originates on a remote dial-up service. This practice is of special importance when public domain and commercial confidential databases are not physically separated. All the procedural aspects of communications lines should be structured, catalogued and traceable by logs as far as operations, problem analysis, and maintenance are concerned. The communication capabilities should be tested periodically to prevent unauthorized access.

Presently two fields can be identified where traffic of confidential data on telecommunications links may occur. These are a.) information exchange with other jurisdictions, and b.) electronic submission of applications or applications correspondence.

i) Database Security

Apart from preventing unauthorized access to data holdings, the integrity of the CPO files must also be ensured. This is accomplished through the following measures.

- Establishment and maintenance of a final archive of documents
- Audit trail of all changes to document files
- Off-site backup of the entire data holdings.

8. BENEFITS ASSESSMENT AND RECOMMENDATIONS

8.1. General Benefits of Automation

A successful automation of the Canadian Patent Office would provide benefits to the Federal Government, the department, the Patent Office (CPO), the CPO's clients and the nation.

Some of the major benefits produce a multiplicity of lesser benefits which may, nonetheless, be of significant financial import. This is particularly true of the benefits accrued from essential improvements in document control and workflow management.

In the following paragraphs, major benefits are discussed in general terms. For those benefits which are found to be quantifiable, dollar estimates are given. Intangible benefits are described in sufficient detail so as to render an appreciation of their intrinsic value.

It will be understood that an incremental approach for implementation is introduced in Chapter 6 and that each of the implementation levels, by varying the pace, depth or comprehensiveness of the project, varies the benefits also. These pages discuss the benefits flowing from the preferred implementation. Options four to two vary the benefits mainly in the area of quality - option one, the base design, offers benefits on a considerably lower scale.

This chapter presents an abstract of results of research into benefits of patent automation which is contained in working papers. Broader research into aggregate economic benefits of patent automation is underway; those results are not addressed here. Each subject of benefits is briefly stated, followed by an explanation and evidence.

8.2. Summary of Tangible Benefits

Savings and Cost Avoidance at the CPO

- Revenue increase or avoidance of loss of future revenue (not estimated)

- Cost Avoidance

    - Planned Essential Improvements $10.9 M
    - Increased Workload $4.0 M
    - Bill C-22 Economies $4.4 M

    $19.3 M

- Cost Reduction

    - Space: Removal of Archives (not estimated)
    - Staff: Savings of support staff and equipment $5.0 M

    $5.0 M

CPO Savings Total	$24.3 M
Social Costs Avoided	
• Reduction in wasted R&D	$200.0 M
• Reduction of patent litigation costs	$65.0 M
Total Social Cost Avoidance	$265.0 M
Total Benefits during Project Period	$290 M
External Economic Benefits (over 5 years after completion)	$163 M

8.3. Fundamental Benefits

The benefits to be derived from automation of the CPO can be considered under eleven headings.

8.3.1. Strategic Maneuverability

The ability to secure strategic maneuverability at a time of unprecedented technological change and of massive change in the patents system on the international scene is a vital benefit for the CPO.

After completion of their automation projects, the trilateral countries will have about ninety percent of the free world's holdings of patents and a very high share (more than 80%) of all technical information in machine-readable form. Members of the trilateral agreement will have significant advantages in patent processing, and new emerging benefits in strategic search for technology, leading to industrial policy. The benefits of access to machine-readable technological information are available only to countries which invest in compatible systems, and which convert their own data.

8.3.2. Removal of Uncertainty

The department, the CPO, management and (particularly) staff benefits from the removal of a climate of uncertainty that presently prevails. The Canadian Patent Office is currently at the lower limit in terms of the size of its examiner corps that is considered necessary for continued viability (WIPO). The staff is faced with rising workload, tools of declining value such as the Canadian classification system, and knowledge of the certain superiority of patent decisions in other countries in the near future. On the other hand, the CPO has rudimentary EDP systems,

which eases the problem of choice of solutions, and the Canadian solution will benefit from new standards and technology. Therefore, the certainty of progress will be essential in maintaining the quality of patent decisions during the years of system development, by nurturing the requisite sacrifices from staff.

8.3.3. Internal Standardization - External Harmonization

Internal effectiveness will be improved by processing more documents in standardized form more quickly. Internal efficiency will grow as paper functions are reduced gradually and a well-planned system replaces present traditional procedures and methods, which were developed over a century. Such systems of mixed methods and processes are no longer adequate to support essential services to the public without significant increase in staff resources given the rate of flow of incoming patent applications and the size of the search folio that has accumulated.

Internationally, electronic data exchange will occur on a large scale, with standards expected in the early 1990s. In particular, access to the data holdings of the members of the trilateral agreement is essential to the long term objectives of the CPO, and may be available ONLY in electronic form in the future. Canada could lose a large part of its patent business if it has no compatible host facility recognized by others.

8.3.4. Improved Quality and Efficiency

Patent examiners have scarce and valuable skills. All patent automation plans and tests performed elsewhere demonstrate that automation leads to access to more documents which are more effectively selected, more rapidly presented, and more efficiently scanned. Index-based or contextual search helps the examiner to narrow down prior art faster, and discover prior art previously overlooked. In-non-automated systems the quality assurance of the search function is sporadic. With automation, quality assurance of the search function can be made concurrent. While measuring increments in quality is not easy, it is clear that a significant improvement of the quality of search is a key benefit of automation. Moreover, with automation, the quality attained is more easily verifiable.

Effectiveness of the examiner role is closely linked to economy of effort. In this area, the numbers are unequivocal. The current delay in application processing of 34 months, increasing annually, cannot be arrested or reversed without automation, except by the heavy (proportional) reinforcement of the corps of examiners. Based on a growth rate of applications of three percent per year, about twenty additional examiners will be needed by 1996, if the technical tools of search and examination are not improved. Beyond that, eleven professional person years are needed to reduce the period pending to just under two years. Support staff for these two categories of professional staff are 16 person years. The cumulative cost of this additional requirement in 1996 will be $4.4 million.

In terms of net savings after the automated solution is installed and stabilized, the working papers show potential economies in terms of person years late in the project period. These are hard to

quantify and depend on approval by management of a human resource retraining plan. Savings at the end of the project period are in the order of 20 person years, many of which can be utilized in automation functions. Finally, there are minor savings due to cost avoidance in replacement of currently installed equipment.

8.3.5. Broader Dissemination Capabilities

The benefits of better dissemination of patent and related technological data to industry are widely documented (The Bottom Line, Ch. 5 and references therein). The Market Demand Assessment showed that between one and three million 'uses of patent information' per year can be expected. It is known that the main beneficiaries will be small and medium firms which have a high rate of job creation and are the main victims of wasted efforts in research and development. Quantitative estimates of benefits through social cost avoidance are harder to find. The German ministry of Science and Technology (source: GPO, 1987) has estimated that up to one third of the R&D expenditure in the Federal Republic has been wasted, mainly due to re-invention and parallel invention. The rate of waste of Canada's R&D cost (currently $7.1 billion per year) has not been estimated, nor do sources from other countries or in Canada tell us which share of the wasted R&D could in fact be saved by a better system of patent and technology dissemination.

A conservative estimate is provided as follows. Assuming that wasted R&D occurs in Canada at the same rate as in Germany, and that only one percent of the R&D cost in Canada would be avoided through a dissemination program based on specific aspects of the proposed automated system, then the annual savings will be in the order of $70 million. Assuming that dissemination benefits are available in year five of the project period (cf. plan), this produces an estimate of external benefits in the order of $200 million by 1996.

Another area of external benefits is in costs of patent litigation avoided by better patent and technology information. The methodology of this argument was developed by the USPTO (Master Plan, April, 1987). A good estimate of the total cost of related litigation in Canada has been $96 million per year (Bolling, January, 1986, CPO internal research). Based on the U.S. debate on this subject, one may make a conservative assumption to the effect that only ten percent of the social cost of litigation can be avoided by improved quality of patents, which reside in a database of improved quality. Applying the rate of conversion of Canadian patent files to automated form proposed here, one derives an estimated cumulative benefit in 1996 of $65 million.

The benefits summarized here result in estimated continuing savings of social costs of $80 million per year. This figure could be doubled without excessive assumptions.

Direct external economic benefits of patent automation have been estimated in the Economic Impact Analysis.

8.3.6. Benefits to Patent Professionals

Benefits to patent agents are seen mainly in the fact that their concerns about loss of patent business to other jurisdictions will be addressed with a Canadian automated system. The proposed system will act as an agency of redistribution of other country data, and give access to international technology data banks. The estimated high incidence of expected use of the Canadian system (1 to 3 million uses of patent-related data per year) is evidence of anticipated benefit. This group of benefits has not been quantified in terms of dollars as no pricing regime has been established for services of an automated CPO.

8.3.7. Secure Backup

The benefits of having a greater level of security for all aspects of Canadian patent holdings are tangible, but not readily quantifiable. Benefits consist of reduction of access of the public to stacks, greater protection of paper documents through reduced handling and eventual removal to off-site storage, and protection of the information as such as a national asset in a secure location on media that are reproducible and have long life. Finally, storage in electronic form will avoid the cost of restoring microfilm copies of older documents (required beyond ten years).

8.3.8. Space Savings

The ability to eliminate the paper file holdings as conversion to other media progresses is not easily estimated, as the incidence of use of collateral documents is not known. Habits of work established by long tradition are probably as important as are actual needs to have paper files close at hand. Current costs of the floor space for on-site paper file holdings are $800,000 per year. The rate of removal is hard to estimate, and should be left to future updates of the Canadian automation plan. However, it is clear that without automation, additional space will be required to store documents in paper form or microfilm.

8.3.9. Bill C-22

The additional duties of the CPO in implementation of Bill C-22 include several functions, such as fee processing and 18 months disclosure, which will be performed more efficiently if they are combined with the automation solution. The proposed system contains related modules. It is not clear at this point whether there are net savings of the costs of implementation of Bill C-22 that can be attributed to automation.

8.3.10. Avoidance of Costs of Essential Improvements

The LRIP identified a program of planned essential improvements with the following components and costs. Updating the Canadian Patent Classification System was estimated by a CPO task force to cost in the order of $1.6 million. Other essential improvements are the Search

File Integrity Program ($2.5 million), a Patent Publication System for a revised Canadian Patent Office Record ($4.6 million), provisions for receipt of foreign patents and applications in electronic form ($1 million), and the provision of a case management and a fee collection system for application under the Patent Cooperation Treaty (PCT) ($0.9 million). The total cost of essential improvements avoidable through automation is $10.9 million.

8.3.11. Derived Benefits

Background papers for the present study have mentioned other, indirect benefits which are not listed here. These benefits are either minor in amount or not clearly identified at this time. They will be identified in updates of the automation plan. Examples are more rapid order fulfillment for orders of patent documents from the public, and better statistical data on the origin and ownership of Canadian patents.

8.4. Conclusions and Recommendations

This chapter has shown that automation of the Canadian Patent Office would result in significant benefits for the Department, for the professional and business communities, and for Canadian society at large.

Automation at the Canadian Patent Office is technically and financially feasible. Foreign offices are planning or implementing similar automation projects. Developments in technology provide the necessary means for constructing a state-of-the-art automated patent system. Financially, technology of reasonable cost is available and further moderate cost reductions in important fields of technology can be anticipated. There are no significant qualitative advances in technology expected which could change the benefit-cost ratio in a major way.

Although the project is not entirely without risk factors, its success is ensured by the current existence of necessary technologies, the development of international standards for electronic document management, the development of patent-related standards, and the experience which can be gained from foreign offices undergoing automation.

Failure to automate, or a major delay would be very costly to Canada. The Patent Office would not be able to face the ever increasing workload, it would not be a partner to automated foreign offices, and it would not fulfill its duties in providing technological information to society. The demand of the patent community for related services would migrate to other jurisdictions, thus rendering a future attempt to automate less effective. A delay would reduce the degree of co-operation from other patent offices.

These considerations show that an attempt to automate the Patent Office is not premature. Automation must begin now.

Recommendations

- The CPO should enter into negotiations with other patent offices over data exchange and standards.

- The CPO should begin contributing to the international standards effort for text and image storage and data exchange.

- The CPO should undertake to standardize the format of submission of applications in concert with the Canadian patent community, and other offices, notably the USPTO.

- The CPO should disseminate its plans to the patent community, and begin marketing its future automated system so as to increase its acceptance in the market and avoid the loss of future demand.

- Consumer and Corporate Affairs and the Government should enable the CPO to embark promptly on the automation project overall, and develop the preferred solution aggressively, with continuous, firm control of ASB over the development process.

APPENDICES

APPENDIX A: VOLUMETRICS AND THROUGHPUT CALCULATIONS

A.1. Tables and Assumptions

TABLE I: PAGE VOLUMETRICS OF CANADIAN PATENTS

Assumption about compression: empty spaces on pages occupy virtually no storage. Area covered by text is not compressible.

A. Text Pages (abstract, disclosure, claims, correspondence)

Page Size:	8.5*14 inches
Total Area of a Page:	119 sq. inches
Text Volume:	1800 chars/page
Character Width/Heigth:	0.1 inches
Area Covered by Text:	18 sq. inches
Compression Ratio:	1:6.61
With 10% Overhead:	1:6

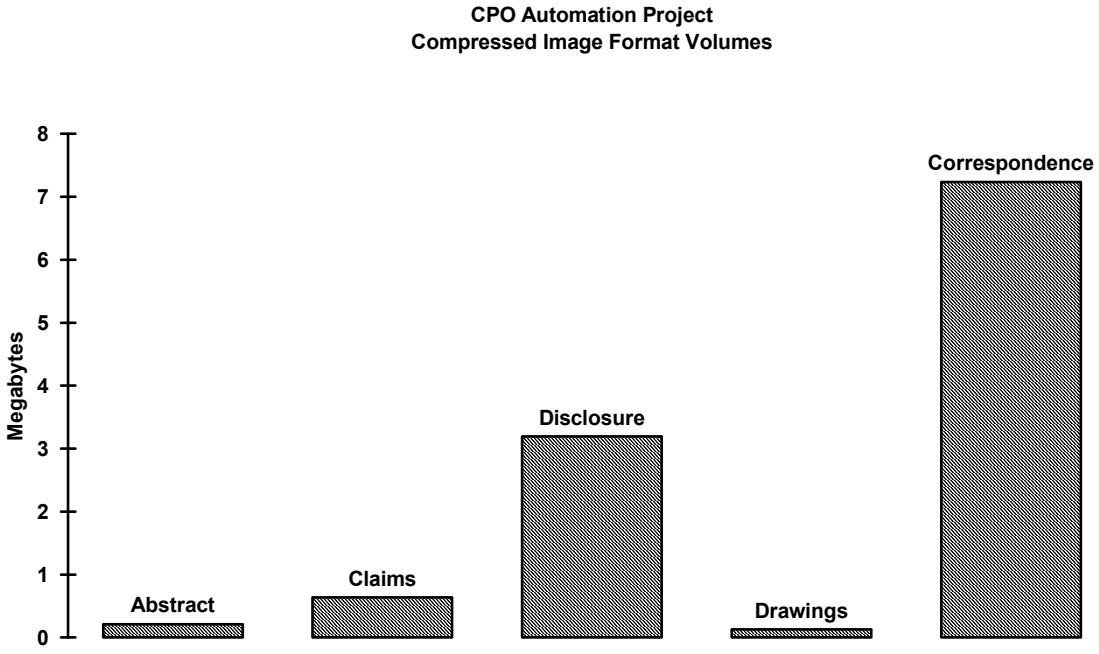
B: Front Cover Page

Page Size:	8.5*14 inches
Text Volume:	274 chars/page
Area Covered by Text:	8.5 sq. inches

Compression Ratio:1:14

C: Drawing Pages

Page Size:	8.5*14 inches
Area Covered by Drawing:	11.9 sq. inches
Compression for Drawing:	1:2
Compression Ratio:	1:20


CPO Automation Project
Compressed Image Format Volumes
Megabytes
0
1
2
3
4
5
6
7
8
Abstract
Claims
Disclosure
Drawings
Correspondence

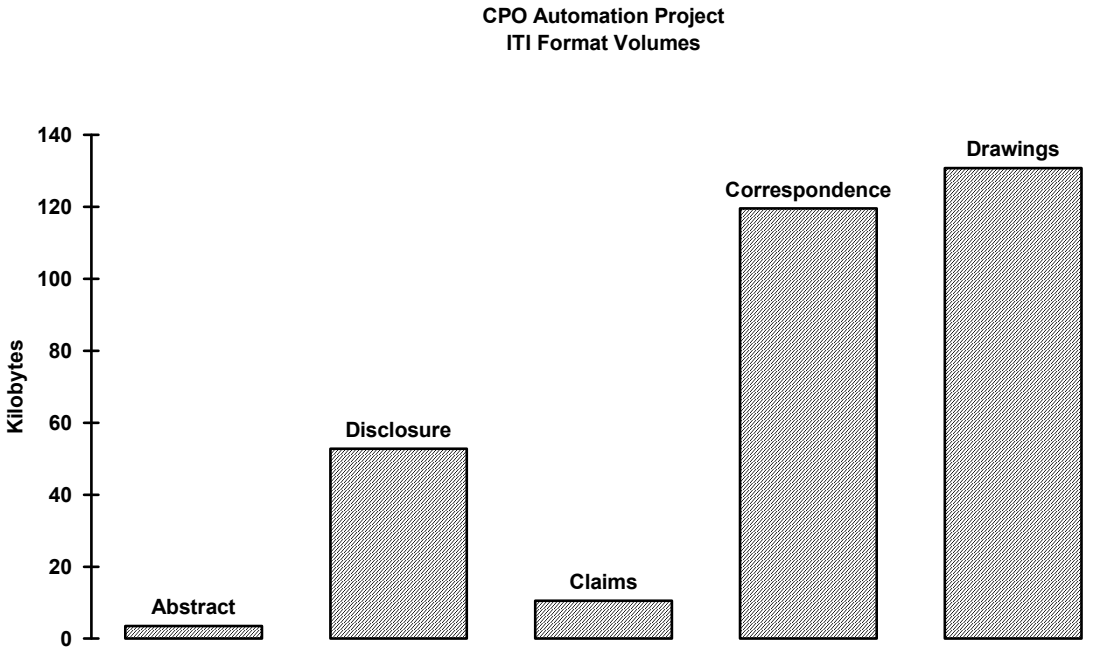

CPO Automation Project
ITI Format Volumes
Kilobytes
0
20
40
60
80
100
120
140
Abstract
Disclosure
Claims
Correspondence
Drawings

TABLE II: DOCUMENT VOLUMETRICS

A. Patent Data Organization:

No	Name	Type	Page	Char	Compr.	Area	KBytes @300dpi	ITI
a.	Patent proper							
1	Front Cover	T	1	274	1:14	8.50	93.38	0.54
2	Abstract	T	1	1800	1:6	19.83	217.90	3.52
3	Disclosure	T	15	1800	1:6	19.83	217.90	3.52
4	Claims	T	3	1800	1:6	19.83	217.90	3.52
5	Drawings	T&I	2	0	1:20	5.95	65.37	65.37
	Equivalent non-compressible area total (22 pages):					397.23	4364.14	198.07
b.	Correspondence							
6	Correspondence	T	34		1:6	19.83	217.90	3.52
7	Corresp. reduced	T	10		1:6	19.83	217.90	3.52
	Correspondence Total					674.23	7408.45	119.53
	Reduced corresp. total:					198.33	2178.96	35.16
	Full document total:					1071.56	11772.59	317.60
	Reduced document total:					595.57	6543.10	233.23

Notes:

- Black&white images (no half-tones) are assumed.
- The 'Integrated text and image' database calculations are made on the following assumptions:
  - 100% overhead compared to a text database on the text portions of a document (format information, occasional images - such as a mathematical formula -, occasional special symbols)
  - Drawings occupy the same space as in an image database
  - Only 300dpi resolution is calculated.

B. Number of Patents:

	Present	Future (10 years)
CANADA pending	90,000	70,000
CANADA active -1978	200,000	0
CANADA in force 1978-	200,000	400,000
CANADA archived	830,000	1,030,000
US.	4,500,000	N/A
Other	20,000,000	N/A

C. Storage requirements (Text&Image @300dpi):

a. CANADA pending (for maximum 100,000 applications)

Document portion	Database type	CAPACITY IN GIGABYTES			
	Text only	Image @200	Image @300	Image @400	Text&Image
Front cover	0.025	3.958	8.905	15.832	0.051
Abstract	0.168	9.236	20.781	36.942	0.335
Disclosure	2.515	138.534	311.708	554.137	5.029
Claims	0.503	27.707	62.342	110.827	1.006
Drawings	N/A	5.541	12.468	22.165	12.468
Corresp. red.	1.676	92.356	207.806	369.425	3.353
Correspondence	5.700	314.011	706.539	1256.044	11.399

b. CANADA in-force -1978 (200,000 at present, will be reduced to 0)

Document portion	Database type	CAPACITY IN GIGABYTES			
	Text only	Image @200	Image @300	Image @400	Text&Image
Front cover	0.051	7.916	17.812	31.665	0.102
Abstract	0.335	18.471	41.560	73.885	0.671
Disclosure	5.029	277.068	623.404	1108.274	10.058
Claims	1.006	55.414	124.681	221.655	2.012
Drawings	N/A	11.083	24.936	44.331	24.936
Corresp. red.	3.353	184.712	415.603	738.849	6.706
Correspondence	11.399	628.022	1413.049	2512.087	22.799

c. CANADA in-force 1978- (200,000 at present, will grow to 400,000)

Document portion	Database type	CAPACITY IN GIGABYTES			
	Text only	Image @200	Image @300	Image @400	Text&Image
Front cover	0.102	15.832	35.623	63.330	0.204
Abstract	0.671	36.942	83.121	147.770	1.341
Disclosure	10.058	554.137	1246.808	2216.548	20.117
Claims	2.012	110.827	249.362	443.310	4.023
Drawings	N/A	22.165	49.872	88.662	49.872
Corresp. red.	6.706	369.425	831.205	1477.698	13.411
Correspondence	22.799	1256.044	2826.098	5024.175	45.598

d. CANADA archive (830,000 at present, calculations for 1,200,000)

Document portion	Database type	CAPACITY IN GIGABYTES			
	Text only	Image @200	Image @300	Image @400	Text&Image
Front cover	0.306	47.497	106.869	189.990	0.612
Abstract	2.012	110.827	249.362	443.310	4.023
Disclosure	30.175	1662.411	3740.424	6649.643	60.350
Claims	6.035	332.482	748.085	1329.929	12.070
Drawings	N/A	66.496	149.617	265.986	149.617
Corresp. red.	20.117	1108.274	2493.616	4433.095	40.233
Correspondence	68.396	3768.131	8478.295	15072.523	136.793

e. US (4,500,000 documents)

Document portion	Database type	CAPACITY IN GIGABYTES	
	Text only	Image @300	Text&Image
Front cover	15.087	400.760	140.384
Abstract	N/A	N/A	N/A
Disclosure	135.787	3819.005	271.574
Claims	N/A	N/A	N/A
Drawings	N/A	661.254	661.254
Corresp. red.	N/A	N/A	N/A
Correspondence	N/A	N/A	N/A

Notes:

- only 300dpi is used, being the standard in US
- US format assumed
- Front cover contains abstracts
- Disclosure includes claims
- No data available on correspondence

f. Foreign (20,000,000 documents)

Document portion	Database type	CAPACITY IN GIGABYTES			
	Text only	Image @200	Image @300	Image @400	Text&Image
Front cover	5.104	791.624	1781.154	3166.497	10.207
Abstract	33.527	1847.123	4156.027	7388.492	67.055
Disclosure	502.914	27706.847	62340.405	110827.386	1105.828
Claims	100.583	5541.369	12468.081	22165.475	201.165
Drawings	N/A	1108.274	2493.616	4433.095	2493.616
Corresp. red.	335.276	18471.231	41560.270	73884.916	670.522
Correspondence	1139.939	62802.186	141304.918	251208.736	2279.878

Note: Canadian format assumed.

TABLE III: COMPARISON TO US DATA

The USPTO experience shows that a US patent, when stored as image, occupies appr. 1 Mbyte @ 300 dpi. The calculations above show, that a Canadian Patent requires appr. 4.5 Mbytes of storage capacity. To test the theory these calculations were based on, here is a comparison.

US Patent:

Page Size:	8.5*11 inches
Front Page Text:	0.5 pages
Front Page Drawing:	0.5 pages
Text Volume:	7200 characters/page
Character width/height:	0.05 inch
Compression for Drawings:	1:20
Compression for Text:	1:5.2
Front Page:	1 page
Text:	4.5 pages
Drawings:	3 pages

Based on this information, here is a comparison table:

Megabytes	CANADA	US
Front Page & Abstract:	0.30	0.12
Disclosure & Claims:	3.83	0.87
Drawings:	0.13	0.15
Total:	4.26	1.14

The higher storage capacity requirement of Canadian Patents is due to larger characters - more graphic information is stored about character images at the same resolution.

TABLE IV: INTERNAL WORKFLOW ESTIMATIONS

A: Correspondence Input/Output Estimations:

The following table describes the operations affecting correspondence, yearly amount, and number of pages likely to be invloved.

Application Processing Functions	Transactions per year	Pages Inputted per Transaction		Pages Outputted per Transaction	
		Each	Annually	Each	Annually
RECEIVING (30+3 persons)					
1 receipt	35,000	1	35,000	0	0
2 ack letter/certificate	35,000	0	0	2	70,000
3 classification	35,000	0	0	0	0
4 registration request	49,000	1	49,000	0	0
5 registration letter	49,000	0	0	1	49,000
6 formality letter	5,250	0	0	1	5,250
7 formality response	5,250	1	5,250	0	0
RECEIVING Total:	213,500		89,250		124,250
18 MONTH PUBLICATION (10+1 persons)					
8 18 month publ.	35,000	0	0	10	350,000
EXAMINATION AREA (30+3 persons)					
9 receipt/classify	35,000	1	35,000	1	35,000
10 allowed search*	35,000	0	0	0	0
11 examination request	24,500	2	49,000	0	0
12 R40 letter	8,750	0	0	2	17,500
13 R40 response	8,750	3	26,250	0	0
14 order art	24,500	0	0	1	24,500
15 examine*	24,500	0	0	0	0
16 copending search*	24,500	0	0	0	0
17 patent search*	24,500	0	0	0	0
18 lapsed search*	24,500	0	0	0	0
19 foreign search*	3,500	0	0	0	0
20 exam report	17,500	0	0	2	35,000
21 exam report resp	17,500	5	87,500	0	0
22 copending search*	24,500	0	0	1	24,500
23 allowance/classify*	24,500	0	0	1	24,500
24 abandonment	10,500	0	0	1	10,500
EXAMINATION area Total	332,500		197,750		196,000
ISSUE (25+2 persons)					
25 allowance response	24,500	2	49,000	0	0
26 grant/publication	24,500	0	0	10	245,000
27 renewal submit	105,000	1	105,000	0	0
28 renewal ack	105,000	0	0	1	105,000
ISSUE Total:	259,000		154,000		350,000
Grand totals:	840,000		441,000		1,020,250
Total Pages Accessed:	1,461,250				

B: Patent/Application Files Access Estimations

The following table describes the operations affecting files, yearly amount, and number of document pages likely to be involved.

Application Processing Functions	Transactions per year	Pages Inputted per Transaction		Pages Outputted per Transaction	
		Each	Annually	Each	Annually
RECEIVING (30+3 persons)					
1 receipt	35,000	1	35,000	30	1,050,000
2 ack letter/certificate	35,000	1	35,000	30	1,050,000
3 classification	35,000	10	350,000	15	525,000
4 registration request	49,000	0	0	0	0
5 registration letter	49,000	0	0	0	0
6 formality letter	5,250	0	0	0	0
7 formality response	5,250	1	5,250	30	157,500
RECEIVING Total:	213,500		425,250		2,782,500
18 MONTH PUBLICATION (10+1 persons)					
8 18 month publ.	35,000	1	35,000	30	1,050,000
EXAMINATION AREA (30+3 persons)					
9 receipt/classify	35,000	1	35,000	30	1,050,000
10 allowed search*	35,000	25	875,000	37.5	1,312,500
11 examination request	24,500	0	0	0	0
12 R40 letter	8,750	1	8,750	30	262,500
13 R40 response	8,750	1	8,750	1.5	13,125
14 order art*	24,500	10	245,000	90	2,205,000
15 examine*	24,500	1	24,500	30	735,000
16 copending search*	24,500	25	612,500	52.5	1,286,250
17 patent search*	24,500	50	1,225,000	84	2,058,000
18 lapsed search*	24,500	100	2,450,000	168	4,116,000
19 foreign search*	3,500	300	1,050,000	540	1,890,000
20 exam report	17,500	0	0	0	0
21 exam report resp	17,500	1	17,500	30	525,000
22 copending search*	24,500	25	612,500	37.5	1,312,500
23 allowance/classify*	24,500	1	24,500	1.5	36,750
24 abandonment	10,500	1	10,500	1.5	15,750
EXAMINATION area Total	332,500		7,199,500		16,818,375
ISSUE (25+2 persons)					
25 allowance response	24,500	0	0	0	0
26 grant/publication	24,500	1	24,500	30	735,000
27 renewal submit	105,000	1	105,000	2	210,000
28 renewal ack	105,000	0	0	0	0
ISSUE Total:	259,000		129,500		945,000
Grand totals:	840,000		7,789,250		21,595,875

*Operations with an asterisk are assumed to be conducted on examiner workstations.

C: Full-text Search Volume Estimations:

The following calculations are for determining the volume affected by full-text search.

35,000 applications/year
20%: full-text search
90%: affects 1000 documents
10%: affects 4000 documents
18 (Canadian) pages affected/document

$$35{,}000 \cdot 0.2 \cdot \left(0.9 \cdot 1{,}000 + 0.1 \cdot 4{,}000\right) = 9{,}100{,}000$$

NOTE: In the tables that follow, 'brute-force' search method will be examined for comparison purposes. 'Intelligent' methods exist, which practically do not access a document, but rather a special index file which is created when the document is stored. This difference is noted by the assumption that a 'brute-force' search method retrieves the whole document (18 pages) while an 'intelligent' method requires no more than the data amount equivalent of one page.

D: Page Access Summary

The following table describes the number of pages accessed by the main user groups.

Store/retrieval volumetrics:

	Pages	Access Requests
Receiving:	2,996,000	638,750
18 month publication:	1,400,000	70,000
Examination area:	2,162,125	203,000
Issue:	1,449,000	388,500
Examination workstations (file retrieval):	15,050,000	7,329,000
Accesses total:	23,057,125	8,629,250
Full-text search, brute force:	163,800,000	9,100,000
Total, with 'brute-force' full-text search:	186,857,125	17,729,250
Full-text search, intelligent:	9,100,000	9,100,000
Total, with 'Intelligent' full-text search:	32,157,125	17,729,250

E: Workload on different workstations:

	Number of workstations (personnel)	Seconds per document per workstation	Seconds per Page
Receiving:	30	278.98	59.48
18 month publication:	10	848.57	42.43
Examination area:	30	877.83	82.42
Issue:	25	382.24	102.48
Examination:	100	81.05	39.47

TABLE V: INTERNAL PEAK-LOAD ESTIMATIONS

Assumption: 35% of the transactions happen during a two-hour peak period.

Volumes processed in peak-time

	Pages	Documents
Receiving:	1,048,600	223,562.5
18 month publication:	490,000	24,500
Examination area:	756,743.75	71,050
Issue:	507,150	135,975
Examination workstations (file retrieval):	5,267,500	2,565,150
Full-text search, brute force:	57,330,000	3,185,000
Full-text search, intelligent:	3,185,000	3,185,000

On a per second per workstation basis:

	Workstations	Pages	Documents
Receiving:	30	0.022066	0.004705
18 month publication:	10	0.030934	0.001547
Examination area:	30	0.015925	0.001495
Issue:	25	0.012807	0.003434
Examination:	100	0.033254	0.016194
Full-text search, brute force:	100	0.361932	0.020108
Full-text search, intelligent:	100	0.020108	0.020108

To calculate system design criteria, the mathematical algorithm described in the next appendix was used.

The following options were considered worth examining:

Option #	Description
1	Groups 1-4 (Application processing)
2	Group 5 (Examination)
3	Groups 5-6 (Examination with brute-force search)
4	Groups 5,7 (Examination with intelligent search)
5	Groups 1-5 (Appl. processing and examination)
6	Groups 1-6 (Appl. proc., exam. and brute-search)
7	Groups 1-5,7 (Appl.p., exam. and intel. search)

The results:

Option #	Paralel throughput in terms of							
	No. of pages				No. of documents			
	90%	95%	99%	99.99%	90%	95%	99%	99.99%
1	4	4	5	8	5	6	7	11
2	6	6	8	12	21	22	25	31
3	46	48	52	59	43	45	49	58
4	8	9	11	16	43	45	49	58
5	8	9	11	15	24	26	29	35
6	48	50	54	62	47	49	53	62
7	11	12	14	19	47	49	53	62

TABLE VI: EXTERNAL USER COMMUNITY WORKLOAD

A. Derivation of Patent Agent Community Workload

Number of Patent Agents: 297
System Usage: 21,348 "uses"
Usage Distribution:

- CPO classification index 37 access/agent/year
- CPO tombstones 55 access/agent/year
- CPO abstracts and 1st claims 42 access/agent/year
- disclosure and claims 48 access/agent/year
- disclosure, claims and drawings 55 access/agent/year

NOTE: The above data results in 70,389 accesses to the CPO database. We can assume, that a "use" by a patent agent means multiple accesses (a search, for example). On the average, a "use" means 3.297 accesses to the CPO system (they use commercial databases and the USPTO system as well).

It is assumed, that the load is equally distributed over 220 working days, 10 hours/day (nationwide service).

The page and document volume totals derived from the above estimations:

	Document	Page
Classification	10,989	10,989
Tombstone	16,335	16,335
Abstracts	12,474	12,474
Text	14,256	256,608
Text & Drawing	16,335	326,700
TOTAL:	70,389	623,106

B. Derivation of the Non-patent-agent Community Workload

Total number of "uses" by the non patent agent community:

1,131,853 (mid-point) or 272,748 (worst-case)

Assumption:

For this user population, a "use" means one access of some kind to the system (unlike the assumption for patent agents).

The distribution over different kind of usages is calculated by using survey data about how important users find specific features. The answer in the questionnare is one of four degrees of importance - this will be used to weight (0, 1/3, 2/3, 1) the results.

	3	2	1	0	Weight
Classification	220	286	147	139	1379/2376
Tombstone	517	519	303	266	2892/4815
Abstract	353	238	97	114	1632/2406
Disclosure	195	281	181	142	1328/2397
Claims	170	215	225	175	1165/2355
Drawings	209	256	186	146	1325/2391

Using these figures, the absolute number of accesses can be calculated:

Access type	Number of access/year		Percent
	Mid-point	Worst-case	
Classification	189,739	45,722	16.76
Tombstone	196,354	47,317	17.35
Abstract	221,750	53,346	19.59
Disclosure	181,121	43,646	16.00
Claims	161,724	38,971	14.29
Drawings	181,165	43,656	16.01
TOTAL:	1,131,853	272,748	100.00

It is assumed, that the workload is equally distributed over 220 working days, 10 hours/day (nationwide service).

TABLE VII: EXTERNAL USERS - PEAK-LOADS

A. Patent Agents

Using the same method as for internal peak-load estimations, the following design criteria were established for various channel speeds:

Speed (Kbaud)	90%	99%	99.99%
1.2	6	9	13
2.4	4	5	8
4.8	2	4	6
9.6	1	3	5
19.2	1	2	4

Note: it is assumed that the channels are fully multiplexed and assigned to a user only when a transaction is taking place. Telephone lines, which are busy throughout the whole session regardless of the data amount transferred represent a counterexample.

B. Non patent agent user community

Two estimates are calculated: the mid-point and worst-case estimate on the number of accesses to the system.

Access type	Number of access/year		Percent
	Mid-point	Worst-case	
Classification	189,739	45,722	16.76
Tombstone	196,354	47,317	17.35
Abstract	221,750	53,346	19.59
Disclosure	181,121	43,646	16.00
Claims	161,724	38,971	14.29
Drawings	181,165	43,656	16.01
TOTAL:	1,131,853	272,748	100.00

Only 19.2 Kbaud circuits are calculated - the large demand rationalizes it.

The resulting number of channels:

	90%	99%	99.99%
Mid-point	4	6	9
Worst-case	1	3	5

TABLE VIII: CPU THROUGHPUT IN BYTES

Average page sizes involved in different options:
(for options, see table V.)

	DATIMTEX (byte)	IMAGE (Kbyte)
O1	7450.05	198.637
O2	7020.93	153.162
O3	5827.93	183.068
O4	6862.70	166.893
O5	7060.54	171.495
O6	5930.83	184.056
O7	6915.70	169.758

DATIMTEX: Data amount/sec (throughput) on the whole system:

Option #	Throughput (Kbytes/sec)			
	90%	95%	99%	99.99%
1	29.10	29.10	36.38	58.20
2	41.14	41.14	54.85	82.28
3	261.80	273.18	295.95	335.79
4	53.61	60.32	73.72	107.23
5	55.16	62.06	75.85	103.43
6	278.01	289.59	312.76	359.09
7	74.29	81.04	94.55	128.32

IMAGE: Data amount/sec (throughput) on the whole system:

Option #	Throughput (Kbytes/sec)			
	90%	95%	99%	99.99%
1	794.5	794.5	993.2	1589.1
2	919.0	919.0	1225.3	1837.9
3	8421.1	8787.3	9519.5	10801.0
4	1335.1	1502.0	1835.8	2670.3
5	1372.0	1543.5	1886.4	2572.4
6	8834.7	9202.8	9939.0	11411.5
7	1867.3	2037.1	2376.6	3225.4

Throughput requirements with external workload

Assumptions:

- DATIMTEX-type database
- "Intelligent" full-text search system
- Integrated system handling appl. proc., exam., and ext.
- "Mid-point" estimate for non-patent-agents
- 35% of local accesses happen during a 2-hour peak period
- External accesses are linearly distributed over 10 hours

Group	Workstats.	Page /Workstat/second	Document
Appl.proc	95	0.018624	0.003024
Exam.	100	0.033254	0.016194
External	100	0.006055	0.001518

Throughput requirements:

	90%	95%	99%	99.99%
Documents	26	28	31	38
Pages	9	10	12	16

Average page size: 7036.88 bytes for these transactions

Local facilities throughput (Kbytes/sec):

90%	95%	99%	99.99%
61.85	68.72	82.46	109.95

A.2. Probabilistic Method for Peak-load Calculations

[Paragraph on ASCII representation of equations erased]

When designing multi-user systems (a local area network, for example) one encounters the problem of capacity requirement calculations. When there is a large number of users accessing the same system, a question arises: what throughput capacity the system should be designed for.

It is very unlikely, that the throughput serving one user is enough. There is a probability that transactions from more than one users overlap. It is also unlikely that all the users happen to access the system at the same time - therefore there is no need to overdesign the system to such a peak. The question is, how can a realistic estimate be derived?

A simple mathematical approach follows to solve this problem.

A.2.1. Transaction Probabilities

The first question to be answered, is that: what is the probability that at a given time, exactly k transactions are in progress, given transaction frequency, elapsed transaction time and the number of users?

Solutions:

A. All users belong to the same group, executing the same kind of transaction.

known data:

number of users: n
transaction frequency: f
transaction length: t (number of transactions a second at full throughput)

The probability, that at a given time, a selected workstation is in process of a transaction is:

$$p_s = ft$$

The probability, that of n workstations, at a given time, exactly k are in process of a transaction, can be calculated using the binomial distribution:

$$p_k = \binom{n}{k} (ft)^k (1 - ft)^{n-k} \qquad (1)$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n)$$

For large n-s (over 10), or when n is not known (therefore f is also unknown, only f*n is known) the above formula can be approximated by the Poisson-distribution:

$$p_k = \frac{(nft)^k}{k!} e^{-nft}$$

B. Users belong to a number of groups where they execute the same kind of transaction within each group but transactions differ between groups.

Known data:

number of subgroups: N
number of workstations in group i: n_i
transaction frequency in group i: f_i
Elapsed transaction time: t_i

The probability of k transactions within a group can be calculated as described above. The probability of exactly K transactions happening at a given time on the whole system is:

$$P_K = \sum_{k_1=0, k_2=0, \ldots k_n=0}^{k_1+k_2+\ldots+k_n=K} \left(\prod_{i=1}^{N} p\left(n_i, k_i, f_i t_i\right) \right) \qquad (2)$$

where

$$p\left(n_i, k_i, f_i t_i\right)$$

is the probability distribution (binomial or Poisson) described above.

In case of large n_i-s, P_K can be approximated with the formula:

$$P_K = p\left(\sum n_i, K, \frac{\sum f_i n_i}{\sum n_i t_i} \right) \qquad (3)$$

A.2.2. System speed calculations when probability is known

The practical problem faced here is somewhat different. The probability (or rather the sum of some probabilities) is well known being a system design criteria (i.e. 95% of requests should be completed within a given time) and the question is, how many simultanous requests the system should be capable to handle.

Known parameters:

Number of subsystems: N
Number of workstations in subsystem i: n_i
Transaction frequency in subsystem i: f_i
Maximum elapsed transaction time: t_i
Probability requirement: d

where d is a number between 0 and 1, expressing the probability that a transaction will not take more than t amount of time.

The requirement is, that in at least $d \cdot 100$ per cent of the cases, a transaction should not last longer than t amount of time.

Question: which is the number K, for which the following are true:

- if the system is capable to handle at least $\frac{n}{\sqrt[n]{dn+k!}}$ transactions, the requirement is met
- if the system is capable to handle no more than $K - 1$ transactions, the requirement is not met.

The system speed can be expressed by dividing the maximum transaction length by the maximum number of simultaneous transactions.

Using (2), this requirement can be expressed:

$$\sum_{k=0}^{K} \left(\sum_{k_1=0, k_2=0, \ldots k_n=0}^{k_1+k_2+\ldots k_n=K} \left(\prod_{i=1}^{n} p(n_i, k_i, f_i t_i) \right) \right) \geq d$$

and

$$\sum_{k=0}^{K-1} \left(\sum_{k_1=0, k_2=0, \ldots k_n=0}^{k_1+k_2+\ldots k_n=K} \left(\prod_{i=1}^{n} p(n_i, k_i, f_i t_i) \right) \right) < d \qquad (4)$$

A.2.3. Calculations performed for CPO throughput estimates

The proposed CPO implementation plan contains two large-capacity multi-user subsystems: a local area network for internal data traffic (patent documents to workstations, for example) and a wide area facility for external users (patent agents, for example).

A tool for both calculations was a 'C' program which implemented formula (4) for different sets of data, trying different K-s (incrementally) until the one confirming to the requirements was found.

Each group of calculations was made for three different d-s (probability requirements): 0.9, 0.99 and 0.9999 (meaning that in 90%, 99% and 99.99% of the cases, respectively, the transaction time will not exceed the criteria).

For LAN capacity requirements, the following data was used:

No. of groups: 7

No.	Group	Workstations	Page freq.	Document freq.
APPL. PROC.				
1	Receiving	30	0.022066	0.004705
2	18 mth publ.	10	0.030934	0.001547
3	Exam. area	30	0.015925	0.001495
4	Issue	25	0.012807	0.003434
EXAMINATION				
5	Examination	100	0.033254	0.033254
FULL-TEXT SRCH				
6	Brute force	100	0.361932	0.020108
7	Intelligent	100	0.020108	0.020108

NOTE: The number of workstations was based on CPO estimates. These numbers may vary (being large) without causing significant changes in the results. An assumption was essential for the use of the binomial distribution which, in turn, was essential for the required calculation accuracy (.9999/1 meaning 0.01% accuracy requirement).

The frequency data was based on CPO estimates about different application processing steps, our assumptions about the meaning of these steps and our assumptions about the data amount involved in examination/search.

Brute-force search was assumed to require the retrieval of the whole document while intelligent search is assumed to require the data amount equivalent of no more than one page.

The following options were considered worth examining:

Option #	Description
1	Groups 1-4 (Application processing)
2	Group 5 (Examination)
3	Groups 5-6 (Examination with brute-force search)
4	Groups 5,7 (Examination with intelligent search)
5	Groups 1-5 (Appl. processing and examination)
6	Groups 1-6 (Appl. proc., exam. and brute-search)
7	Groups 1-5,7 (Appl.p., exam. and intel. search)

The results:

Option #	Paralel throughput in terms of							
	No. of pages				No. of documents			
	90%	95%	99%	99.99%	90%	95%	99%	99.99%
1	4	4	5	8	5	6	7	11
2	6	6	8	12	21	22	25	31
3	46	48	52	59	43	45	49	58
4	8	9	11	16	43	45	49	58
5	8	9	11	15	24	26	29	35
6	48	50	54	62	47	49	53	62
7	11	12	14	19	47	49	53	62

The actual database implementation defines page sizes and therefore data amounts which, in turn, define LAN requirements explicitly (in Kbytes/sec instead of pages/sec). Also, the implementation gives meaning to the number of simultaneous accesses.

The WAN capacity measures were taken as follows:

First, two main user groups are distinguished: Patent Agents and other users. The capacity requirements presented here are conclusions drawn from the Task 4 report.

The methodology is similar to the one used document access estimations.

A. Patent Agents

The number of patent agents is 297, distributed throughout the country. The following data is available:

Type of access	Projected number of access/year/agent
CPO Classification	37
Tombstone Index	55
Abstract, 1st claim	42
Disclosure & Claims	48
Disclosure, Claims, Drawings	55

The number of workstations in each group equals to the number of agents (297).

The transaction length (and therefore the speed) can be determined by knowing the data amount involved in each type of access, which is determined by our preferred implementation strategy.

The resulting number of channels:

Speed (Kbaud)	90%	99%	99.99%
1.2	6	9	13
2.4	4	5	8
4.8	2	4	6
9.6	1	3	5
19.2	1	2	4

Note, that it is assumed that the channels are fully multiplexed and assigned to a user only when a transaction is taking place. Telephone lines, which are busy throughout the whole session regardless of the data amount transferred are a counterexample.

B. Non patent agent user community

Two estimates are calculated: the mid-point and worst-case estimate on the number of accesses to the system. The distribution over different access types is calculated using some Task 5 survey results, as it is described elsewhere. The projected estimates are:

Access type	Number of access/year		Percent
	Mid-point	Worst-case	
Classification	189,739	45,722	16.76
Tombstone	196,354	47,317	17.35
Abstract	221,750	53,346	19.59
Disclosure	181,121	43,646	16.00
Claims	161,724	38,971	14.29
Drawings	181,165	43,656	16.01
TOTAL:	1,131,853	272,748	100.00

Not knowing the number of users, a large number (100) is assumed, when the binomial distribution will be fairly close to Poisson-distribution results.

Only 19.2 Kbaud circuits are calculated - the large demand rationalizes it.

The resulting number of channels:

	90%	99%	99.99%
Mid-point	4	6	9
Worst-case	1	3	5

The meaning of these results is:

- Assuming that a document access request is serviced in 10 seconds, the probability of no more than 38 overlapping requests is more than 99.99%. The probability of more than 38 overlapping requests is therefore less than 0.01%. The hardware designer should take into account that there is a probability, that two request arrive to the same device and one of them

must be queued - the number of storage devices (with 10 second access time, e.g. jukebox) is certainly more than 38 if the requirements are to be met.

- A network of more than 109.95 Kbyte (0.86 Mbit) per second effective throughput in worst-case ensures, that the probability of a transmit request being serviced within 1 second is more than 99.99%. With other words, less than 0.01% of the requests will not be serviced in one second. Naturally, the file server(s) should be compatible to these throughput figures.

APPENDIX B: COST COMPARISON OF THE OPTIONS

OPTION 1: Index database

	1988/89	1989/90	1990/91	1991-96	Total
1. Salaries - ASB Staff	600	900	900	4500	6900
2. Project Management	550	890	300	0	1740
3. Hardware	150	4000	1000	0	5150
4. System Software		500			500
5. Appl. SW Development	100	400	0	0	500
6. Implementation		50	0	0	50
7. Facilities/Preparation	180				180
8. Backfile Conversion	500	500	0		1000
9. Extern Database Access Setup	10	50			60
10. Marketing & Promotion				150	150
11. File Translation	0				0
12. Operations Management	110	100	185	547	942
13. Computer Operations		200	200	1000	1400
14. Hardware Maintenance			240	1500	1740
15. Communications	10	10	10	150	180
16. Software Licences			100	500	600
17. Extern Database Usage				300	300
18. Appl. SW Maintenance		50	50	250	350
19. Training & User Support		25	25	100	150
20. Facilities		150	150	750	1050
21. On-going Data Capture		0	0	0	0
22. Publishing					0
23. Translation			100	0	100
Totals:	2210	7825	3260	9747	23042

Summary

Equipment & Maintenance	150	4000	1240	1500	6890
Backfile Conversion	500	500	0	0	1000
On-going Data Capture	0	0	0	0	0
Software	100	1000	150	750	2000
Facilities & Training	180	175	175	850	1380
Operations & Communications	10	210	210	1150	1580
Project Consulting Support	550	890	300	0	1740
Salaries	600	900	900	4500	6900
Operations Management	110	100	185	547	942
Other	10	50	100	450	610
Total:	2210	7825	3260	9747	23042

Summary 2

Salaries	600	900	900	4500	6900
Capital	330	4500	1100	500	6430
O&M	1280	2425	1260	4747	9712
Total:	2210	7825	3260	9747	23042

OPTION 2: Separate text and image holdings

	1988/89	1989/90	1990/91	1991-96	Total
1. Salaries - ASB Staff	600	900	1400	7000	9900
2. Project Management	550	890	600	500	2540
3. Hardware	150	8500	5000	3000	16650
4. System Software		500			500
5. Appl. SW Development	100	800	710	2237	3847
6. Implementation		50	50	250	350
7. Facilities/Preparation	180				180
8. Backfile Conversion	500	2600	2000		5100
9. Extern Database Access Setup	10	50			60
10. Marketing & Promotion				300	300
11. File Translation	10				10
12. Operations Management	110	100	370	2183	2763
13. Computer Operations		200	600	3000	3800
14. Hardware Maintenance			510	4122	4632
15. Communications	10	10	10	1500	1530
16. Software Licences			100	500	600
17. Extern Database Usage				300	300
18. Appl. SW Maintenance		150	300	750	1200
19. Training & User Support		25	25	300	350
20. Facilities		250	250	1250	1750
21. On-going Data Capture		970	510	2550	4030
22. Publishing					0
23. Translation			100	500	600
Totals:	2220	15995	12535	30242	60992

Summary

Equipment & Maintenance	150	8500	5510	7122	21282
Backfile Conversion	500	2600	2000	0	5100
On-going Data Capture	0	970	510	2550	4030
Software	100	1500	1160	3737	6497
Facilities & Training	180	275	275	1550	2280
Operations & Communications	10	210	610	4500	5330
Project Consulting Support	550	890	600	500	2540
Salaries	600	900	1400	7000	9900
Operations Management	110	100	370	2183	2763
Other	20	50	100	1100	1270
Total:	2220	15995	12535	30242	60992

Summary 2

Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	19742	33162
Total:	2220	15995	12535	30242	60992

OPTION 3: Functionally integrated text and image database

	1988/89	1989/90	1990/91	1991-96	Total
1. Salaries - ASB Staff	600	900	1400	7000	9900
2. Project Management	550	890	600	750	2790
3. Hardware	150	8500	5000	3000	16650
4. System Software		500			500
5. Appl. SW Development	100	800	710	3356	4966
6. Implementation		50	50	250	350
7. Facilities/Preparation	180				180
8. Backfile Conversion	500	2600	2000		5100
9. Extern Database Access Setup	10	50			60
10. Marketing & Promotion				600	600
11. File Translation	10				10
12. Operations Management	110	100	370	2183	2763
13. Computer Operations		200	600	3000	3800
14. Hardware Maintenance			510	4122	4632
15. Communications	10	10	10	1500	1530
16. Software Licences			100	500	600
17. Extern Database Usage				300	300
18. Appl. SW Maintenance		150	300	1125	1575
19. Training & User Support		25	25	300	350
20. Facilities		250	250	1250	1750
21. On-going Data Capture		970	510	2550	4030
22. Publishing					0
23. Translation			100	500	600
Totals:	2220	15995	12535	32286	63036

Summary

Equipment & Maintenance	150	8500	5510	7122	21282
Backfile Conversion	500	2600	2000	0	5100
On-going Data Capture	0	970	510	2550	4030
Software	100	1500	1160	5231	7991
Facilities & Training	180	275	275	1550	2280
Operations & Communications	10	210	610	4500	5330
Project Consulting Support	550	890	600	750	2790
Salaries	600	900	1400	7000	9900
Operations Management	110	100	370	2183	2763
Other	20	50	100	1400	1570
Total:	2220	15995	12535	32286	63036

Summary 2

Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	21786	35206
Total:	2220	15995	12535	32286	63036

OPTION 4: Fully integrated text and image database

	1988/89	1989/90	1990/91	1991-96	Total
1. Salaries - ASB Staff	600	900	1400	7000	9900
2. Project Management	550	890	600	1000	3040
3. Hardware	150	8500	5000	3000	16650
4. System Software		500			500
5. Appl. SW Development	100	800	710	4475	6085
6. Implementation		50	50	250	350
7. Facilities/Preparation	180				180
8. Backfile Conversion	500	2600	2000		5100
9. Extern Database Access Setup	10	50			60
10. Marketing & Promotion				750	750
11. File Translation	10				10
12. Operations Management	110	100	370	2183	2763
13. Computer Operations		200	600	3000	3800
14. Hardware Maintenance			510	4122	4632
15. Communications	10	10	10	1500	1530
16. Software Licences			100	500	600
17. Extern Database Usage				300	300
18. Appl. SW Maintenance		150	300	1500	1950
19. Training & User Support		25	25	300	350
20. Facilities		250	250	1250	1750
21. On-going Data Capture		970	510	2550	4030
22. Publishing					0
23. Translation			100	500	600
Totals:	2220	15995	12535	34180	64930

Summary

Equipment & Maintenance	150	8500	5510	7122	21282
Backfile Conversion	500	2600	2000	0	5100
On-going Data Capture	0	970	510	2550	4030
Software	100	1500	1160	6725	9485
Facilities & Training	180	275	275	1550	2280
Operations & Communications	10	210	610	4500	5330
Project Consulting Support	550	890	600	1000	3040
Salaries	600	900	1400	7000	9900
Operations Management	110	100	370	2183	2763
Other	20	50	100	1550	1720
Total:	2220	15995	12535	34180	64930

Summary 2

Salaries	600	900	1400	7000	9900
Capital	330	9000	5100	3500	17930
O&M	1290	6095	6035	23680	37100
Total:	2220	15995	12535	34180	64930

APPENDIX C: DETAILED IMPLEMENTATION PLAN

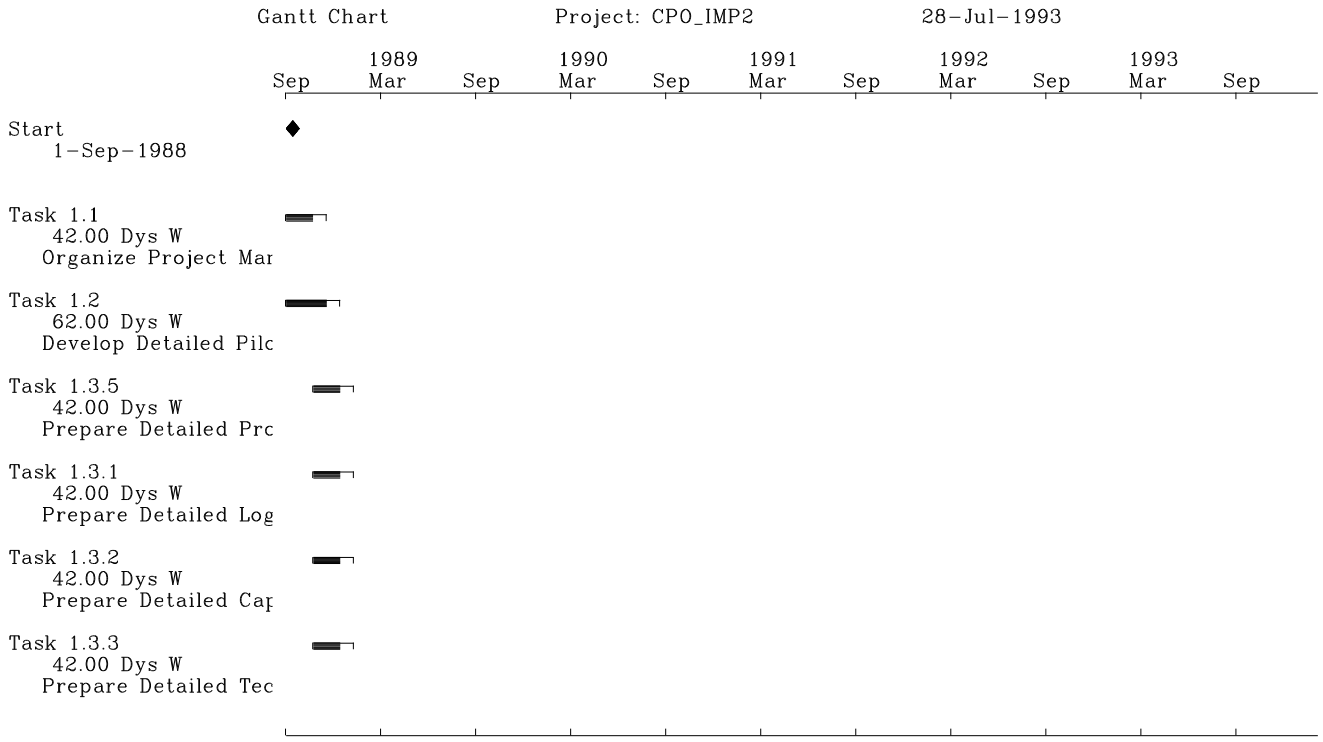
C.1. Implementation Plan: Task and Milestone List

Name	Start Date	Duration (approximate)	Description
Start	88-09-01	(M)	Project Approval
Task 1.1	88-09-01	42.00	Organize Project Management Team
Task 1.2	88-09-01	62.00	Develop Detailed Pilot Project Team
Task 1.3.1	88-11-01	42.00	Prepare Detailed Logical Design
Task 1.3.2	88-11-01	42.00	Prepare Detailed Capacity
Task 1.3.3	88-11-01	42.00	Prepare Detailed Technology Plan
Task 1.3.4	88-11-01	42.00	Prepare Detailed Design Requirements
Task 1.3.5	88-11-01	42.00	Prepare Detailed Project Program Schedules
Task 1.4	88-11-29	62.00	Award Vendor Contracts for Pilot Projects
Task 1.5	88-12-30	42.00	Prepare, Issue & Evaluate RFP's to design, engineer, furnish & install
Task 1.6	89-02-27	62.00	Perform Pilot Projects & Acceptance Testing
Task 1.7	89-05-24	62.00	Evaluate Pilot Projects
Task 1.8	89-03-01	625.00	Backfile Conversion
Task 2.1	89-08-17	(M)	Review Project Implementation Plan
Task 2.2	89-08-18	50.00	Issue Vendor Contracts
Task 2.3	89-10-30	50.00	Prepare Facilities & Install Hardware
Task 2.4	90-01-12	50.00	Perform Hardware & Systems Software Acceptance
Task 2.5	90-03-23	50.00	Develop Operations Practices & Procedures
Task 2.6	90-06-01	50.00	Recruit, Train & Staff Operations Functions
Task 2.7	90-03-2	50.00	Proceed with Applications Software Development
Task 2.8	90-03-23	150.00	Develop Testbed Databases
Task 2.9	90-06-01	50.00	Complete Unit & Integration Testing for Applications Processing Systems on Image, Text & ITI Databases
Task 2.10	90-06-01	50.00	Complete Unit & Integration Testing for Index & Classification Search Systems & Full-Text Search on Text Databases
Task 2.11	90-08-09	(M)	Review Implementation Plan
Task 3.1	90-10-22	50.00	Test & Evaluate Applications Processing Systems on the Image, Text and ITI Databases in the Testbed
Task 3.2	90-10-22	50.00	Test & Evaluate Index & Classification Search Systems & Full-Text Search Systems on the Text Database in the Testbed
Task 3.3	90-08-10	50.00	Proceed with Development of ITI Full-Text Search Systems
Task 3.5	90-10-22	50.00	Complete Unit & Integration Testing for Full-Text Search on the ITI Database

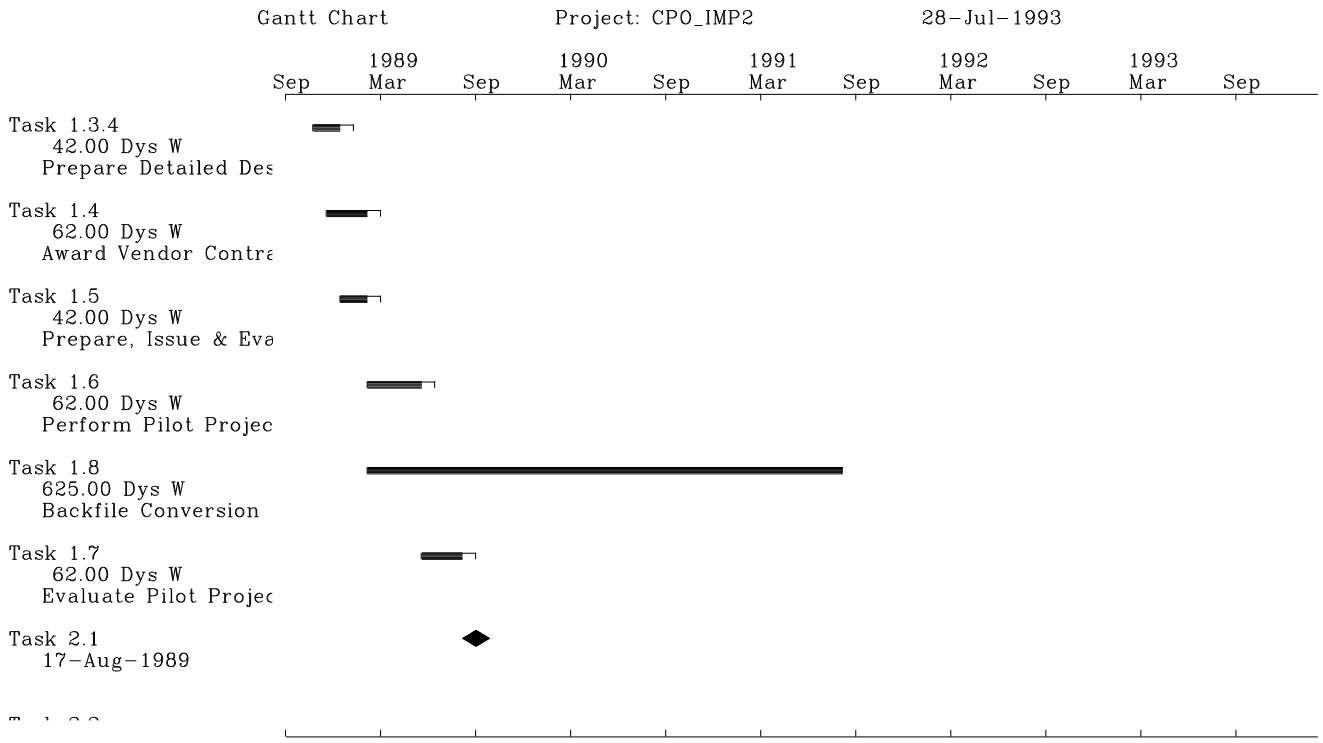
Task 3.6	90-08-10	250.00	Proceed with Development of Dissemination Systems
Task 3.7	91-01-04	50.00	Develop Operations Integration Plan
Task 3.8	91-03-15	50.00	Develop User Training Programs for Applications Processing and Search Systems
Task 3.9	91-05-24	50.00	Implement Applications Processing & Search Systems User Training Programs
Task 3.10	91-08-07	(M)	Review Implememtation Plan
Task 4.1	91-08-08	250.00	Deploy Applications Processing Systems
Task 4.2	91-08-08	125.00	Deploy Index & Classification Search Systems & Full-Text Search Systems on the Text Database
Task 4.3	92-02-06	125.00	Test & Evaluate Index & Classification Search Systems & Full-Text Search Systems on the ITI Database in the Testbed
Task 4.4	91-08-05	62.00	Test & Evaluate Dissemination Systems on the ITI Database in the Testbed
Task 4.5	91-10-31	62.00	Develop User Training Programs for Dissemination Systems
Task 4.6	92-01-31	62.00	Implement User Training Programs for Dissemination Systems
Task 4.7	92-08-21	(M)	Review Implementation Plan
Task 5.1	92-08-24	250.00	Deploy Index & Classification Search Systems & Full-Text Search Systems on the ITI Database
Task 5.2	92-08-24	250.00	Deploy Dissemination Systems for Internal Users
Task 5.3	92-08-24	125.00	Proceed with Definition of External User Interfaces to the Patent Information Processing System
Task 5.4	93-02-18	125.00	Plan Technical Field Trial for External Users
Task 5.5	93-08-20	(M)	Review Implementation Plan
Task 6.1	93-08-13	84.00	Proceed with Development of External User Interfaces to Full-Text Search Systems on the ITI Database
Task 6.2	93-08-13	84.00	Proceed with Development of External User Interfaces to the Application Processing Systems on the ITI Database
Task 6.3	93-08-13	84.00	Install Communications Facilities for Remote Access to Patent Information Processing Systems
Task 6.4	93-12-30	84.00	Complete Unit & Integration Testing of External User Interfaces to Full-Text Search Systems
Task 6.5	93-12-30	84.00	Complete Unit & Integration Testing of External User Interfaces to Applications Processing Systems
Task 6.6	93-08-23	250.00	Proceed with Development of User Accounting Systems
Task 6.7	94-04-27	84.00	Develop Training Program for External Users
Task 6.8	94-04-27	84.00	Deploy Remote Workstations for Technical Field Trial

Task 6.9	94-08-23	(M)	Review Implementation Plan
Task 7.1	94-08-24	42.00	Deliver Training Program to External Users Participating in Technical Field Trial
Task 7.2	94-11-01	42.00	Proceed with Technical Field Trial to Evaluate User Interfaces to the the ITI Databases
Task 7.3	94-11-01	42.00	Proceed with Technical Field Trial to Evaluate User Interfaces to the Full-Text Search System on the ITI Database
Task 7.4	94-11-01	42.00	Proceed with Technical Field Trial to Evaluate User Interfaces to the Information Dissemination Systems on the ITI Database
Task 7.5	94-08-24	84.00	Complete Unit & Integration Testing of User Accounting Systems
Task 7.6	94-12-30	42.00	Test User Accounting Systems on Field Trial Users
Task 7.7	95-03-01	42.00	Complete Technical Field Trial Evaluations
Task 7.8	95-04-28	42.00	Develop Communications Program for General Users
Task 7.9	95-04-28	42.00	Develop Training Programs for General Users
Task 7.10	95-04-28	42.00	Expand Communications Facilities for Remote Users
Task 7.11	95-06-27	42.00	Deploy External User Interfaces for General Users
Task 7.12	95-03-01	125.00	Deploy User Accounting Systems
Task 7.13	95-08-23	(M)	Review Implementation Plan
Task 8.1	95-08-24	250.00	Promote Communications Program
Task 8.2	95-08-24	250.00	Deliver Training Programs to General Users
Task 8.3	95-08-24	250.00	Evaluate External Usage
Task 8.4	96-08-22	(M)	Review Implementation Plan
End	96-08-22	(M)	Completion of Implementation of the Automation Project for then Canadian Patent Office (CCAC)

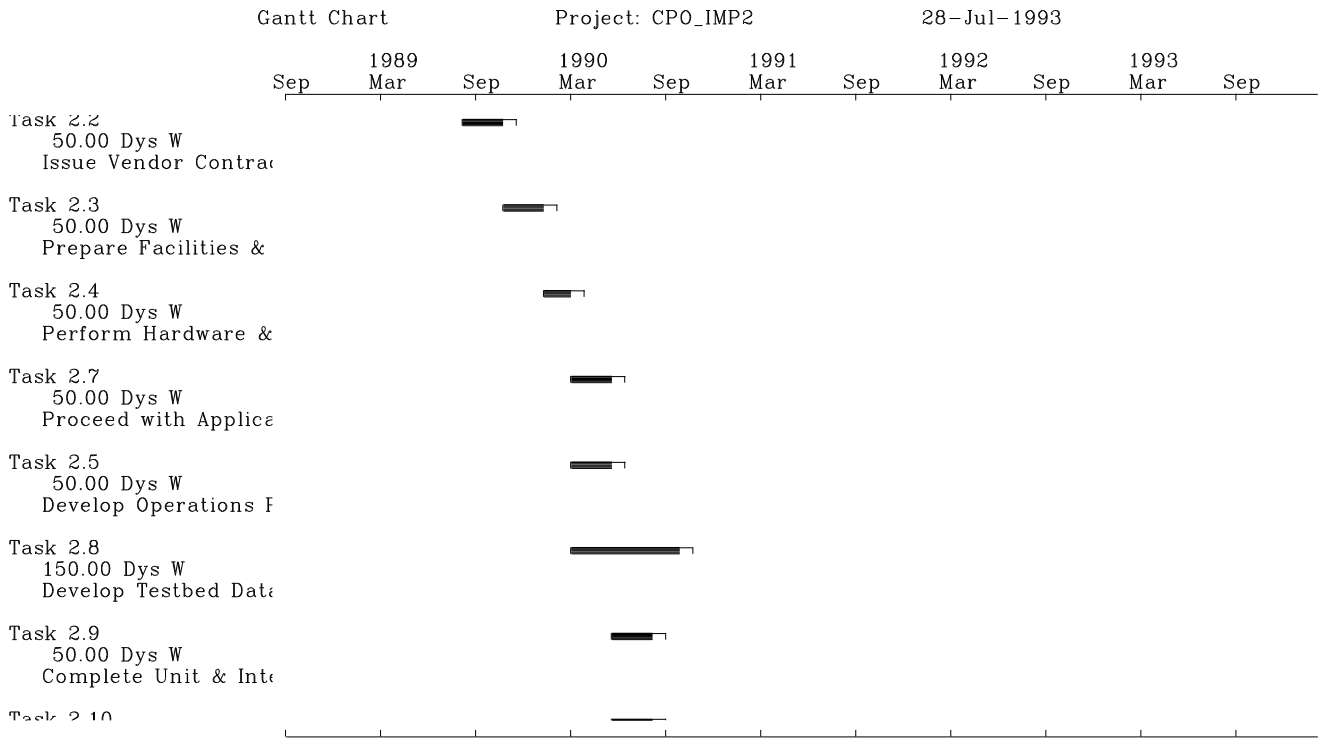
C.2. Gantt Chart



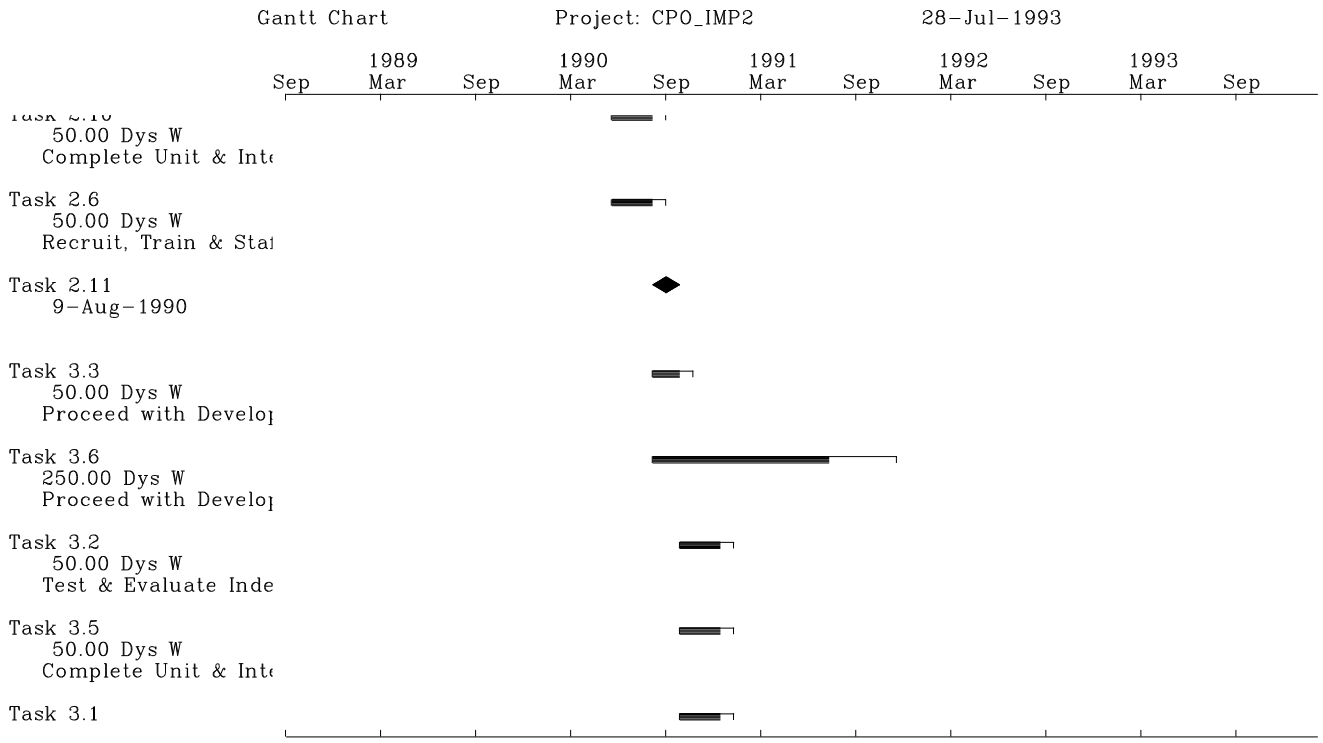
Page 1-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Task 1.3.4
42.00 Dys W
Prepare Detailed Des
Task 1.4
62.00 Dys W
Award Vendor Contra
Task 1.5
42.00 Dys W
Prepare, Issue & Eva
Task 1.6
62.00 Dys W
Perform Pilot Projec
Task 1.8
625.00 Dys W
Backfile Conversion
Task 1.7
62.00 Dys W
Evaluate Pilot Projec
Task 2.1
17-Aug-1989

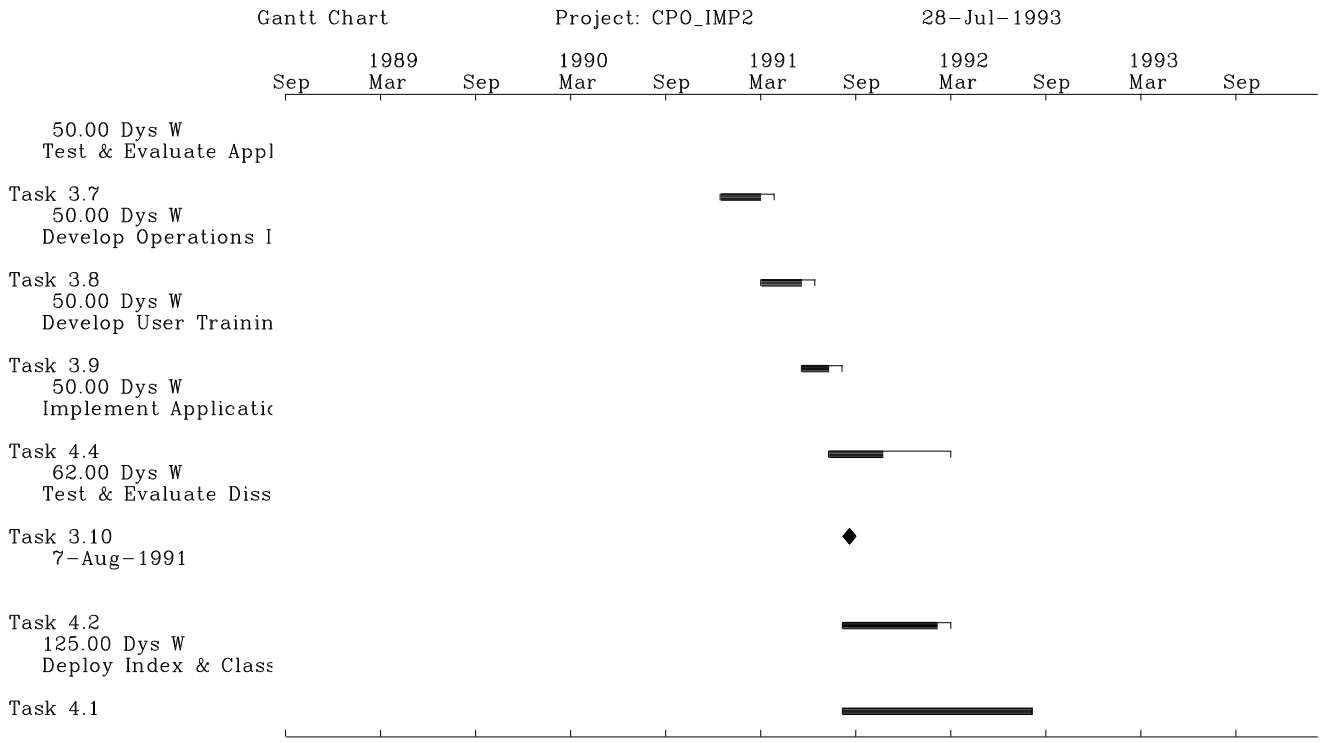
Page 2-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Task 2.2
50.00 Dys W
Issue Vendor Contra
Task 2.3
50.00 Dys W
Prepare Facilities &
Task 2.4
50.00 Dys W
Perform Hardware &
Task 2.7
50.00 Dys W
Proceed with Applica
Task 2.5
50.00 Dys W
Develop Operations F
Task 2.8
150.00 Dys W
Develop Testbed Data
Task 2.9
50.00 Dys W
Complete Unit & Inte
Task 2.10

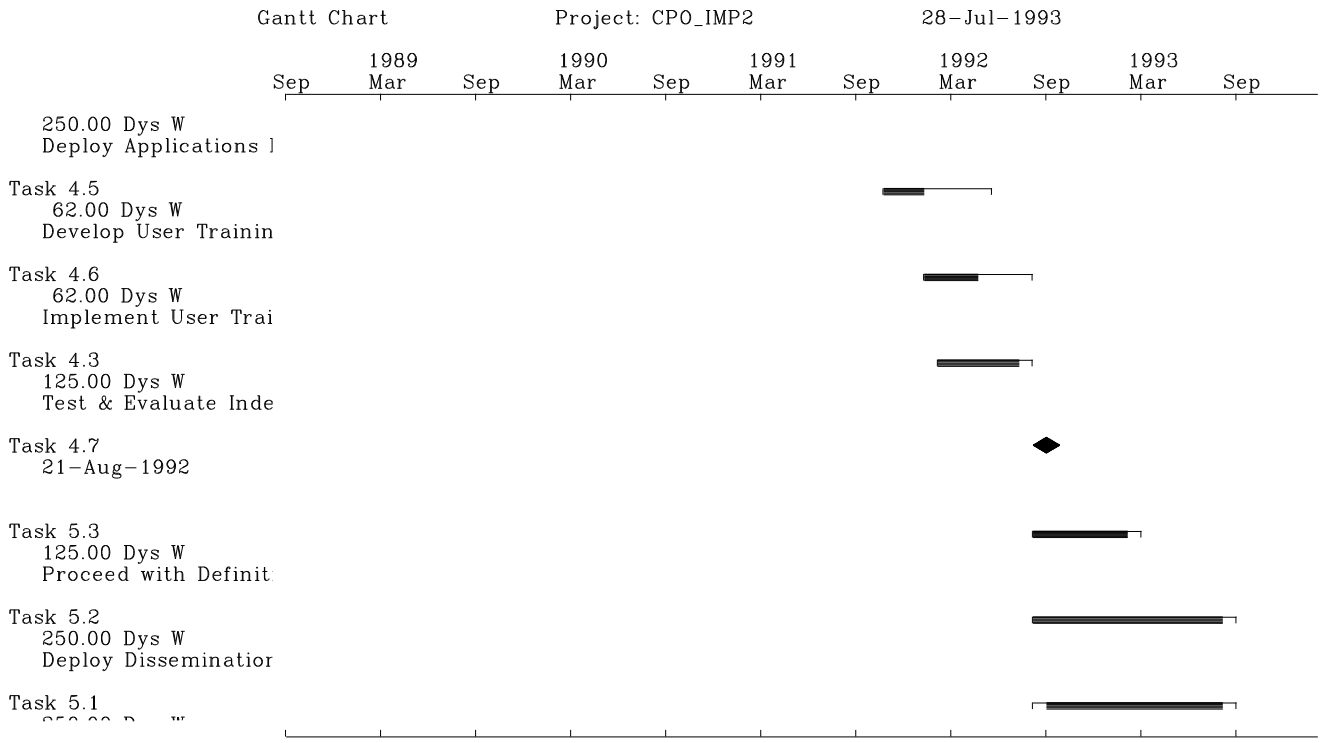
Page 3-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
50.00 Dys W
Complete Unit & Int
Task 2.6
50.00 Dys W
Recruit, Train & Sta
Task 2.11
9-Aug-1990
Task 3.3
50.00 Dys W
Proceed with Develo
Task 3.6
250.00 Dys W
Proceed with Develo
Task 3.2
50.00 Dys W
Test & Evaluate Inde
Task 3.5
50.00 Dys W
Complete Unit & Int
Task 3.1

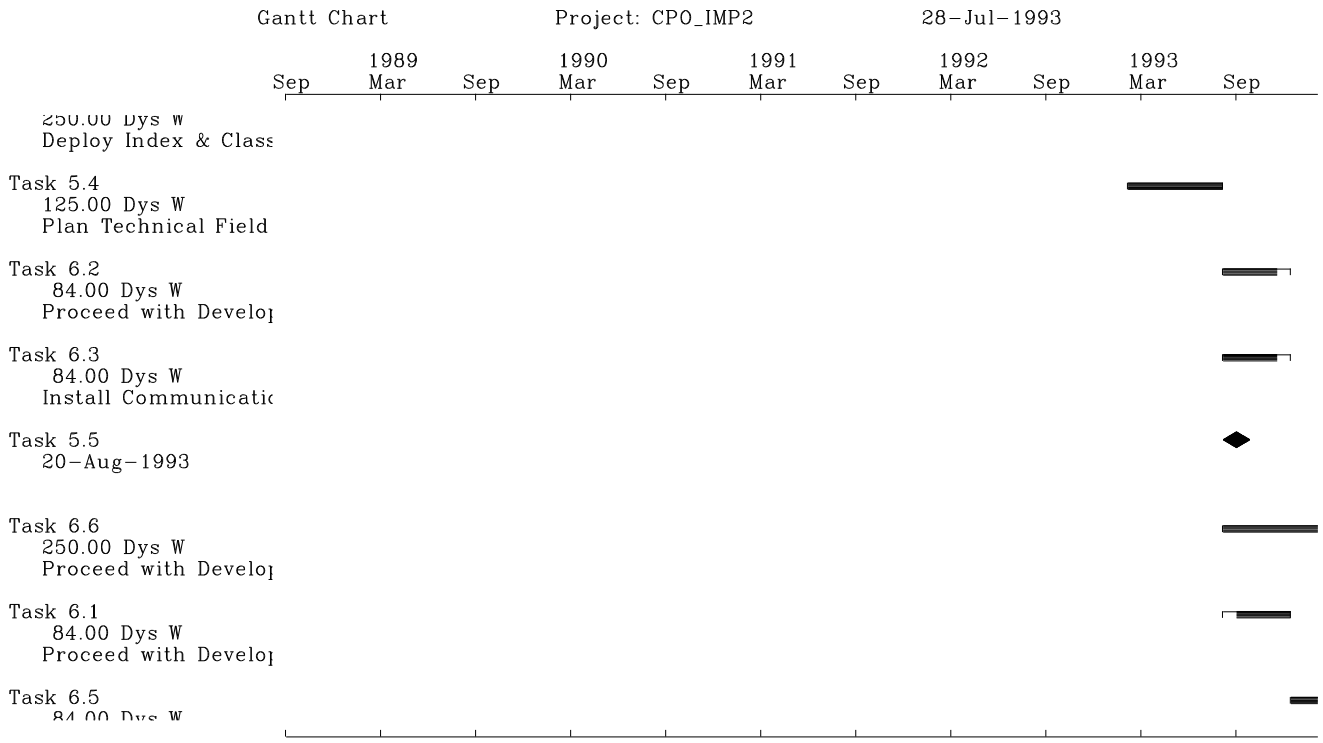
Page 4-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
50.00 Dys W
Test & Evaluate Appl
Task 3.7
50.00 Dys W
Develop Operations I
Task 3.8
50.00 Dys W
Develop User Trainin
Task 3.9
50.00 Dys W
Implement Applicatic
Task 4.4
62.00 Dys W
Test & Evaluate Diss
Task 3.10
7-Aug-1991
Task 4.2
125.00 Dys W
Deploy Index & Class
Task 4.1

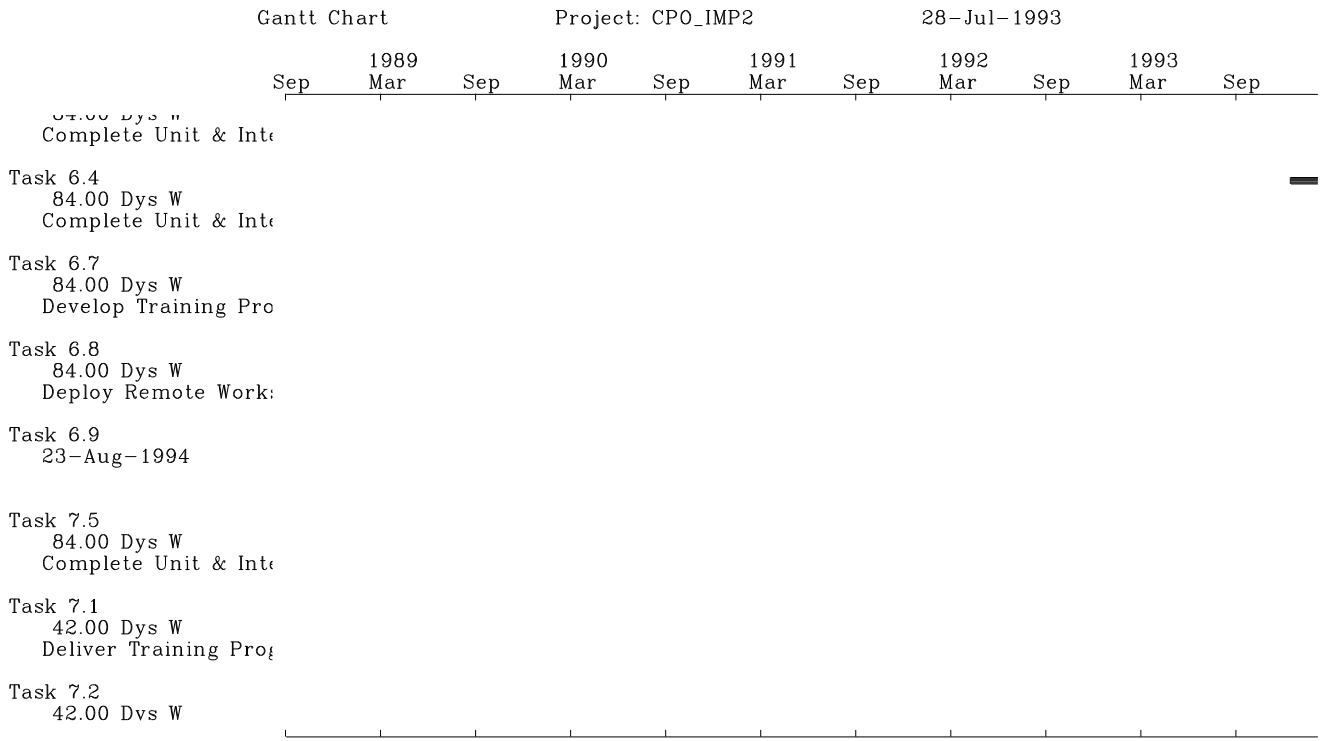
Page 5-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
250.00 Dys W
Deploy Applications
Task 4.5
62.00 Dys W
Develop User Trainin
Task 4.6
62.00 Dys W
Implement User Trai
Task 4.3
125.00 Dys W
Test & Evaluate Inde
Task 4.7
21-Aug-1992
Task 5.3
125.00 Dys W
Proceed with Definit
Task 5.2
250.00 Dys W
Deploy Dissemination
Task 5.1

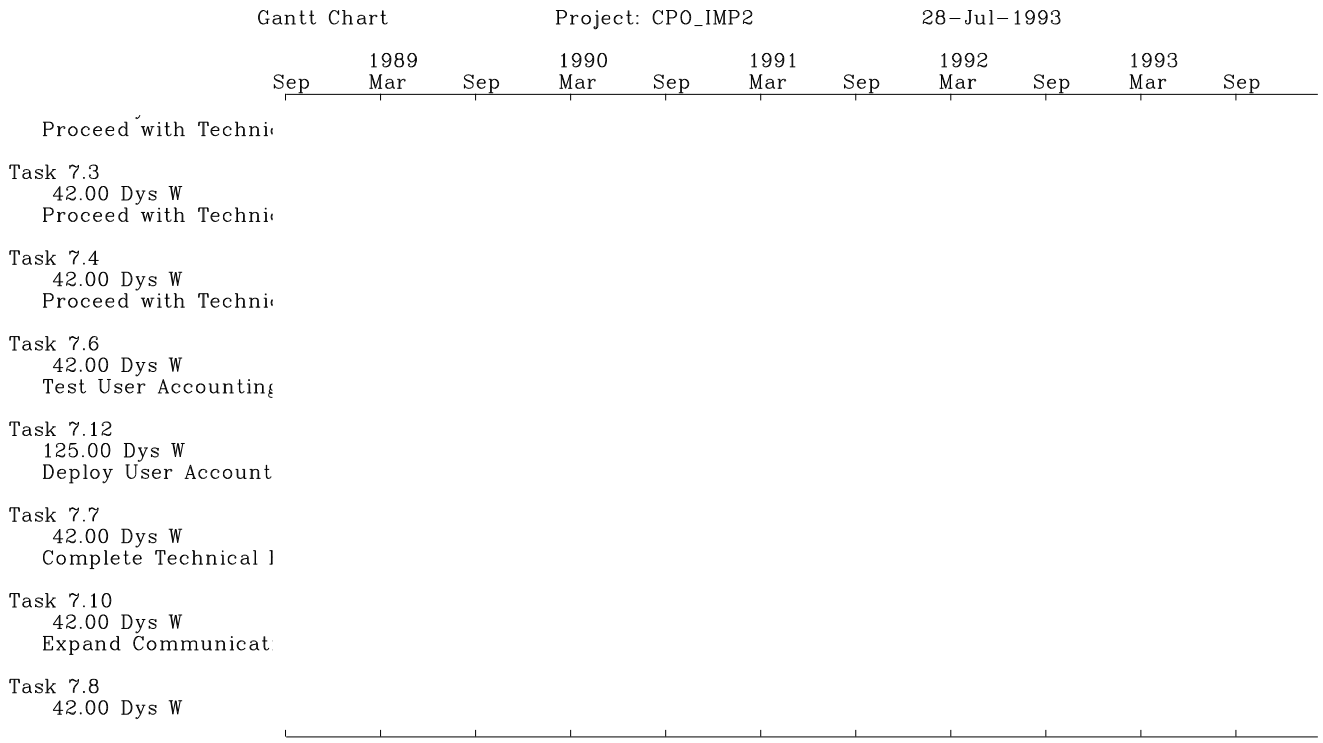
Page 6-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
250.00 Dys W
Deploy Index & Class
Task 5.4
125.00 Dys W
Plan Technical Field
Task 6.2
84.00 Dys W
Proceed with Develo
Task 6.3
84.00 Dys W
Install Communicatic
Task 5.5
20-Aug-1993
Task 6.6
250.00 Dys W
Proceed with Develo
Task 6.1
84.00 Dys W
Proceed with Develo
Task 6.5
84.00 Dys W

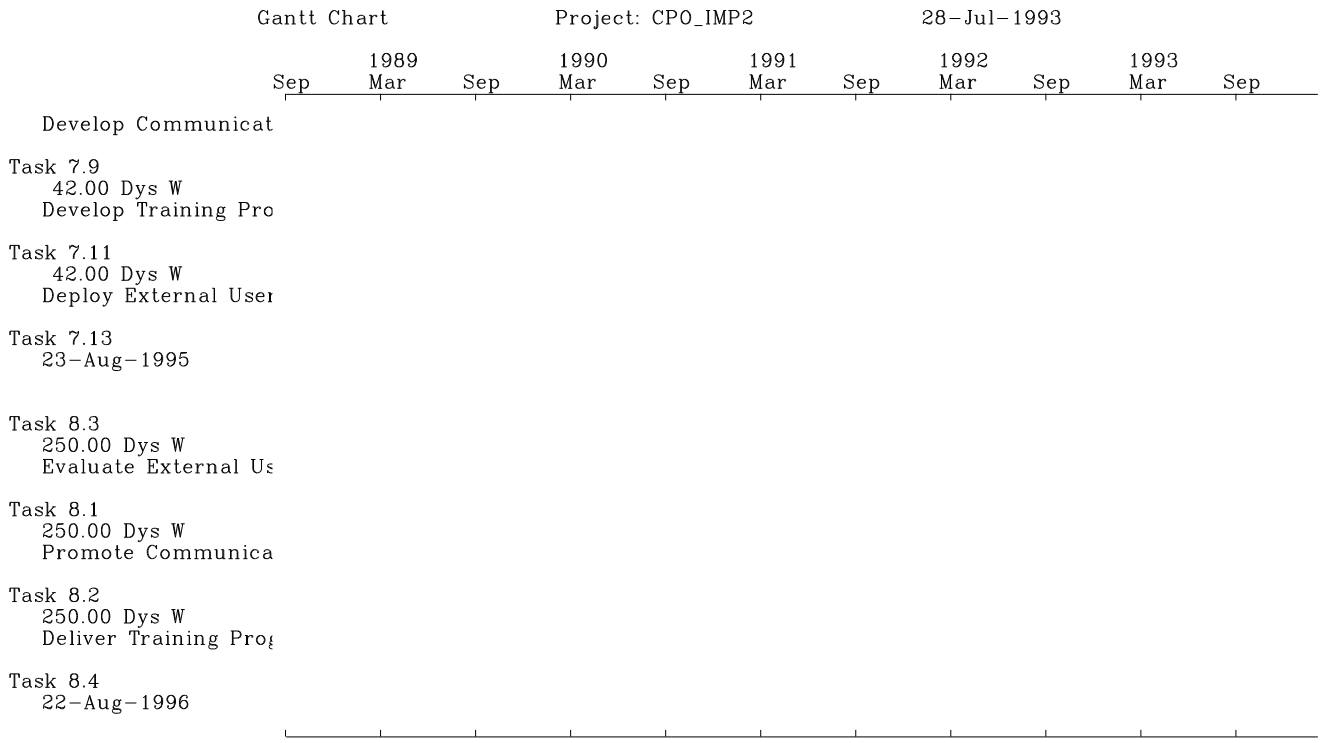
Page 7-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Complete Unit & Inte
Task 6.4
84.00 Dys W
Complete Unit & Inte
Task 6.7
84.00 Dys W
Develop Training Pro
Task 6.8
84.00 Dys W
Deploy Remote Work
Task 6.9
23-Aug-1994
Task 7.5
84.00 Dys W
Complete Unit & Inte
Task 7.1
42.00 Dys W
Deliver Training Pro
Task 7.2
42.00 Dys W

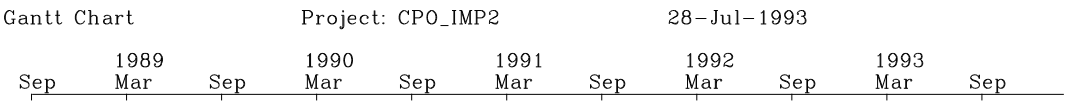
Page 8-1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Proceed with Techni
Task 7.3
42.00 Dys W
Proceed with Techni
Task 7.4
42.00 Dys W
Proceed with Techni
Task 7.6
42.00 Dys W
Test User Accounting
Task 7.12
125.00 Dys W
Deploy User Account
Task 7.7
42.00 Dys W
Complete Technical I
Task 7.10
42.00 Dys W
Expand Communicat
Task 7.8
42.00 Dys W

Page 9-1

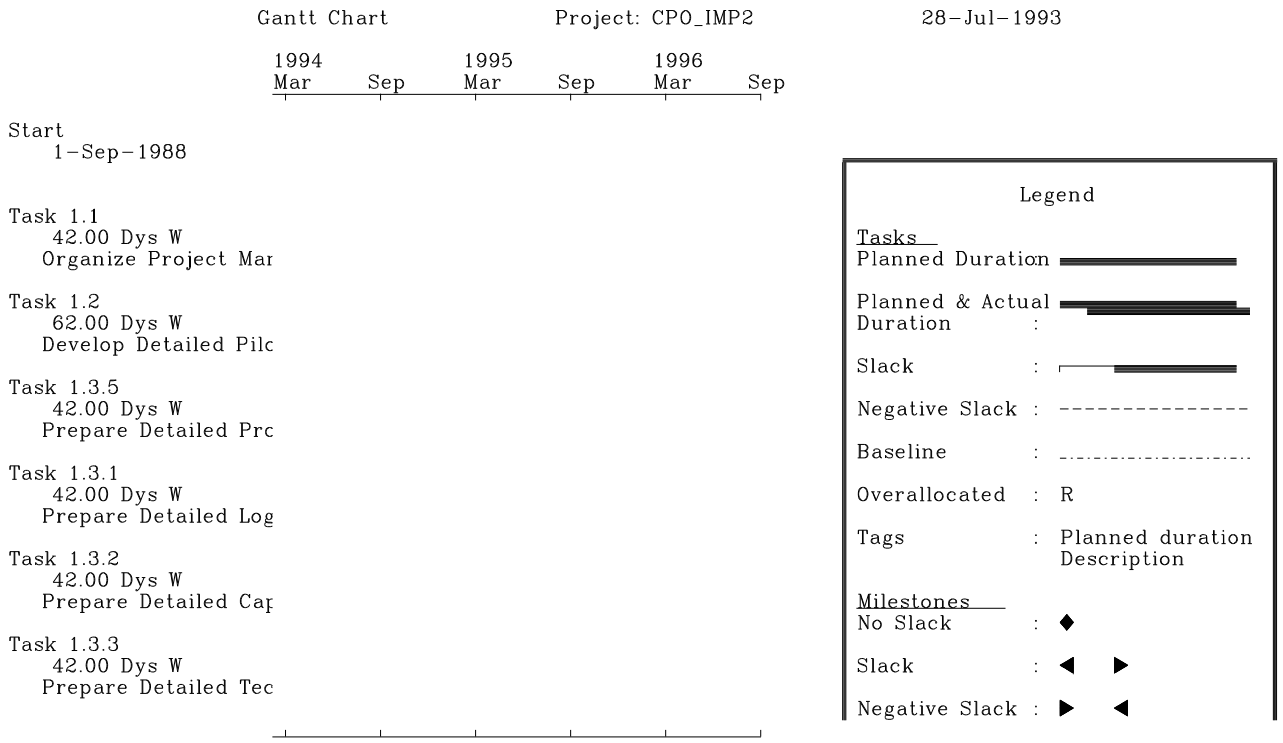

Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1989
1990
1991
1992
1993
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Mar
Sep
Develop Communicat
Task 7.9
42.00 Dys W
Develop Training Pro
Task 7.11
42.00 Dys W
Deploy External User
Task 7.13
23-Aug-1995
Task 8.3
250.00 Dys W
Evaluate External Us
Task 8.1
250.00 Dys W
Promote Communica
Task 8.2
250.00 Dys W
Deliver Training Pro
Task 8.4
22-Aug-1996

Page 10-1

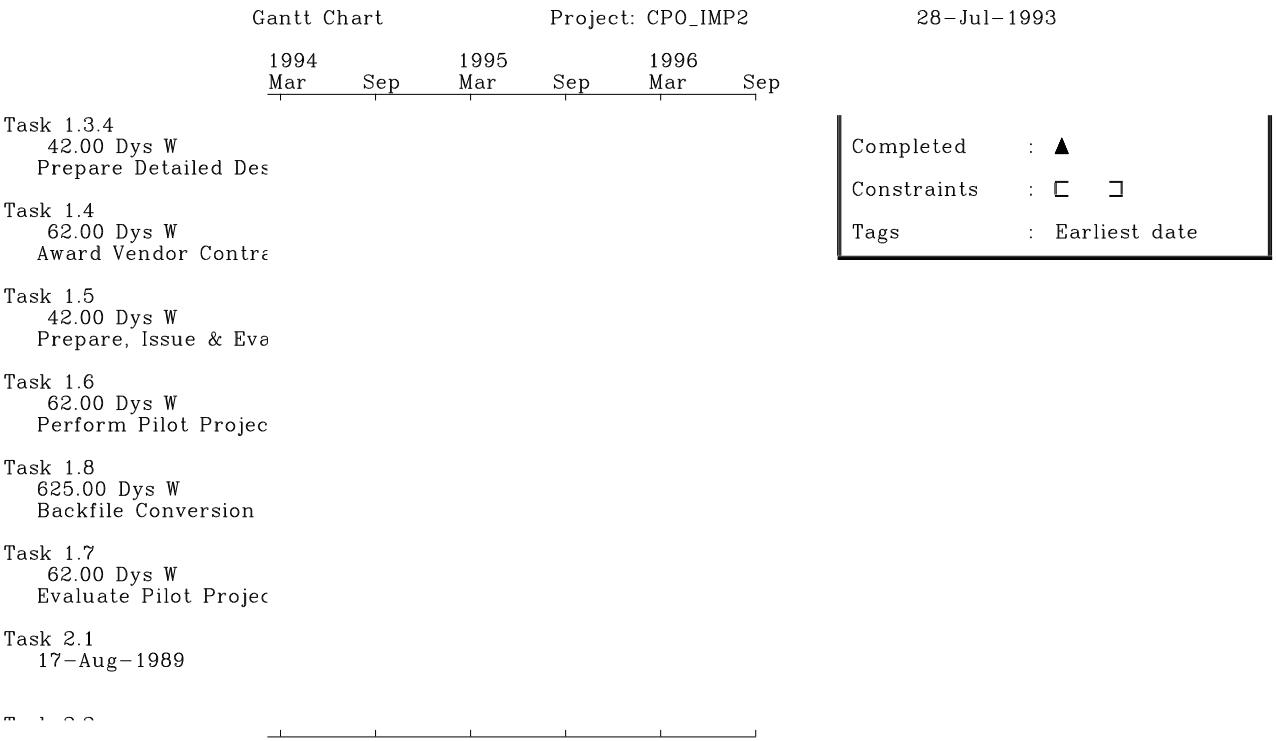


End
22-Aug-1996

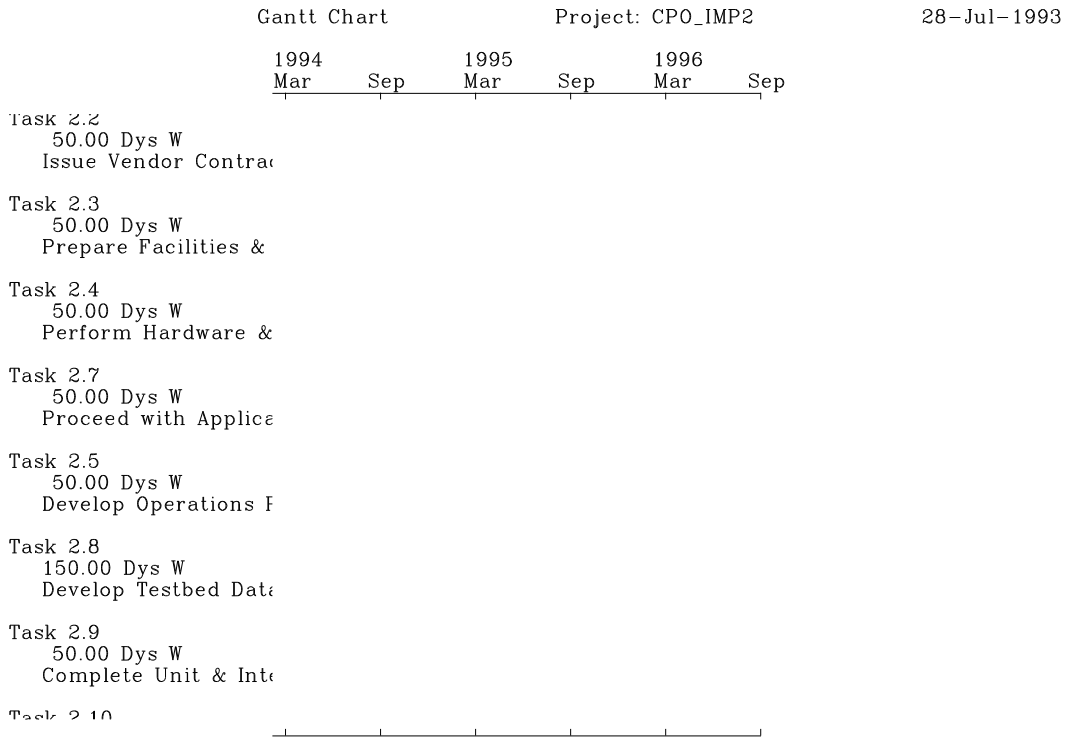
Page 11-1

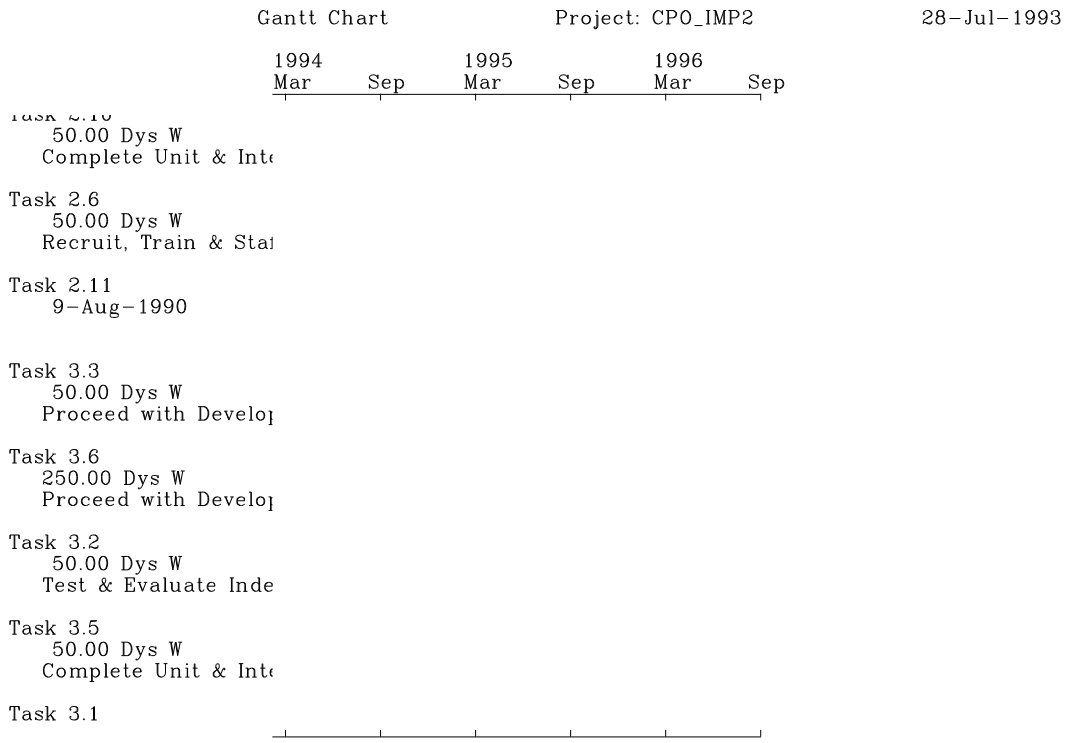

Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
Start
1-Sep-1988
Task 1.1
42.00 Dys W
Organize Project Mar
Task 1.2
62.00 Dys W
Develop Detailed Pilc
Task 1.3.5
42.00 Dys W
Prepare Detailed Prc
Task 1.3.1
42.00 Dys W
Prepare Detailed Log
Task 1.3.2
42.00 Dys W
Prepare Detailed Cap
Task 1.3.3
42.00 Dys W
Prepare Detailed Tec
Legend
Tasks
Planned Duration
Planned & Actual Duration :
Slack :
Negative Slack :
Baseline :
Overallocated : R
Tags : Planned duration
Description
Milestones
No Slack :
Slack :
Negative Slack :

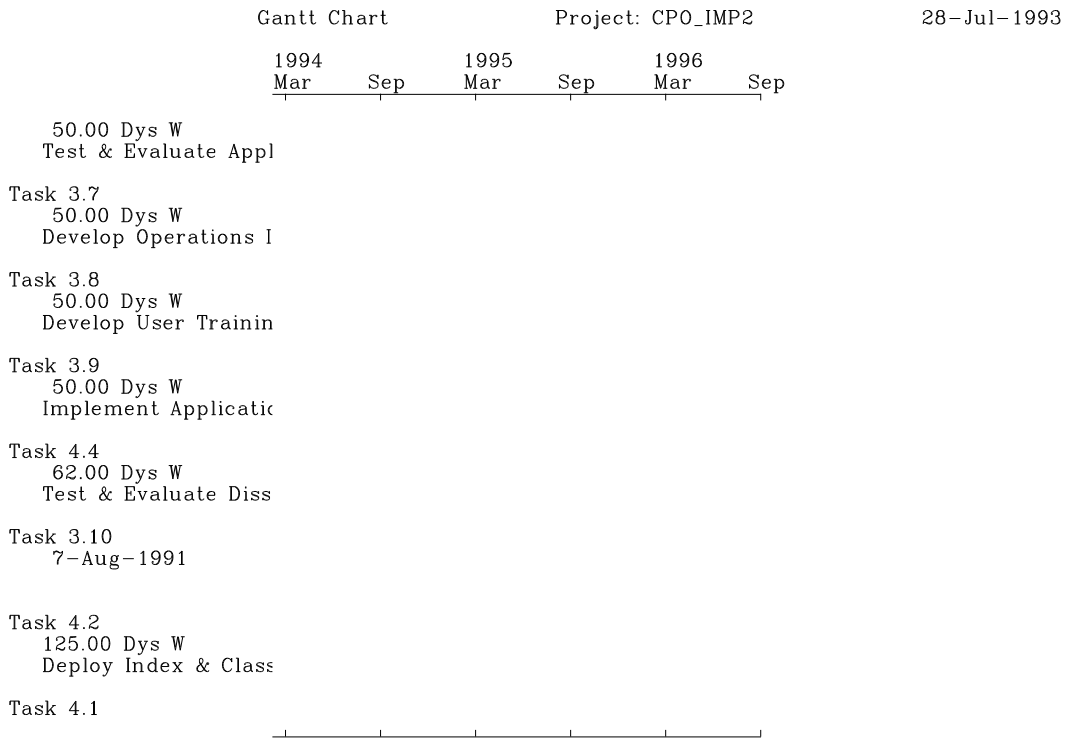
Page 1-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
Task 1.3.4
42.00 Dys W
Prepare Detailed Des
Task 1.4
62.00 Dys W
Award Vendor Contra
Task 1.5
42.00 Dys W
Prepare, Issue & Eva
Task 1.6
62.00 Dys W
Perform Pilot Projec
Task 1.8
625.00 Dys W
Backfile Conversion
Task 1.7
62.00 Dys W
Evaluate Pilot Projec
Task 2.1
17-Aug-1989
Completed : ▲
Constraints : ⊏ ⊐
Tags : Earliest date

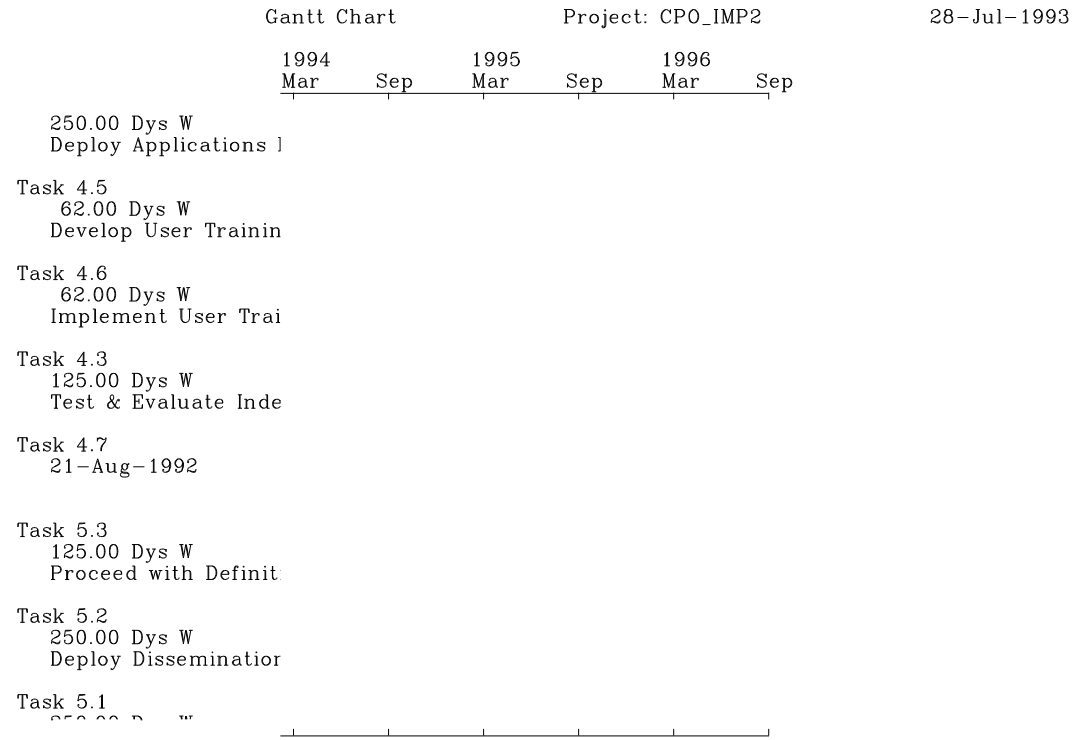
Page 2-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
Task 2.2
50.00 Dys W
Issue Vendor Contra
Task 2.3
50.00 Dys W
Prepare Facilities &
Task 2.4
50.00 Dys W
Perform Hardware &
Task 2.7
50.00 Dys W
Proceed with Applic
Task 2.5
50.00 Dys W
Develop Operations F
Task 2.8
150.00 Dys W
Develop Testbed Dat
Task 2.9
50.00 Dys W
Complete Unit & Int

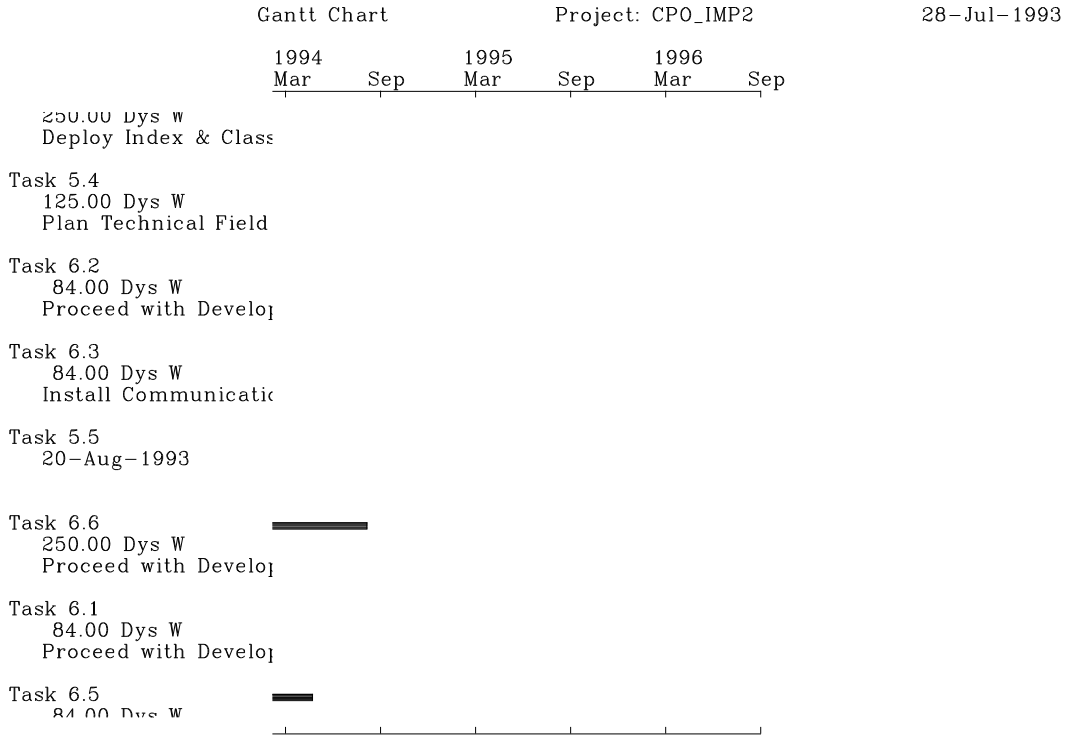

Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
50.00 Dys W
Complete Unit & Int
Task 2.6
50.00 Dys W
Recruit, Train & Sta
Task 2.11
9-Aug-1990
Task 3.3
50.00 Dys W
Proceed with Develo
Task 3.6
250.00 Dys W
Proceed with Develo
Task 3.2
50.00 Dys W
Test & Evaluate Inde
Task 3.5
50.00 Dys W
Complete Unit & Int
Task 3.1


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
50.00 Dys W
Test & Evaluate Appl
Task 3.7
50.00 Dys W
Develop Operations I
Task 3.8
50.00 Dys W
Develop User Trainin
Task 3.9
50.00 Dys W
Implement Applicatic
Task 4.4
62.00 Dys W
Test & Evaluate Diss
Task 3.10
7-Aug-1991
Task 4.2
125.00 Dys W
Deploy Index & Class
Task 4.1

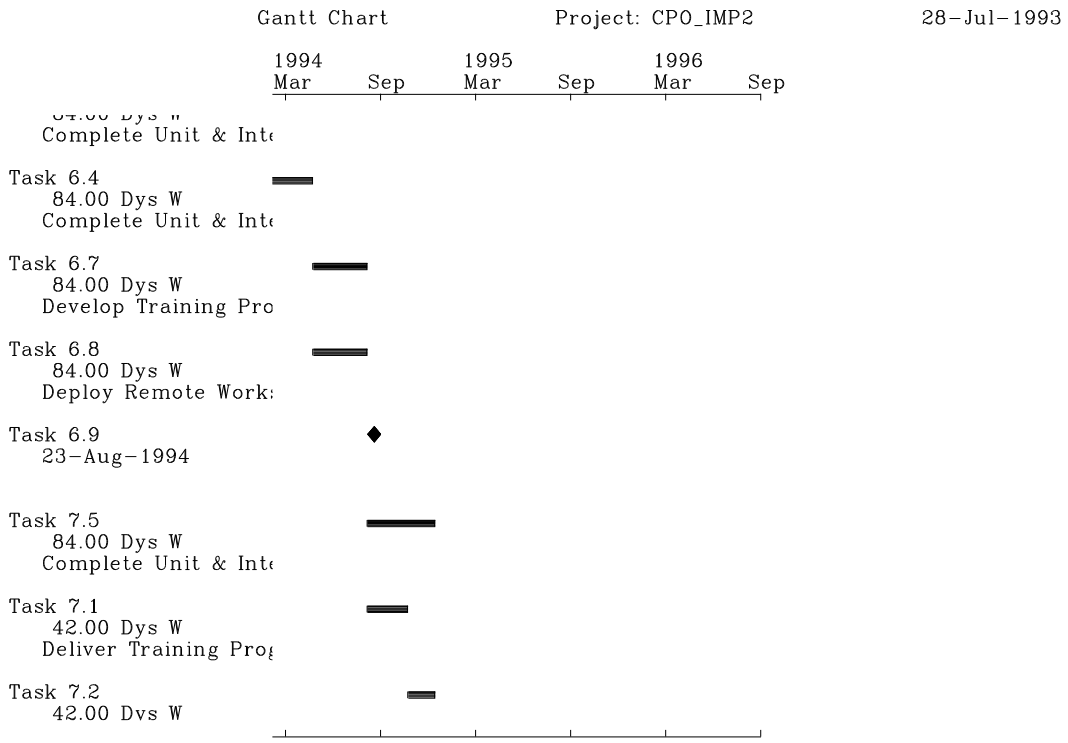
Page 5-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
250.00 Dys W
Deploy Applications
Task 4.5
62.00 Dys W
Develop User Trainin
Task 4.6
62.00 Dys W
Implement User Trai
Task 4.3
125.00 Dys W
Test & Evaluate Inde
Task 4.7
21-Aug-1992
Task 5.3
125.00 Dys W
Proceed with Definit
Task 5.2
250.00 Dys W
Deploy Disseminatior
Task 5.1

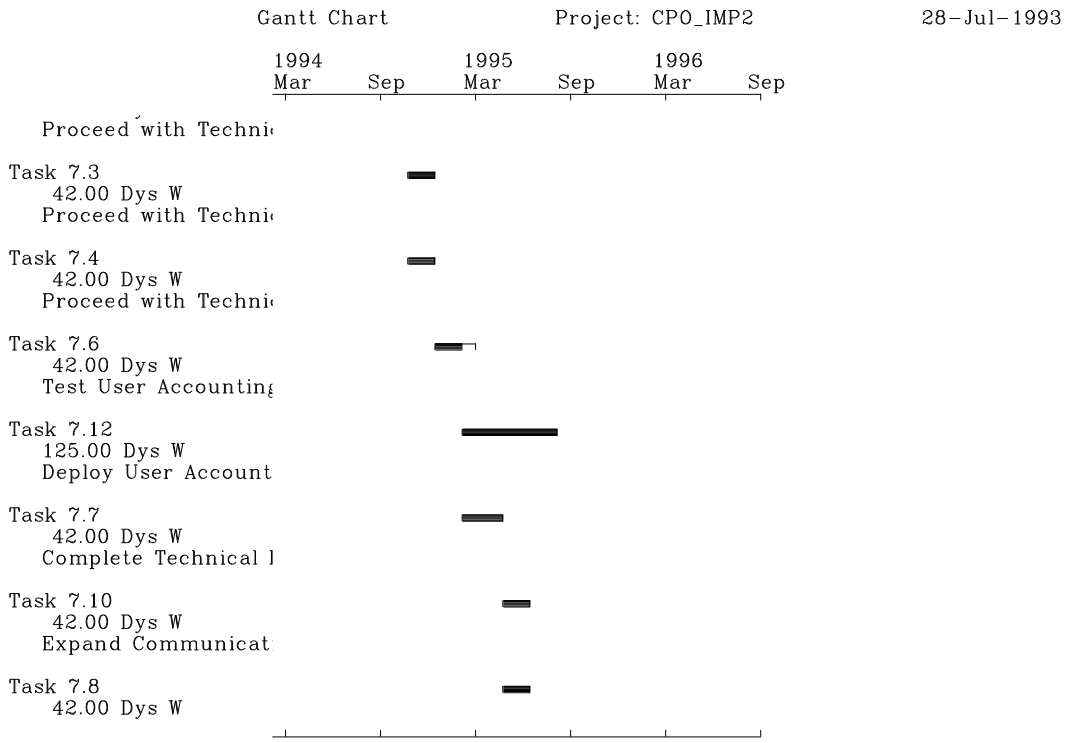
Page 6-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
250.00 Dys W
Deploy Index & Class
Task 5.4
125.00 Dys W
Plan Technical Field
Task 6.2
84.00 Dys W
Proceed with Develo
Task 6.3
84.00 Dys W
Install Communicatic
Task 5.5
20-Aug-1993
Task 6.6
250.00 Dys W
Proceed with Develo
Task 6.1
84.00 Dys W
Proceed with Develo
Task 6.5
84.00 Dys W

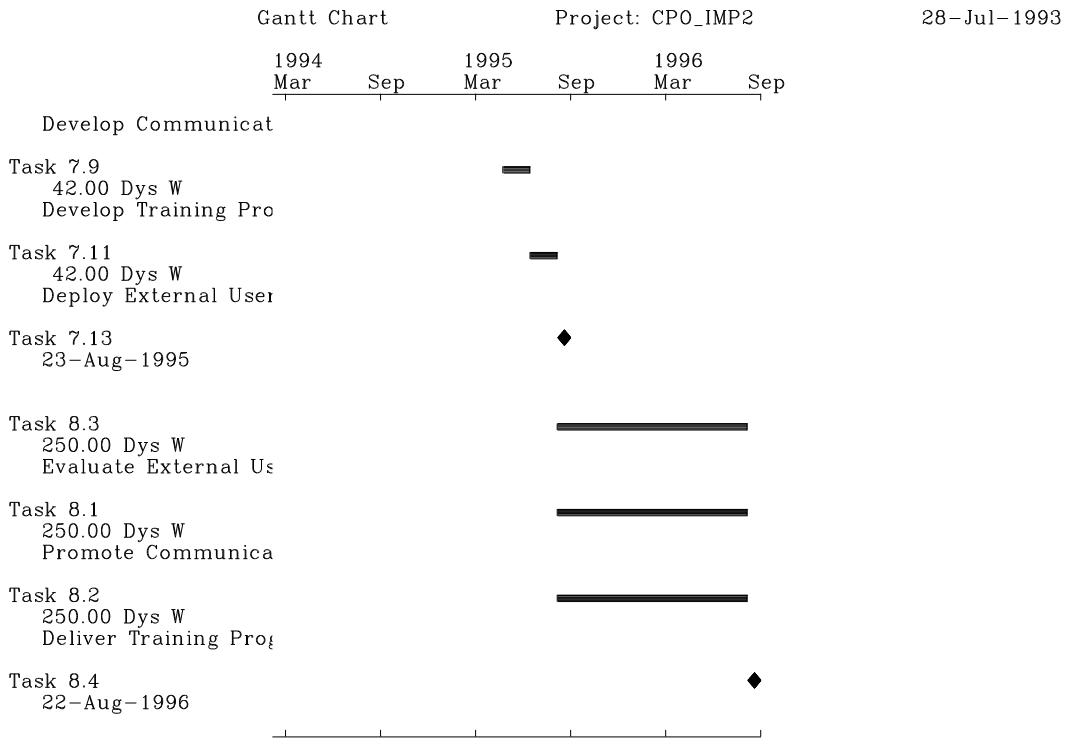
Page 7-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
Complete Unit & Inte
Task 6.4
84.00 Dys W
Complete Unit & Inte
Task 6.7
84.00 Dys W
Develop Training Pro
Task 6.8
84.00 Dys W
Deploy Remote Work
Task 6.9
23-Aug-1994
Task 7.5
84.00 Dys W
Complete Unit & Inte
Task 7.1
42.00 Dys W
Deliver Training Pro
Task 7.2
42.00 Dys W

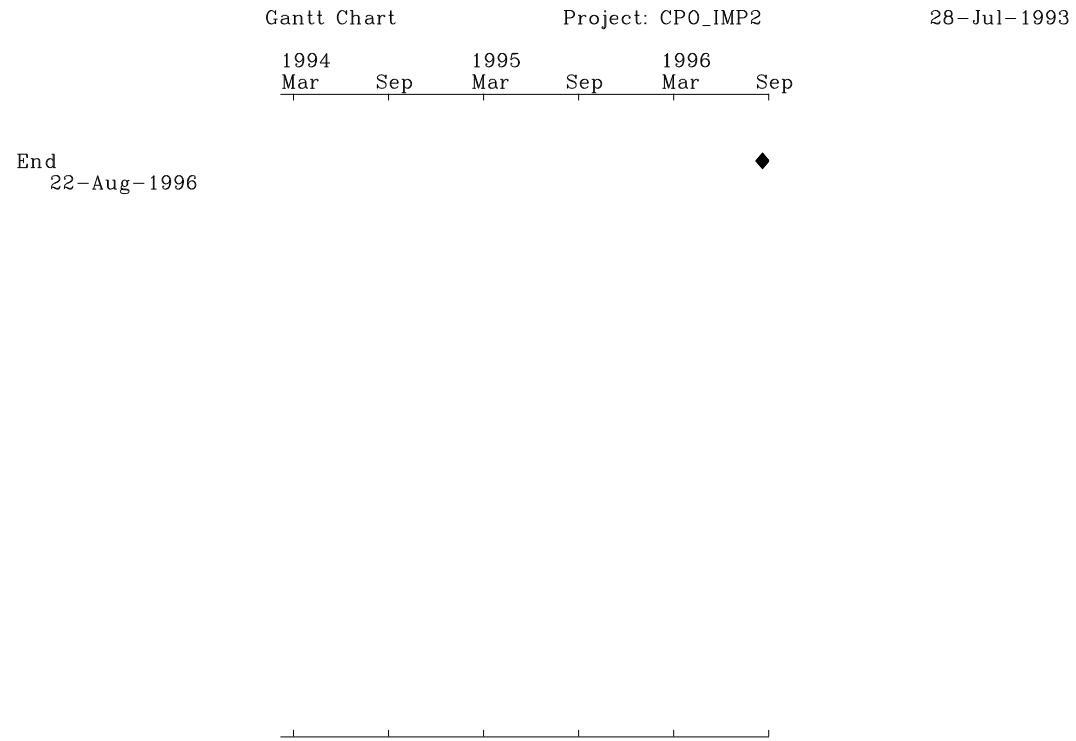
Page 8-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
Proceed with Techni
Task 7.3
42.00 Dys W
Proceed with Techni
Task 7.4
42.00 Dys W
Proceed with Techni
Task 7.6
42.00 Dys W
Test User Accounting
Task 7.12
125.00 Dys W
Deploy User Account
Task 7.7
42.00 Dys W
Complete Technical I
Task 7.10
42.00 Dys W
Expand Communicat
Task 7.8
42.00 Dys W

Page 9-2

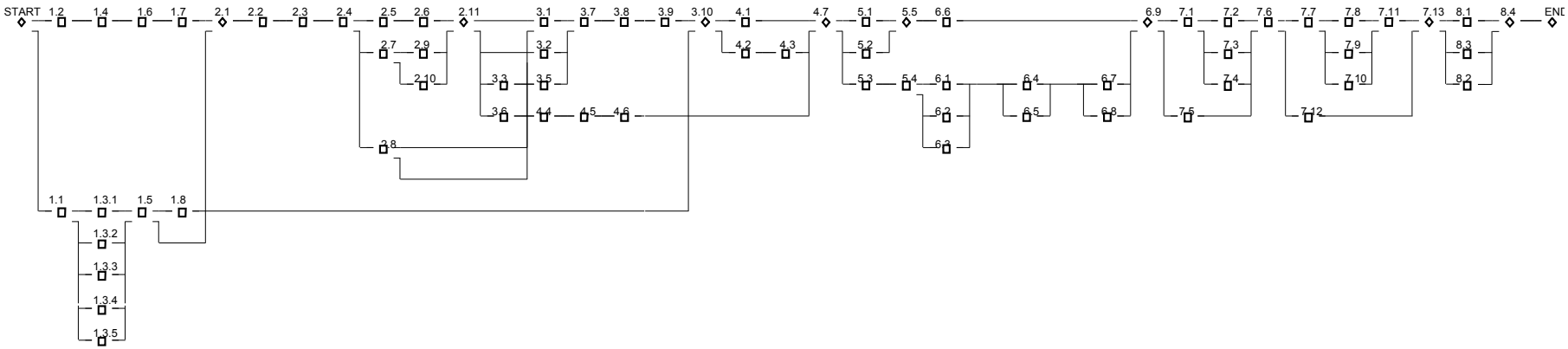

Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
Develop Communicat
Task 7.9
42.00 Dys W
Develop Training Pro
Task 7.11
42.00 Dys W
Deploy External User
Task 7.13
23-Aug-1995
Task 8.3
250.00 Dys W
Evaluate External Us
Task 8.1
250.00 Dys W
Promote Communica
Task 8.2
250.00 Dys W
Deliver Training Pro
Task 8.4
22-Aug-1996

Page 10-2


Gantt Chart
Project: CPO_IMP2
28-Jul-1993
1994
1995
1996
Mar
Sep
Mar
Sep
Mar
Sep
End
22-Aug-1996

Page 11-2

C.3. Pert Chart



C.4. Detailed Cost Estimates

C.4.1. Cost Estimates

	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94	1994-95	1995-96	Total
A. Capital Cost Components	1961	14001	9468	3980	3683	3248	2748	1823	40911
A.1. Project Management	885	1925	2200	2200	2150	2075	1400	1000	13835
A.1.1. Government Resource Costs	325	325	350	350	350	350	350	350	2750
A.1.1.1. Project Director	50	50	50	50	50	50	50	50	400
A.1.1.2. Executive Assistant	50	50	50	50	50	50	50	50	400
A.1.1.3. Executive Secretary	50	50	50	50	50	50	50	50	400
A.1.1.4. Asst. Proj. Dir. (Plng & Admin)	50	50	50	50	50	50	50	50	400
A.1.1.5. Asst. Proj. Dir. (Contract Adm.)	25	25	50	50	50	50	50	50	350
A.1.1.6. Asst. Proj. Dir. (User Srvcs)	50	50	50	50	50	50	50	50	400
A.1.1.7. Clerical Support	50	50	50	50	50	50	50	50	400
A.1.2. Contractor Resource Costs	525	1600	1850	1850	1800	1725	1050	650	11050
A.1.2.1. Proj. Mngr. (QA & Methods)	50	50	100	100	50	25	25	25	425
A.1.2.2. Proj. Mngr. (Recr. & Staff)	25	25	25	25	25	50	100	50	325
A.1.2.3. Proj. Mngr. (Training Adm.)	25	25	25	25	25	50	100	50	325
A.1.2.4. Proj. Mngr. (Fin. Accnting)	25	100	100	100	100	50	50	50	575
A.1.2.5. Admin Assistant	25	100	100	100	100	50	25	25	525
A.1.2.6. Proj. Mngr. (QC)	50	100	100	100	100	100	50	50	650
A.1.2.7. Proj. Mngr. (Fac. & Equip.)	50	100	100	100	100	100	50	50	650
A.1.2.8. Proj. Mngr. (Contr. Ops.)	50	100	100	100	100	100	50	50	650
A.1.2.9. Sen. Sys. An. (Inf. Stor. & Retr.)	25	50	100	100	100	100	50	25	550
A.1.2.10. Systems Analyst (2)	50	200	200	200	200	200	100	50	1200
A.1.2.11. Sen. Sys. An. (Reg. Systems)	25	100	100	100	100	100	50	25	600
A.1.2.12. Systems Analyst (2)	50	200	200	200	200	200	100	50	1200
A.1.2.13. Sen. Sys. An. (Diss. Systems)	0	50	100	100	100	100	50	25	525
A.1.2.14. Systems Analyst (2)	0	100	200	200	200	200	100	50	1050
A.1.2.15. Sen. Sys. An. (Mgmt. Inf. Sys.)	25	100	100	100	100	100	50	25	600
A.1.2.16. Systems Analyst (2)	50	200	200	200	200	200	100	50	1200
A.1.3. Equipment and Supplies	35	0	0	0	0	0	0	0	35
A.2. Hardware Costs	100	8665	3570	530	333	73	73	73	13415
A.2.1. CPU and Associated Hardware	0	8000	0	0	0	0	0	0	8000
A.2.2. On-line Storage Systems	0	500	1000	500	313	63	63	63	2500
A.2.3. Off-line Storage Systems	100	65	70	30	20	10	10	10	315
A.2.4. Workstations	0	0	2500	0	0	0	0	0	2500
A.2.5. Test Equipment	0	100	0	0	0	0	0	0	100
A.3. Systems Software	0	500	0	0	0	0	0	0	500
A.3.1. Operating System	0	200	0	0	0	0	0	0	200
A.3.2. System Utilities and Tools	0	50	0	0	0	0	0	0	50
A.3.3. Database Software	0	250	0	0	0	0	0	0	250
A.4. Application Software	43	255	1300	1050	1000	850	925	400	5823
A.4.1. Government Resources	38	100	100	100	100	100	100	100	738
A.4.1.1. Asst. Proj. Dir. (Systems Devlpmnt.)	38	50	50	50	50	50	50	50	388
A.4.1.2. Clerical Support	0	50	50	50	50	50	50	50	350
A.4.2. Contractor Resources	0	150	1200	950	900	750	825	300	5075
A.4.2.1. Sen. Design Specialist (DB)	0	0	100	100	100	100	100	50	550
A.4.2.2. DB Design Specialist (2)	0	0	200	200	200	200	200	100	1100
A.4.2.3. Sen. Design Specialist (Exam./Srch.)	0	0	100	100	100	50	25	0	375
A.4.2.4. Appl. Software Specialist (2)	0	0	200	200	200	100	50	0	750
A.4.2.5. Sen. Design Specialist (Appl. Proc.)	0	50	100	50	50	25	25	0	300
A.4.2.6. Appl. Software Specialist (2)	0	100	200	100	100	50	50	0	600
A.4.2.7. Sen. Design Specialist (Inf. Diss.)	0	0	100	100	50	25	25	0	300
A.4.2.8. Appl. Software Specialist (2)	0	0	200	100	100	50	50	0	500
A.4.2.9. Sen. Design Specialist (Mgmt. Inf.)	0	0	0	0	0	50	100	50	200
A.4.2.10. Appl. Software Specialist (2)	0	0	0	0	0	100	200	100	400
A.4.3. Development Tools and Supplies	5	5	0	0	0	0	0	0	10
A.4.4. Software Purchase	0	0	0	0	0	0	0	0	0
A.5. Implementation Costs	0	135	50	50	50	100	200	200	785
A.5.1. Delivery, Installation and Insurance	0	85	0	0	0	0	0	0	85
A.5.2. Documentation and Training Courses	0	50	50	50	50	100	200	200	700
A.5.2.1. Contracted Trainer/Developer (2)	0	50	50	50	50	100	200	200	700

A.6. Facilities Costs	0	654	0	0	0	0	0	0	654
A.6.1. Space Conditioning (5,000 sq. ft.)	0	184	0	0	0	0	0	0	184
A.6.1.1. Design & Engineering	0	15	0	0	0	0	0	0	15
A.6.1.2. Raised Floor	0	70	0	0	0	0	0	0	70
A.6.1.3. Suspended Ceiling	0	10	0	0	0	0	0	0	10
A.6.1.4. Paint	0	5	0	0	0	0	0	0	5
A.6.1.5. Operator Workstation	0	40	0	0	0	0	0	0	40
A.6.1.6. Office Partition	0	3	0	0	0	0	0	0	3
A.6.1.7. Security System	0	1	0	0	0	0	0	0	1
A.6.1.8. Heating & Air-conditioning	0	30	0	0	0	0	0	0	30
A.6.1.9. Sprinkling/Halon	0	5	0	0	0	0	0	0	5
A.6.1.10.Lighting	0	5	0	0	0	0	0	0	5
A.6.2. Air Conditioning - User Environment	0	0	0	0	0	0	0	0	0
A.6.3. Furniture	0	400	0	0	0	0	0	0	400
A.6.3.1. Workstation Desks (200 pcs)	0	200	0	0	0	0	0	0	200
A.6.3.2. Ergonomic Chairs (200 pcs)	0	100	0	0	0	0	0	0	100
A.6.3.3. Filing Cabinets (200 pcs)	0	100	0	0	0	0	0	0	100
A.6.4. Electrical - User Environment	0	70	0	0	0	0	0	0	70
A.7. Backfile Data Conversion	934	1868	1868	0	0	0	0	0	4669
A.7.1. Scanning/OCR-ing	417	833	833	0	0	0	0	0	2083
A.7.2. Document preparation/re-assembly	200	400	400	0	0	0	0	0	1000
A.7.3. Manual Indexing	75	150	150	0	0	0	0	0	375
A.7.4. Equipment Rental	242	484	484	0	0	0	0	0	1211
A.8. External Database Access Setup	0	0	130	0	0	0	0	0	130
A.8.1. Software Licence	0	0	130	0	0	0	0	0	130
A.9. Marketing and Promotion	0	0	0	150	150	150	150	150	750
A.9.1. Remote Access Services	0	0	0	150	0	0	0	0	150
A.9.2. Public Search Room Services	0	0	0	0	150	0	0	0	150
A.9.3. Technical Information Search Program	0	0	0	0	0	150	0	0	150
A.9.4. Public Education Program	0	0	0	0	0	0	150	0	150
A.9.5. Publication Services	0	0	0	0	0	0	0	150	150
A.10. File Translation	0	0	350	0	0	0	0	0	350
A.10.1. Tape-related Costs	0	0	300	0	0	0	0	0	300
A.10.2. Contracted Resources	0	0	50	0	0	0	0	0	50
B. On-going Cost Elements	336	1990	3069	3524	3555	3575	3835	4134	24019
B.1. Systems Management	90	165	300	300	300	300	300	300	2055
B.1.1. Government Resources	75	150	300	300	300	300	300	300	2025
B.1.1.1. Asst. Proj. Dir. (Sys. Mngmt.)	25	25	50	50	50	50	50	50	350
B.1.1.2. Systems Mngr. (Sys. Plng.)	25	25	50	50	50	50	50	50	350
B.1.1.3. Sys. Specialist (Cap & Perf. Meas.)	0	25	50	50	50	50	50	50	325
B.1.1.4. Systems Mngr. (Security)	25	25	50	50	50	50	50	50	350
B.1.1.5. Sys. Specialist (DB & Network Sec.)	0	25	50	50	50	50	50	50	325
B.1.1.6. Clerical Support	0	25	50	50	50	50	50	50	325
B.1.2. Contractor Resources	0	0	0	0	0	0	0	0	0
B.1.3. Equipment and Supplies	15	15	0	0	0	0	0	0	30
B.2. Computer Operations	163	680	1200	1200	1200	1200	1200	1200	8043
B.2.1. Government Resources	138	225	300	300	300	300	300	300	2163
B.2.1.1. Asst. Proj. Dir. (Sys & Ops. Support)	25	50	50	50	50	50	50	50	375
B.2.1.2. Systems Mngr. (DB Administration)	25	25	50	50	50	50	50	50	350
B.2.1.3. DB Systems Specialist	13	25	50	50	50	50	50	50	338
B.2.1.4. Systems Mngr. (Systems Software)	25	50	50	50	50	50	50	50	375
B.2.1.5. Systems Software Programmer	0	25	50	50	50	50	50	50	325
B.2.1.6. Clerical Support	50	50	50	50	50	50	50	50	400
B.2.2. Contractor Resources	0	450	900	900	900	900	900	900	5850
B.2.2.1. Systems Mngr. (Computer Operations)	0	50	100	100	100	100	100	100	650
B.2.2.2. Computer Operator (4)	0	200	400	400	400	400	400	400	2600
B.2.2.3. Data Entry Operator (2)	0	100	200	200	200	200	200	200	1300
B.2.2.4. Systems Mngr. (Data Communications)	0	50	100	100	100	100	100	100	650
B.2.2.5. Data Communications Technician	0	50	100	100	100	100	100	100	650
B.2.3. Equipment and Supplies	25	5	0	0	0	0	0	0	30
B.3. Hardware Maintenance	0	6	526	740	772	792	796	801	4433
B.3.1. CPU and Associated Hardware	0	0	480	480	480	480	480	480	2880
B.3.2. On-line Storage Systems	0	0	30	90	120	139	143	146	668
B.3.3. Off-line Storage Systems	0	6	10	14	16	17	18	18	99
B.3.4. Workstations	0	0	0	150	150	150	150	150	750
B.3.5. Test Equipment	0	0	6	6	6	6	6	6	36
B.4. Communications	0	0	0	40	40	40	295	590	1005
B.4.1. CPO X.25 Circuits	0	0	0	10	10	10	70	140	240
B.4.2. X.25 Circuit Usage (External Users)	0	0	0	30	30	30	225	450	765

B.5. Software Licences	0	0	25	25	25	25	25	25	150
B.5.1. Systems Software Licences	0	0	25	25	25	25	25	25	150
B.5.2. Applications Software Licenses	0	0	0	0	0	0	0	0	0
B.6. External Database Usage									0
B.7. Application Software Maintenance	25	63	100	200	200	200	200	200	1188
B.7.1. Government Resources	25	63	100	200	200	200	200	200	1188
B.7.1.1. Sen. Systems Analyst (SW Maint.)	25	25	25	50	50	50	50	50	325
B.7.1.2. SW Specialist (DB Maintenance)	0	13	25	50	50	50	50	50	288
B.7.1.3. SW Specialist (Exm/Srch & Ap. Proc)	0	13	25	50	50	50	50	50	288
B.7.1.4. SW Specialist (Diss. & Mngmt. Inf.)	0	13	25	50	50	50	50	50	288
B.8. Training and User Support	0	38	50	150	150	150	150	150	838
B.8.1. Government Resources	0	38	50	150	150	150	150	150	838
B.8.1.1. Systems Mngr. (Training & Support)	0	13	25	50	50	50	50	50	288
B.8.1.2. Instructor (2)	0	25	25	100	100	100	100	100	550
B.9. Facilities Maintenance	59	237	257	257	257	257	257	257	1837
B.9.1. Lease/Rental	59	215	235	235	235	235	235	235	1682
B.9.1.1. Computer Room (5,000 sq. ft.)	0	98	98	98	98	98	98	98	685
B.9.1.2. Project Staff (6,000 sq. ft.)	59	117	117	117	117	117	117	117	880
B.9.1.3. Training Space (1,000 sq. ft.)	0	0	20	20	20	20	20	20	117
B.9.2. Maintenance	0	22	22	22	22	22	22	22	154
B.9.2.1. Design & Engineering	0	2	2	2	2	2	2	2	13
B.9.2.2. Raised Floor	0	8	8	8	8	8	8	8	59
B.9.2.3. Suspended Ceiling	0	1	1	1	1	1	1	1	8
B.9.2.4. Paint	0	1	1	1	1	1	1	1	4
B.9.2.5. Operator Workstation	0	5	5	5	5	5	5	5	34
B.9.2.6. Office Partition	0	0	0	0	0	0	0	0	3
B.9.2.7. Security System	0	0	0	0	0	0	0	0	1
B.9.2.8. Heating & Air-conditioning	0	4	4	4	4	4	4	4	25
B.9.2.9. Sprinkling/Halon	0	1	1	1	1	1	1	1	4
B.9.2.10.Lighting	0	1	1	1	1	1	1	1	4
B.10. On-going Data Capture	0	702	512	512	512	512	512	512	3773
B.10.1. Equipment Purchase	0	460	0	0	0	0	0	0	460
B.10.2. Equipment Maintenance	0	0	28	28	28	28	28	28	166
B.10.3. Scanning/OCR-ing	0	146	292	292	292	292	292	292	1896
B.10.4. Document preparation/re-assembly	0	70	140	140	140	140	140	140	910
B.10.5. Manual Indexing	0	26	53	53	53	53	53	53	341
B.11. Publishing	0	0	0	0	0	0	0	0	0
B.12. Translation	0	100	100	100	100	100	100	100	700
Total Project Costs	2298	15992	12537	7504	7238	6823	6582	5957	64930
A. Capital Cost Components	1961	14001	9468	3980	3683	3248	2748	1823	40911
A.1. Project Management	885	1925	2200	2200	2150	2075	1400	1000	13835
A.2. Hardware Costs	100	8665	3570	530	333	73	73	73	13415
A.3. Systems Software	0	500	0	0	0	0	0	0	500
A.4. Application Software	43	255	1300	1050	1000	850	925	400	5823
A.5. Implementation Costs	0	135	50	50	50	100	200	200	785
A.6. Facilities Costs	0	654	0	0	0	0	0	0	654
A.7. Backfile Data Conversion	934	1868	1868	0	0	0	0	0	4669
A.8. External Database Access Setup	0	0	130	0	0	0	0	0	130
A.9. Marketing and Promotion	0	0	0	150	150	150	150	150	750
A.10. File Translation	0	0	350	0	0	0	0	0	350
B. On-going Cost Elements	336	1990	3069	3524	3555	3575	3835	4134	24019
B.1. Systems Management	90	165	300	300	300	300	300	300	2055
B.2. Computer Operations	163	680	1200	1200	1200	1200	1200	1200	8043
B.3. Hardware Maintenance	0	6	526	740	772	792	796	801	4433
B.4. Communications	0	0	0	40	40	40	295	590	1005
B.5. Software Licences	0	0	25	25	25	25	25	25	150
B.6. External Database Usage	0	0	0	0	0	0	0	0	0
B.7. Application Software Maintenance	25	63	100	200	200	200	200	200	1188
B.8. Training and User Support	0	38	50	150	150	150	150	150	838
B.9. Facilities Maintenance	59	237	257	257	257	257	257	257	1837
B.10. On-going Data Capture	0	702	512	512	512	512	512	512	3773
B.11. Publishing	0	0	0	0	0	0	0	0	0
B.12. Translation	0	100	100	100	100	100	100	100	700
	2298	15992	12537	7504	7238	6823	6582	5957	64930
Equipment and Maintenance	422	9640	4608	1298	1132	892	896	901	19789
Backfile Conversion	692	1383	1383	0	0	0	0	0	3458
On-going Data Capture	0	242	484	484	484	484	484	484	3147
Software	150	1350	2155	1775	1725	1575	1250	525	10505
Facilities and Training	59	941	307	307	307	357	457	457	3190
Operations and Communications	0	450	900	940	940	940	1195	1490	6855
Project Consulting Support	300	600	650	650	600	525	450	350	4125
Salaries	600	900	1200	1400	1400	1400	1400	1400	9700
Operations Management	75	300	400	400	400	400	200	100	2275
Other	0	185	450	250	250	250	250	250	1885
Control Total	2298	15992	12537	7504	7238	6823	6582	5957	64930

Salaries	600	900	1200	1400	1400	1400	1400	1400	9700
Capital	180	9844	3595	555	358	98	98	98	14824
O&M	1518	5248	7742	5549	5480	5325	5085	4459	40406
	2298	15992	12537	7504	7238	6823	6582	5957	64930
Number of ASB positions:	15	23	23	24	24	24	24	24	
Number of clerical positions:	2	4	4	4	4	4	4	4	
Number of ASB person-years:	10.00	14.50	20.00	24.00	24.00	24.00	24.00	24.00	
Number of clerical person-years:	2.00	3.50	4.00	4.00	4.00	4.00	4.00	4.00	

Definitions:

Equipment and Maintenance	A.1.3.+A.2.+A.4.3.+A.7.4.+B.1.3.+B.2.3.+B.3.+B.10.1.+B.10.2
Backfile Conversion	A.7.1.+A.7.2.+A.7.3.
On-going Data Capture	B.10.3.+B.10.4.+B.10.5.
Software	A.1.2.10.+A.1.2.12.+A.1.2.14.+A.1.2.16.+A.3.+A.4.2.+A.4.4.+A.8.1.+B.5.
Facilities and Training	A.5.2.+A.6.+B.9.
Operations and Communications	B.2.2.+B.4.
Project Consulting Support	A.1.2.1.+A.1.2.2.+A.1.2.3.+A.1.2.4.+A.1.2.5.+A.1.2.6.+A.1.2.7.+A.1.2.8.
Salaries	A.1.1.+A.4.1.+B.1.1.+B.2.1.+B.7.1.+B.8.1.
Operations Management	A.1.2.9.+A.1.2.11.+A.1.2.13.+A.1.2.15.+B.1.2.
Other	A.5.1.+A.9.+A.10.+B.6.+B.11.+B.12.

C.4.2. Summary of Costs

C.4.2.1. Capital/Project Costs

C.4.2.1.1. Project Management Costs

- Description

  These costs encompass all costs associated with the overall project management for the implementation of the CPO automation project. Specifically, they cover

  a. compensation costs for government staff required to plan, oversee and monitor the implementation effort,

  b. compensation costs for contracted staff required to plan, oversee and monitor the implementation effort, and

  c. the cost of project management supplies and tools such as microcomputers and project management software.

- Costing Assumptions and Methods

  The reference for these costs is based on the method described below using the following rates supplied by the CPO management:

  a. Average compensation for government resources were based on an overall rate of $50,000 per man-year. This figure was assumed to include gross salary and 15% of gross salary for benefits. Allocations for overhead were not included.

b. An average fee of $100,000 per man-year was used to estimate the costs for contracted resources.

The project management staff costs to implement the project have been derived by first developing the schedule of major activities for each year of the project. This schedule is contained in Appendix C.1.-C.3. A generic set of postitions for the organization of the team required to implement the project was also defined. The project team organization is contained in Appendix C.5. Estimates in person-years of support required from each position were then computed using the schedule of major activities to be performed.

c. An amount of $5,000 for each resource position defined in the project team was used to estimate the costs for microcomputers, project management software and miscellaneous supplies. These costs were applied to each position in the year in which man-months for the position first appeared appeared in the schedule.

- Cost Summary of Project Management Costs

	Cost Component	Estimated Amount
a.	Government Resource Costs	$2,750,000
b.	Contracted resource Costs	$11,050,000
c.	Equipment and Supplies	$35,000
	Total Project Management Costs	$13,835,000

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Government Resource Costs	A.1.1.
b.	Contracted Resource Costs	A.1.2.
c.	Project Mgmt Tools and Supplies	A.1.3.

C.4.2.1.2. Hardware Costs

- Description

These costs encompass all hardware acquisition costs. It includes the cost of

a. CPU and associated hardware

b. on-line storage for operational and user data

c. off-line storage for back-up and archive data

d. workstations

e. test equipment

- Costing Assumptions and Methods

The cost estimates are based on quotations obtained from vendors by RES, the CPO, and other consulting teams. Test equipment costs were based on our own experience in preparing network control facilities cost budgets.

CPU, associated hardware, workstations, peripheral equipment and test equipment are installed as soon as possible to meet requirements for data conversion, software development, testbed development and testing.

On-line mass storage equipment is installed incrementally according to the rate of development of the databases.

Off-line mass storage media is purchased according to the growth of archive and back-up databases. Materials, such as magnetic tape required for archive or back-up, are included as part of on-going operations expenses.

- Summary of Hardware Costs

	Cost Component	Estimated Amount
a.	CPU and associated hardware	$8,000,000
b.	On-line storage	$2,500,000
c.	Off-line storage	$315,000
d.	Workstations	$2,500,000
e.	Test equipment	$100,000
	Total Hardware Costs	$13,415,000

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	CPU and associated hardware	A.2.1.
b.	On-line storage	A.2.2.
c.	Off-line storage	A.2.3.
d.	Workstations	A.2.4.
e.	Test equipment	A.2.5.

C.4.2.1.3. Systems Software

- Description

Systems software costs encompass the acquisition costs of the following components.

a. The computer operating system

b. System software utilities and software development tools

c. Database management system software

- Costing Assumptions and Methods

The cost estimates are based on quotations obtained from vendors by RES, the CPO, and other consulting teams.

These costs cover initial purchase of a license to use the said software. Annual license fees/charges where applicable are accounted for in component B.5. Installation charges are included in cost component A.5.

- Summary of Systems Software Costs

	Cost Component	Estimated Amount
a.	The computer operating system	$200,000
b.	Systems utilities and tools	$50,000
c.	DBMS software	$250,000
	Total Systems Software Costs	$500,000

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	The computer operating system	A.3.1.
b.	Systems utilities and tools	A.3.2.
c.	DBMS software	A.3.3.

C.4.2.1.4. Applications Software Development/Acquisition

- Description

The costs associated with the development and testing of the applications software for the CPO system. These costs would include the costs of acquiring off-the-shelf applications software and adapting it to the local environment.

- Costing Assumptions and Methods

As no fully integrated system for the development of a fully integrated patent processing system is known, the worst case assumption of development of the application software from scratch was used. A large number of elements are known to exist, such as fourth generation

languages, document management software, full-text search tools etc. These tools may help to decrease the costs. Several of these tools have been researched and their costs range in the order of $60,000 to $100,000.

The software development exercise was assumed to be contracted out and the rate of $100,000 per man-year was used as an average estimate for each contracted resource required.

The applications systems development costs were derived by first developing the schedule of major activites for the project, defining the generic positions required for the systems development function and loading these positions with estimates of the support required in man-years from each position given the schedule.

- Summary of Applications Software Costs

	Cost Component	Estimated Amount
a.	Government Resource Costs	$737,500
b.	Contracted Resource Costs	$5,075,000
c.	Development Tools, Supplies	$10,000
	Total Applications Software Costs	$5,822,500

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Government Resource Costs	A.4.1.
b.	Contracted Resource Costs	A.4.2.
c.	Software Development Tools	A.4.3.

C.4.2.1.5. Implementation Costs

- Description

The costs associated with the installation of equipment Included in these costs are all costs associated with development of user documentation and training programs. These costs are sub-divided into components as follows:

a. Delivery installation and insurance of equipment

b. Development of applications software documentation and development of training courses and materials

- Costing Assumptions and Methods

Costs estimates for delivery, installation and insurance are based on estimates developed by Task 5. Development of documentation and training costs are based on the resource costs of a course development specialist for 2.5 years.

- Summary of Implementation Costs

	Cost Component	Estimated Amount
a.	Delivery, installation and insurance	$85,000
b.	Documentation and training courses	$700,000
	Total Installation Costs	$785,000

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Delivery, installation and insurance	A.5.1.
b.	Documentation and training courses	A.5.2.

C.4.2.1.6. Facilities Costs

- Description

  The costs associated with modifying existing facilities and setting up new facilities in order to accommodate staff and equipment for the CPO system.

  The specific cost elements included for consideration are

  a. Space Conditioning Costs

  b. Air Conditioning (User Environment)

  c. Furniture

  d. Electrical (in User Environment)

- Costing Assumptions and Methods

  Square footage requirements are based on an estimated allocation of 5000 square feet for a computer room including 1600 square feet for storage space and 400 square feet for operations staff offices, 6000 square feet for project staff and 1000 square feet for a training centre. The square footage allocations for computer facilities assume that an uninterruptible power supply is not a requirement. These estimates were compared with the actual space allocations in existing sites and found to be compatible. An estimate provided by one vendor estimates 1800 square feet are required for machinery.

  The costs provided here are based on earlier estimates developed in the General Requirements study with the following revisions.

a. Space Conditioning Costs

Space conditioning costs for computer operations were researched independently by RES by interviewing a local contractor specializing in the business of computer room construction. The contractor was provided with the estimated floor space requirements and developed from these the space conditioning costs for the computer room. These costs are comparable to an estimate provided by one vendor. They total $184,000.

No space conditioning requirements for the space occupied by the project staff are foreseen.

b. Air Conditioning (User Environment)

Upgrading of the air-conditioning for the user environment is not considered necessary. The heat dissipation of the workstations required will not be above that given off by standard microcomputers.

c. Furniture

Ergonomic furniture requirements are based on catalogue prices for the necessary furniture units. One hundred application processing workstations and one hundred examiner workstations amount to approximately $300,000.

Miscellaneous elements (e.g. filing cabinets) amount to an additional $90,000. Furniture for the training centre (5 workstations) amount to $10,000, resulting in a total of $400,000 required for office furniture (volume discounts not included).

d. Electrical (in User Environment)

The estimate in the General Requirements study for this cost to be $70,000.

- Summary of Facilities Costs

	Cost Component	Estimated Amount
a.	Space Conditioning Costs	$184,000
b.	Air Conditioning (User Environment)	N/A
c.	Furniture	$400,000
d.	Electrical (in User Environment)	$70,000
	Total Space Conditioning Costs	$654,000

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Space Conditioning Costs	A.6.1.
b.	Air Conditioning (User Environment)	A.6.2.
c.	Furniture	A.6.3.
d.	Electrical (in User Environment)	A.6.4.

C.4.2.1.7. Backfile Data Conversion Costs

- Description

  Costs assiciated with backfile data conversion activities, in particular:

  a. Scanning/OCR-ing

  b. Document preparation/re-assembly

  c. Manual indexing

  d. Equipment rental

- Costing Assumptions and Methods

The cost estimates are based on studies conducted Pilot Backfile Conversion study team. In particular, it was assumed, that approximately 250,000 patent documents (after 1978) will be converted. It was also assumed, that no equipment will be purchased for this purpose.

- Summary of Costs

	Cost Component	Estimated Amount
a.	Scanning/OCR-ing	$2,083,333
b.	Document preparation/re-assembly	$1,000,000
c.	Manual indexing	$375,000
d.	Equipment rental	$1,210,914
	Total Backfile Conversion Costs	$4,669,247

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Scanning/OCR-ing	A.7.1.
b.	Document preparation/re-assembly	A.7.2.
c.	Manual indexing	A.7.3.
d.	Equipment rental	A.7.4.

C.4.2.1.8. Capital Set-up Charges For External Database Access

- Description

  Any initial charges associated with the set-up of access or purchase of a "license to use" external databases.

- Costing Assumptions and Methods

  The cost estimates are based on quotations obtained from the CPO, and other consulting teams. The estimate provided is based on the cost of acquiring the Messenger search software package from Chemical Abstracts, the supplier of such search software to the USPTO.

- Summary of Charges

	Cost Component	Estimated Amount
a.	Messenger perpetual license	$130,000

- References

  See spreadsheet reference A.8.1. Annual charges for software maintenance are included in cost component B.5.

C.4.2.1.9. Marketing and Promotion

- Description

  These costs encompass the development and delivery of a communications program to in form the Canadian public of the new system and services.

- Costing Assumptions and Methods

  The cost are based on the estimated funding requirements to implement the following 5 communications programs in years 4 to 8 of the project:

  a. Remote Access Services

  b. Public Search Room Services

  c. Technical Information Search Program

  d. Public Education Program

  e. Publication Services

An estimated implementation cost of $150,000 was allocated to implement each program. These costs cover the development of informational brochures and on-line help services and user manuals.

- Summary of Marketing and Promotion Costs

	Cost Component	Estimated Amount
a.	Remote Access Services	$150,000
b.	Public Search Room Services	$150,000
c.	Technical Information Search Program	$150,000
d.	Public Education Program	$150,000
e.	Publication Services	$150,000
	Total Marketing and Promotion	$750,000

- References

These cost estimates were provided by RES.

	Cost Component	Spreadsheet Reference
a.	Remote Access Services	A.9.1.
b.	Public Search Room Services	A.9.2.
c.	Technical Information Search Program	A.9.3.
d.	Public Education Program	A.9.4.
e.	Publication Services	A.9.5.

C.4.2.1.10. File Translation Costs

- Description

These costs encompass all expenses related to the implementation of an in-house database of in-force USPTO abstracts. This represents 1,000,000 documents.

- Costing Assumptions and Methods

Pilot Backfile Conversion results were used to estimate tape-related charges and the necessary workload for identification of files and conversion to Canadian format.

Two cost components were identified:

a. Tape-related charges including rental, storage, mounting, run, data conversion software, etc.

b. Person-years for identification of files and indexing.

- Summary of File Translation Costs

	Cost Component	Estimated Amount
a.	Tape-related Charges	$300,000
b.	Personnel Costs	$50,000
	Total Marketing and Promotion	$350,000

- References

	Cost Component	Spreadsheet Reference
a.	Tape-related Costs	A.10.1.
b.	Personnel Costs	A.10.2.

C.4.2.2. On-going Annual Operating Costs

C.4.2.2.1. Systems Management Costs

- Description

  These costs encompass all costs associated with the on-going costs of database administration, maintenance of systems software and the performance measurement and capacity planning functions of systems management. These costs are essentially human resource costs. Specifically, they cover

  a. compensation costs of government staff associated with the systems management function

  b. compensation costs of contracted staff associated with the systems management function

- Costing Assumptions and Methods

  The reference for these costs is based on the method described below using the following rates supplied by the CPO management.

  a. Compensation for government resources were based on an overall rate of $50,000 per man-year. This figure was assumed to include gross salary and 15% of gross salary for benefits. Allocations for overhead were not included.

  b. An average fee of $100,000 per person-year was used to estimate the costs for contracted resources.

  The systems management staff costs have been derived by first developing the schedule of major activities for each year of the project. This schedule is contained in section X of Volume Two of this report, (attached). A generic set of postitions for the organization of the team required to support the systems management functions was then defined. The project

team organization is contained in section X of Volume Two of this report, (attached). Estimates in person-years of support required from each position were then computed using the schedule of major activities to be performed.

- Cost Summary of Systems Management Costs

	Cost Component	Estimated Amount
a.	Government resource costs	$2,025,000
b.	Contracted resource costs	$0
c.	Equipment and Supplies	$30,000
	Total Systems Management Costs	$2,055,000

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Government Resource Costs	A.1.1.
b.	Contracted Resource Costs	A.1.2.

C.4.2.2.2. Computer Operations Costs

- Description

These costs encompass all costs associated with the on-going operation of the computer facility. These costs are made up of the following components:

a. compensation costs of Government staff associated with the computer operations function

b. compensation costs of Contracted staff associated with the computer operations function

- Costing Assumptions and Methods

The reference for these costs is based on the method described below using the following rates supplied by the CPO management.

a. Compensation for government resources were based on an overall rate of $50,000 per man-year. This figure was assumed to include gross salary and 15% of gross salary for benefits. Allocations for overhead were not included.

b. An average fee of $100,000 per man-year was used to estimate the costs for contracted resources.

The computer operations staff costs have been derived by first developing the schedule of major activities for each year of the project. A generic set of positions for the organization of

the team required to provide the systems operations functions was then defined. Estimates in person-years of support required from each position were then computed using the schedule of major activities to be performed.

- Cost Summary of Computer Operations Costs

	Cost Component	Estimated Amount
a.	Government Resource Costs	$2,162,500
b.	Contracted resource Costs	$5,850,000
c.	Supplies	$30,000
	Total Computer Operations Costs	$8,042,500

- Reference Calculations

	Cost Component	Spreadsheet Reference
a.	Government Resource Costs	B.2.1.
b.	Contracted Resource Costs	B.2.2.
c.	Supplies	B.2.3.

C.4.2.2.3. Hardware Maintenance Costs

- Description

  These costs encompass all costs associated with the on-going maintenance of the computer facility.

- Costing Assumptions and Methods

  Maintenance costs were estimated at 6% of the capital costs of the hardware on an annual basis and applied commencing 1 year after the purchase of the equipment.

- Cost Summary of Hardware Maintenance Costs

	Cost Component	Estimated Amount
a.	CPU and associated hardware	$2,880,000
b.	On-line storage	$668,000
c.	Off-line storage	$99,000
d.	Workstations	$750,000
e.	Test equipment	$36,000
	Hardware Maintenance	$4,433,000

- Reference

	Cost Component	Spreadsheet Reference
a.	CPU and associated hardware	B.3.1.
b.	On-line storage	B.3.2.
d.	Off-line storage	B.3.3.
e.	Workstations	B.3.4.
g.	Test equipment	B.3.5.

Hardware maintenance costs were developed from vendor quotations.

C.4.2.2.4. Telecommunications Costs

- Description

These costs encompass all costs associated with the provision of wide area communications services to external users.

- Costing Assumptions and Methods

The cost estimates are based on the assumptions that the CPO will subsidize the usage costs of external users during the project period. The external facilities are introduced (according to the schedule of major activities), in the 7th year of the project. Costs of providing communications services were based on the Coopers & Lybrand Market Analysis of external user needs. A traffic model was developed from the survey of demand and rated to determine basic packet switched service requirements.

An X.25 test circuit was costed in for each year prior to the introduction of general service. This test circuit was estimated to cost $10,000 per year and its use was intended to support development and evaluation activities.

- Cost Summary of Telecommunications Costs

	Cost Component	Estimated Amount
a.	CPO X.25 Circuits	$240,000
b.	X.25 Usage (External Users)	$765,000
		$1,005,000

- Reference

	Cost Component	Spreadsheet Reference
a.	CPO X.25 Circuits	B.4.1.
b.	X.25 Usage (External Users)	B.4.2.

The detailed costing analysis for this component can be found in Appendix D.

C.4.2.2.5. Software Licence Costs

- Description

  These costs contain all on-going licence fees paid after systems, application and miscellaneous software. The software purchased for external database access is not included - that cost assumption was based on assuming an eternal licence.

- Costing Assumptions and Methods

  It was assumed in section C.4.2.1.4. (Application Software Develoment/Acquisition) that there all application software will be developed from scratch. Therefore there are no application software licences to be payed. Note, that this is the worst case - purchase of software tools would be cheaper than development.

  According to this assumption, the only software licence to be paid will be that of the system software.

  Licence costs were estimated at 5% of the software purchase prices on an annual basis and applied commencing 1 year after the purchase.

- Cost Summary of Software Licences

	Cost Component	Estimated Amount
a.	System Software Licence	$150,000
b.	Application Software Licence	$0
c.	Miscellanous Software Licence	N/A
	Software Licences Total:	$150,000

- Reference

	Cost Component	Spreadsheet Reference
a.	System Software Licence	B.5.1.
b.	Application Software Licence	B.5.2.
c.	Miscellanous Software Licence	N/A

C.4.2.2.6. External Database Usage Costs

External database usage, on its present level, was not costed in the project budget. For a possible future demand, no estimate is available, therefore this element was not costed.

C.4.2.2.7. Application Software Maintenance Costs

- Description

  The costs associated with the on-going maintenance of the applications software for the CPO system.

- Costing Assumptions and Methods

  The software development exercise was assumed to be the duty of CPO staff. The average yearly compensation of Government staff is estimated to be $50,000.

  The applications software maintenance costs were derived by first developing the schedule of major activites in respect to the development of the applications software, and then deriving the maintenance workload expected. This workload defines the generic positions required for maintenance activities.

  The development tools and supplies, such as workstations and development languages are included under Application Software Development - it is assumed that these supplies will be inherited by the software maintenance staff when development work is finished.

- Cost Summary of Application Software Maintenance

	Cost Component	Estimated Amount
a.	Government Resource Costs	$1,188,000

- Reference

	Cost Component	Spreadsheet Reference
a.	Government Resource Costs	B.7.1.

C.4.2.2.8. Training and User Support Costs

- Description

  These cost element includes the compensation of user training and support resources. It does not include costs associated with the development of training programs or documentation - these are considered capital cost components and listed under the implementation costs.

- Costing Assumptions and Methods

  It is assumed that the resources for on-going training and user support are government staff. The average compensation is estimated to be $50,000/year/resource.

- Cost Summary of Training and User Support

	Cost Component	Estimated Amount
a.	Government Resource Costs	$837,500

- Reference

	Cost Component	Spreadsheet Reference
a.	Government Resource Costs	B.8.1.

C.4.2.2.9. Facilities Maintenance Costs

- Description

  This cost element includes all rental and maintenance costs of floor space and computer room facilities.

- Costing Assumptions and Methods

  Floor rental cost is estimated to be $1.63 per square feet per month. Maintenance costs are estimated to be 12% of the facilities setup costs annually.

- Cost Summary of Facilities Maintenance

	Cost Component	Estimated Amount
a.	Space Rental/Lease Costs	$1,682,000
b.	Facilities Maintenance	$154,000
	Facilities Maintenance Total:	$1,837,000

- Reference

	Cost Component	Spreadsheet Reference
a.	Space Rental/Lease Costs	B.9.1.
b.	Facilities Maintenance	B.9.2.

C.4.2.2.10. On-going Data Capture Costs

- Description

  This cost element includes all costs associated with the data capture of patent applications.

- Costing Assumptions and Methods

  All assumtions are the same as for the backfile conversion except that in this case purchased equipment is used.

The number of applications is expected to grow up to 35,000 a year by the end of the project period. The calculation of these costs assumes this number as a basis.

- Cost Summary of On-going Data Cpture

	Cost Component	Estimated Amount
a.	Equipment Purchase Costs	$460,000
b.	Equipment Maintenance Costs	$166,000
c.	Scanning, OCR-ing	$1,896,000
d.	Preparation	$910,000
e.	Indexing	$341,000
		3,773,000

- Reference

	Cost Component	Spreadsheet Reference
a.	Equipment Purchase Costs	B.10.1.
b.	Equipment Maintenance Costs	B.10.2.
c.	Scanning, OCR-ing	B.10.3.
d.	Preparation	B.10.4.
e.	Indexing	B.10.5.

C.4.2.2.11. Publishing Costs

In this preliminary cost estimation, publishing costs are assumed to be zero, for the following reasons:

- Report-like publications (CPOR) can be produced as part of the computer operations, with virtually no extra cost
- Publication of new patents/applications (18 month) does not represent any significant cost
- Extra publications may (and should) be turned into a revenue-generating enterprise.

This cost element is referred as B.11. on the spreadsheet.

C.4.2.2.12. Translation Costs

- Description

  This cost elements includes all costs associated with the language translation of software (user interfaces) and manuals.

- Costing Assumptions and Methods

  For this cost calculation, it is assumed that 500,000 words will annually be translated, for $0.20/word, starting in year two of the project.

- Cost Summary of Translation

	Cost Component	Estimated Amount
a.	Translation	$700,000

- Reference

	Cost Component	Spreadsheet Reference
a.	Translation	B.12.1.

C.5. Detailed Staffing Plan

Government Staff Reguirements

Position Name	No. of positions	Total Man-Mths
Project Director	1	96
Executive Assistant	1	96
Executive Secretary	1	96
Asst. Proj. Dir. (Plng & Admin)	1	96
Asst. Proj. Dir. (Contract Adm.)	1	84
Asst. Proj. Dir. (User Srvcs)	1	96
Clerical Support	1	96
Asst. Proj. Dir. (Systems Devlpmnt.)	1	93
Clerical Support	1	84
Asst. Proj. Dir. (Sys. Mngmt.)	1	84
Systems Mngr. (Sys. Plng.)	1	84
Sys. Specialist (Cap. & Perf. Meas.)	1	78
Systems Mngr. (Security)	1	84
Sys. Specialist (DB & Network Sec.)	1	78
Clerical Support	1	78
Asst. Proj. Dir. (Sys. & Ops. Support)	1	90
Systems Mngr. (DB Administration)	1	84
DB Systems Specialist	1	81
Systems Mngr. (Systems Software)	1	90
Systems Software Programmer	1	78
Clerical Support	1	96
Sen. Systems Analyst (SW Maint.)	1	78
SW Specialist (DB Maintenance)	1	69
SW Specialist (Exm/Srch & Ap. Proc.)	1	69
SW Specialist (Diss. & Mngmt. Inf.)	1	69
Systems Mngr. (Training & Support)	1	69
Instructor (2)	2	132
Project Management Staff Total	28	2,328

Contracted Staff Requirements

Position Name	No. of positions	Total Man-Mths
Proj. Mngr. (QA & Methods)	1	51
Proj. Mngr. (Recr. & Staff)	1	39
Proj. Mngr. (Training Adm.)	1	39
Proj. Mngr. (Fin. Accnting)	1	69
Admin Assistant	1	63
Proj. Mngr. (QC)	1	78
Proj. Mngr. (Fac. & Equip.)	1	78
Proj. Mngr. (Contr. Ops.)	1	78
Sen. Sys. An. (Inf. Stor. & Retr.)	1	66
Systems Analyst (2)	2	144
Sen. Sys. An. (Reg. Systems)	1	72
Systems Analyst (2)	2	144
Sen. Sys. An. (Diss. Systems)	1	63
Systems Analyst (2)	2	126
Sen. Sys. An. (Mgmt. Inf. Sys.)	1	72
Systems Analyst (2)	2	144
Sen. Design Specialist (DB)	1	66
DB Design Specialist (2)	2	132
Sen. Design Specialist (Exam./Srch.)	1	45
Appl. Software Specialist (2)	2	90
Sen. Design Specialist (Appl. Proc.)	1	30
Appl. Software Specialist (2)	2	60
Sen. Design Specialist (Inf. Diss.)	1	36
Appl. Software Specialist (2)	2	60
Sen. Design Specialist (Mgmt. Inf.)	1	24
Appl. Software Specialist (2)	2	48
Contracted Trainer/Developer (2)	2	84
Systems Mngr. (Computer Operations)	1	78
Computer Operator (4)	4	312
Data Entry Operator (2)	2	156
Systems Mngr. (Data Communications)	1	78
Data Communications Technician	1	78
Contracted Staff Total	46	2,703

APPENDIX D: WIDE AREA NETWORK REQUIREMENTS

It is consistent with the automation objectives of the CPO in view of its dissemination function, to provide low cost remote access to the Automated Patent Information Processing System and services to all regions.

On the basis of analysis incorporated within this document, it is recommended that the following implementation strategy be adopted for all options.

a. The automated patent system provides universal remote network access services to external users only after it has become fully operational for internal user needs.

b. Network services will be introduced on a pilot basis following the introduction of the fully integrated text and image database. In order to minimize costs, services will be expanded on an incremental basis subject to increased usage and demand as well as the development and introduction of differentiated value-added information services.

c. In this phase, Datapac network access arrangements will provide connections from the workstations to the host facility. This phase is estimated to last until the completion of the automation project in 1996, at which time the requirement for further expansion of the services can be evaluated from the historical usage data obtained during the previous three years of the pilot study.

D.1. Model of the Network Usage of General Users

D.1.1. Market Demand Represented by General Users

Data for developing a usage model for general users of patent information was derived from projections of demand developed in the Market Demand Assessment study. That study reported a total of 1,131,853 "uses of patent information" by the non patent agent community on an annual basis. The median estimate was based on standard statistical survey assumptions including 95% confidence intervals. The assumptions were tested for sensitivity to yield a "worst case" estimate of 272,748 projected uses. These survey results are summarized in the following table.

Estimates of Total Annual Demand For Patent Information
Current Demand

Group	Frequency of Use of Patent Information Per Year	
	Standard Estimate	Worst Case Estimate
Small Business	54,285	5,428
Medium Business	23,316	2,961
Large Business	40,349	6,578
R & D Firms	8,619	3,060
Chemists	208,505	67,332
Engineers	788,655	185,442
TIS Users	8,123	1,944
Total	1,131,853	272,748
Patent Agents	21,348	21,348
Grand Total	1,153,201	294,096

These results were used together with the assumption that a "use" (in the case of the general users) could be interpreted to correspond to an access to the automated database (to get the information to use). A further study of the survey data revealed the following distribution of the utility to users of the following patent information data types.

Utility of Patent Information

	Preference Weights				Total (weighted)	Weighted Average
Data Type	3/3	2/3	1/3	0/3	(%)	
Classification	220	286	147	139	460	16.76%
Tombstone	517	519	303	266	954	17.35%
Abstract	353	238	97	114	544	19.59%
Disclosure	195	281	181	142	443	16.00%
Claims	170	215	225	175	388	14.29%
Drawings	209	256	186	146	442	16.01%
Accesses Total:	1664	1197	380	0	3231	100.00%

These data types can be taken as typical components of the patent database. For this reason, the total number of accesses for both the median and worst case estimates were distributed using these percentages to arrive at a model for the number of accesses by general users on each of the database component information sources. The distribution is illustrated below.

Projected Database Accesses

	Number of Accesses per Year		
Access Type	Median	Worst Case	Percent
Classification	189,739	45,772	16.76%
Tombstone	196,354	47,317	17.35%
Abstract	221,750	53,346	19.59%
Disclosure	181,121	43,646	16.00%
Claims	161,724	38,971	14.29%
Drawings	181,165	43,656	16.01%
Total	1,131,850	272,748	100.00%

D.1.2. Assumptions Concerning Network Facilities of General Users

For cost estimating purposes, it was assumed that general users would connect to the CPO system using Datapac public dial access arrangements at a data transfer speed of 2400 bits per second (bps). This arrangement was selected as the most probable choice of potential users for the following reasons.

- As the number of users is not known in this group, a monthly access charge for a dedicated line or private dial arrangement cannot be computed. Thus public dial with holding time charging is substituted for access charges on dedicated arrangements.

- The volume of usage is too low to favour dedicated access over public dial if the number of users is over 3,000. The ratio of the projected number of uses of patent information to the number of patent agents as reported in the Market Demand Assessment is 71.8 "uses" per agent per year. This group represents the heaviest group of users and as such we would expect that general users would use patent information less frequently. This assumption would suggest that the number of general users is larger than (1,131,853/71.8=15,764) 16,000 in the median estimate and (272,748/ 71.8= 3799) 3,800 users in the worst case estimate. Dedicated access at 2400 bps, however, becomes more cost effective in this scenario when the number of users is less than 53.

- For general users, 53% of their searches are projected to be searches on the classification, tombstone data and abstract. The time to transmit results over the network for these search activities varies from less than 3 seconds for classification and tombstone data to 12 seconds for the transmission of abstracts. For this reason, 2400 bps service can be considered as a somewhat slow but very cost effective access arrangement for an occassional user.

- The vast majority of users of data network services use 1200/2400 bps public dial-up to access host computers.

D.2. Model of the Network Usage of Patent Agents

D.2.1. Market Demand Represented by Patent Agents

The Market Demand Assessment reported that the community of 297 patent agents within Canada has a current total demand of 21,348 uses of patent information annually. This translates to 71.88 uses per year per agent. The survey also reported the following statistics describing the sources used for searching and the average number of searches performed annually per agent:

Search Statistics

Type of Search	Average Number of Searches/year
Searches on Commercial Databases	128
Searches Using In-house System Consisting of a File System and Index	69
Searches Conducted at CPO	229
Searches Conducted at USPTO	76
Total	502

The 297 patent agents perform an average of 502 searches per year each and use patent information a total of 21,348 times annually. It was therefore concluded that each use of the patent system corresponds to 8.3752 searches on the average. It should also be noted that patent agents perform a total of 178,794 searches per year of which 68,013 are searches performed on the CPO patent files.

In the Market Demand Assessment, patent agents were asked how often they would search an on-line patent information system to access specific features which could be included. The resulting potential distribution of projected demand is illustrated on the following table.

Projected Demand

Access Type	Average Number Searches /Year	Per Cent of Total	
		Indiv.	Cumul.
CPO Classification Index	37	10.22	10.22
Tombstone Index for 20 years of Canadian Patents	55	15.19	25.41
Abstract,1st Claim for 20 Years of Canadian Patents	42	11.60	37.01
Full text & Claims for 20 years of Canadian Patents	48	13.26	50.27
Full text, Claims & Drawings for 20 years of Canadian Patents	55	15.19	65.46
Full text & Claims for 20 years of U.S. Patents	56	15.47	80.93
Full text & Claims for 20 years of EPO & JPO Patents	22	6.08	87.02
Full text, Claims & Drawings for 20 years of U.S.,EPO & JPO Patents	47	12.98	100.00
Total	362		

It is clear from these results that 65% of patent agents searches would be performed on the database of Canadian in-force patents which would number some 400,000. Since U.S. and foreign patents would require the addition of several million patents to the database at proportionate costs and increase demand by only 35%, we have used the assumption that these databases would be searched separately by accessing other systems. Under the assumption that these databases would be included or that on-line access to the CPO, EPO and JPO database, the resulting base estimate should be increased by the corresponding percentages shown in the above table.

The projected total annual demand of 21,348 uses of patent information by 297 agents on an automated CPO system containing only 20 years of Canadian patents would then be described by the following table.

Projected Searches of Patent Information

Access Type	Average Number Searches/Year	Percent of Total	
		Individual	Cumul.
CPO Classification Index	37	15.61	15.61
Tombstone Index	55	23.21	38.82
Abstract,1st Claim	42	17.72	56.54
Full text & Claims	48	20.25	76.79
Full text, Claims & Drawings	55	23.21	100.00
Total	237		

The projected number of searches on such an automated system corresponds closely to the actual number (229) reported on the current CPO file. The total number of projected searches assuming the number of agents stays constant is 70,389 which of course, correlates to the current number of 68,013. Using the projected figures, then each "use" of patent information corresponds to 3.2972 searches. Using these estimates, the following page and document volumes are established.

Projected Page and Document Volumes

Access Type	Search per Year		Accesses per Year	
	Pages	Documents	Pages	Documents
CPO Classification Index	1	37	10,989	10,989
Tombstone Index	1	55	16,335	16,335
Abstract, 1st Claim	1	42	12,474	12,474
Full text & Claims	18	48	256,608	14,256
Full text, Claims & Drawings	20	55	326,700	16,335
Total		237	623,106	70,389

D.2.2. Assumptions Concerning Network Facilities for Patent Agents

Since patents are the business of patent agents, it can be assumed that they will perform searches and retreive documents much more frequently than general users. Although 56% of their searches will be searches on the classification, tombstone and abstract, searching and retreival of disclosure, claims and drawings will consume more than 94% of their session time. Thus, trading off increased transmission speed against cost will be a critical factor for this group. The Market Demand Assessment reports that 49% of the population will be willing to pay more than twice for what they currently get from the CPOR (tombstone data). We assume therefore, that some of the population will access the CPO on 2400 bps dial-up and some on higher speed dedicated access arrangements. Since there are 62 independent patent agents and 72 firms representing the population of 297 known patent agents for a total of 134 sites, we assumed that 62 agents will access the CPO on 2400 bps dial-up facilities and 235 agents will access the CPO on dedicated facilities from 72 sites. Since there are an average of 3.26 agents per firm, we assumed that these users would multiplex 235 dedicated 2400 bps lines onto 72 9600 bps lines to realize further economies of scale. The distribution of page volumes is described in the following table.

Document Accesses

Access Type	Dial calls per Year	Accesses @2400bps	9.6 Calls per Year	Accesses @9600bps
CPO Classification Index	2,294	2,294	8,695	8,695
Tombstone Index	3,410	3,410	12,925	12,925
Abstract,1st Claim	2,609	2,604	9,865	9,870
Text & Claims	2,976	53,568	11,280	203,040
Text, Claims, Drawings	3,410	68,201	12,925	258,500
Total	14,699	130,076	55,690	493,030

D.3. Traffic Characteristics

In order to translate accesses to the CPO system into network traffic characteristics, the following assumptions were used.

Each access was assumed to represent one terminal session or one call by a user. It was further assumed that during each call only one type of information was searched for and transferred to the terminal user. The number of pages transmitted to the user was based on standard document statistics used in this study. The classification data transmitted was assumed to consist of one page of text.

The data representing page transmissions was assumed to be in integrated text and image formats. This assumption provides a measure in absolute units for conversion of pages to packets. For comparison purposes, the corresponding page sizes in compressed image format at 300 dpi resolution were also used to demonstrate that the wide area transmission of data in this format is not feasible from a cost perspective.

Twenty-four overhead control bits were added to the estimated information bits for each packet to derive call durations at given line speeds.

D.3.1. General Users: Traffic Volumes

On the basis of the previous assumptions, the table describing the traffic charactistics of the general users was derived from the distribution of their annual accesses and is illustrated below. The first table describes unit conversions for converting data in both integrated text and image (ITI), and compressed image (CI), formats into packets. For each data type the corresponding packet ratios were computed. The second table describes the resulting network traffic in terms of kilopackets (1000 packets) and call durations for integrated text and image formats. The third table describes the traffic volumes assuming compressed image formats. The call durations for this last table were computed using the packet ratios in Table 1 and applying them to the call durations computed in Table 2.

Conversion of Pages To Packets

		CI		ITI		Packet Ratios
Call Type	Pages	Kbytes	Packets	Kbytes	Packets	
Classification	1	93.38	374	0.54	3	124.67
Tombstone	1	93.38	374	0.54	3	124.67
Abstract	1	217.90	872	3.52	15	58.13
Disclosure	15	3268.50	13,074	52.73	211	61.96
Claims	3	653.70	2,615	10.56	43	60.81
Drawimgs	2	130.74	523	130.74	523	1.00
Total	23	4457.60	17,832	77.01	798	22.35

Traffic Volumes Assuming ITI Format

Call Type	Pages	Packets	Packets/year	Calls/year	Duration 2400 bps
Classification	1	3	569,217	189,739	2.59
Tombstone	1	3	589,062	196,354	2.59
Abstract	1	15	3,326,250	221,750	12.95
Disclosure	15	211	38,216,531	181,121	182.16
Claims	3	43	6,954,132	161,724	37.12
Drawimgs	2	523	94,749,295	181,165	451.52
Total	23	798	144,431,487	1,131,853	688.93

Traffic Volumes Assuming CI Formatted Data

Call Type	Pages	Packets	Packets/year	Calls/year	Duration 2400 bps
Classification	1	374	70,962,386	189,739	322.90
Tombstone	1	374	73,436,396	196,354	322.90
Abstract	1	872	193,366,000	221,750	752.78
Disclosure	15	13,074	2,367,975,954	181,121	11,286.82
Claims	3	2,615	422,908,260	161,724	2,257.45
Drawimgs	2	523	94,749,295	181,165	451.52
Total	23	17,832	3,223,398,291	1,131,853	688.94

D.3.2. Patent Agents: Traffic Volumes

The following table provides a conversion to packets for the pages accessed and transferred to the agent during a call or session.

Conversion of Pages Accessed to Packets

Access Type		CI		ITI		Packet Ratios
	Pages	Kbytes	Packets	Kbytes	Packets	
Classification	1	93.38	374	0.54	3	124.67
Tombstone	1	93.38	374	0.54	3	124.67
Abstr. & Claim	1	217.90	872	3.52	15	58.13
Full Text	18	3922.20	15,689	63.36	254	61.77
Text, Drawings	20	4052.94	16,212	194.10	777	20.86

The following tables describe traffic volumes for the 62 patent agents using Datapac 2400 dial-up service and the 235 patent agents using 9600 bps access arrangements on 72 dedicated lines.

Traffic Volumes Assuming Integrated Text and Image Format

2400 Dial-Up Sercice

Call Type	Pacs/Page	Accesses @2400bps	Packets/year	Calls/year	Duration @2400bps
Classification	3	2,294	6,882	2,294	2.59
Tombstone	3	3,410	10,230	3,410	2.59
Abstr. & Claim	15	2,604	39,060	2,609	12.95
Full Text	254	53,568	13,606,272	2,976	3,948.16
Text, Drawings	777	68,201	52,992,177	3,410	13,416.39
Total	798	130,076	66,654,621	14,699	

9600 bps Dedicated Access Arrangement

Call Type	Pacs/Page	Accesses @9600bps	Packets/year	Calls/year	Duration @2400bps
Classification	3	8,695	26,085	8,695	2.59
Tombstone	3	12,925	38,775	12,925	2.59
Abstr. & Claim	15	9,870	147,975	9,865	12.95
Full Text	254	203,040	51,572,160	11,280	3,948.16
Text, Drawings	777	258,500	200,854,545	12,925	13,416.39
Total	798	493,030	252,639,540	55,690	

D.4. Estimates of Network Costs

D.4.1. Estimating Procedures

1. An average packet rate of .75 dollars per kilopacket is used for estimating network usage charges.

2. Datapac (1987) rates provided by Telecom Canada were used to compute usage charges. In particular, it should be noted that call-based accounting is used in which the procedures are as follows.

    a. The kpac rate(s) is (are) applied to fractional kpacs on a per call basis and rounded to the next higher cent. The number of charges that apply per call depend on the service arrangement.

    b. The resultant charge(s) for each call is(are) rounded to the nearest cent with a minimum charge of one cent(per chargable item) for each call.

    c. On dial access arrangements, the holding time of a call is rounded to the next higher minute and the holding time rate applied. Thus a minimum charge of 0.04 dollars per minute applies.

    d. One minute was added to the holding time estimates for each typical call to account for interactive communications to establish transactions. The data amount these interactive communications represent is considered to be insignificantly small.

Using the above costing procedures and assumptions the network charges illustrated on the following tables were determined from the traffic projections. Charges are calculated first for data assuming integrated text and image format and then for data assuming it is transferred on the network in compressed image format. The results are then compared in tabular form. It is clear from this last table that transmitting data in CI format is about 20 times as expensive as transmitting it in ITI format. In the previous calculations of traffic volumes using these formats, it can be observed that for a given line speed, CI formatted data takes as much as 124 times as long as transmission of the same information in ITI format. This clearly demonstrates the futility

of basing a design on compressed image format in the view of the evolving standards and the OCR technology available today.

D.4.2. General Users: Network Charges

D.4.2.1. General Users: Network Charges Assuming ITI Format

Call Type	kpacs/call	Calls/year	Network Usage	Pad Usage	Call Setups
		.75/kpac		.47/kpac	.0067/call
Classification	.003	189,739	426.93	267.54	1,271.25
Tombstone	.003	196,354	441.80	276.86	1,315.57
Abstract	.015	221,750	2,494.68	1,563.34	1,485.73
Disclosure	.211	181,121	28,662.40	17,961.77	1,213.51
Claims	.043	161,724	5,215.60	3,268.45	1,083.55
Drawings	.523	181,165	71,061.98	44,532.17	1,213.81
Averages/Totals	.128	1,131,853	108,303.39	67,870.13	7,583.42

Dial Access

Public Dial (.04/min)

Calls/Year	Rated Minutes	H.T. Charges
189,739	1.00+1	15,179.12
196,354	1.00+1	15,179.12
221,750	1.00+1	17,740.00
181,121	4.00+1	36,224.20
161,724	1.00+1	12,937.92
181,165	8.00+1	65,219.40
1,131,853	3.60	162,479.76
Private Dial on-net	184/mon.	N/A
Dedicated on-net	193/mon.	N/A

D.4.2.2. General Users: Network Charges Assuming CI Format

Call Type	kpacs/call	Calls/year	Network Usage	Pad Usage	Call Setups
		.75/kpac		.47/kpac	.0067/call
Classification	.374	189,739	53,127	34,153	1,271
Tombstone	.374	196,354	54,979	35,344	1,316
Abstract	.872	221,750	144,138	90,918	1,486
Disclosure	13.074	181,121	1,776,797	1,112,083	1,214
Claims	2.615	161,724	316,979	198,921	1,084
Drawings	.523	181,165	70,654	45,291	1,214
Totals		1,131,853	2,416,674	1,516,710	7,585

Dial Access
Public Dial (.04/min)

	Calls/year	Holding Time (minutes)	Rated Minutes	H.T. Charges
Classification	189,739	5.38	6.00+1	53,126.92
Tombstone	196,354	5.38	6.00+1	54,979.12
Abstract	221,750	12.55	13.00+1	124,180.00
Disclosure	181,121	188.11	189.00+1	1,376,519.60
Claims	161,724	37.62	38.00+1	252,289.44
Drawings	181,165	7.53	8.00+1	65,219.40
Totals	1,131,853			1,926,314.48
Private Dial on-net	184/mon.			
Dedicated on-net	193/mon.			

D.4.2.3. Contribution of Annual Network Usage Cost of General Users

The following table provides a comparision of the usage costs of remote access facilities for the ITI and CI formats.

	Integrated Text & Image	Compressed Image
Network Usage	$108,303.39	$2,416,673.94
Pad Usage	$67,870.13	$1,516,708.95
Call Set-up	$7,583.42	$7,583.42
Holding Time	$162,479.76	$1,926,314.48
Total Charges	$346,236.70	$5,867,280.79

It is clear from the above table that the usage costs of transmission of CI formatted data will be 17 times the cost of transmitting the same information in ITI format.

D.4.3. Patent Agents: Network Charges

D.4.3.1. Patent Agents: Network Charges Assuming ITI Format

The following table describes traffic charges for the 62 patent agents using Datapac 2400 dial-up service.

Call Type	kpacs/call	Calls/year	Network Usage	Pad Usage	Call Setups
		.75/kpac		.47/kpac	.0067/call
Classification	.003	2,294	22.94	22.94	15.37
Tombstone	.003	3,410	34.10	34.10	22.85
Abstr. & Claim	.015	2,609	52.18	26.09	17.48
Full Text	4.572	2,976	10,207.68	6,398.40	19.94
Text, Drawings	15.540	3,410	39,760.60	24,927.10	22.85
Totals		14,699	50,077.50	31,408.63	98.49

Dial Access
Public Dial (.04/min)

	Calls/year	Call Duration	Rated Minutes	H.T. Charges
Classification	2,294	2.59	1.00+1	183.52
Tombstone	3,410	2.59	1.00+1	272.80
Abstr. & Claim	2,609	12.95	1.00+1	208.72
Full Text	2,976	3,948.16	66.00+1	7,975.68
Text, Drawings	3,410	13,416.39	224.00+1	30,690.00
Totals	14,699			39,330.72

The following table describes traffic charges for the 235 patent agents using 72 Datapac 9600 circuits. Each circuit is assumed to be configured with 4 virtual circuits.

Call Type	kpacs/call	Calls/year	Pad Usage	Network Usage	Call Setups
		.75/kpac		.47/kpac	.0067/call
Classification	.003	8,695	86.95	N/A	58.25
Tombstone	.003	12,925	129.25	N/A	86.59
Abstr. & Claim	.015	9,865	197.30	N/A	66.10
Full Text	4.572	11,280	38,690.40	N/A	75.58
Text, Drawings	15.540	12,925	150,705.50	N/A	86.59
Totals		55,690	189,809.40	N/A	373.11

Dedicated on-net: 456.00/mon. × 12 mon. × 72 circuits = $393,984

Switched Virtual Circuits: 4 svc.'s/circ. × 72 circ.'s × 12 mon.'s × 3.65/mon. = $12,614.40

D.4.3.2. Contribution of Annual Network Usage Cost of Patent Agents

	Dial-Up Access @2400bps	Ded. Access @9600bps	
Network Usage	$50,077.50	$189,809.40	
Pad Usage	$31,408.63	N/A	
Call Set-up	$98.49	$373.11	
Holding Time	$39,330.72	N/A	
Access Charges	N/A	$406,598.40	
Total Charges	$120,915.34	$596,780.91	$717,696.25

D.4.4. Network Usage Cost

D.4.4.1. Total Network Usage Costs Assuming ITI Format

The following table summarizes the total costs of the projected usage by both groups of general users and patent agents.

1. General Users	$346,236.70
2. Patent Agents	$717,696.25
Total	$1,063,932.90

D.4.4.2. Total Network Usage Costs Assuming CI Format

Using the 1 to 18 ratio derived previously in comparing general users network costs in the different formats, we estimate that the total cost in CI format would be approximately 19 million dollars in annual usage costs.

D.4.4.3. Central Facilities Costs

This section deals with the costs of provisioning the central CPO computer with Datapac synchronous circuits required to handle the projected external market traffic. The charges that apply here deal only with the fixed monthly cost of the rental of the circuit. This charge depends only on the line speed subscribed to and the number of virtual circuits configured on the physical link at installation time. We shall assume that only one switched virtual circuit is defined for any given line speed selected.

In order to price out the circuits, the number of circuits required must be determined. This has been done by considering the call durations for each type of transaction at each available line speed and computing from the transaction rates, the probabilities of a given number of concurrent

calls placed to the host computer in the peak period (see Appendix A). The results, assuming either Poisson or polynomial distributions for the treatment of transaction probabilities, are illustrated in the table below for a line speed of 19,200 bits per second. The computation of the probabilities was applied to the groups of general users and patent agents separately and are given for the 90%, 99% and 99.99% confidence intervals.

Concurrent Requests from	90%	95%	99%	99.99%
General Users	4	4	6	9
Patent Agents	8	10	12	18
Total	1	14	18	27

From the table calculated, it can be observed that no more than 18 concurrent requests will occur 99% of the time. Thus, if this confidence interval is taken as the peak load design requirement, then 18 19.2 kbs circuits are required to handle the load. The cost of these facilities is included in the following table which prices out the circuits for each of the preceding results.

Alternative Design Requirements Circuits @ 19,200 bps	90%	95%	99%	99.99%
General Users	$36,480	$36,480	$54,720	$82,080
Patent Agents	$72,960	$91,200	$109,440	$164,160
Total	$109,440	$127,680	$164,160	$246,240

A monthly fixed rate of $750.00 was used for the purposes of these calculations.

Assuming a 95% confidence level, the cost of provisioning the 14 required circuits is estimated to be $127,680.

D.5. Total Usage and Facilities Costs

The total usage and facilities costs are projected to be

	USAGE COSTS	ACCESS COSTS	
General Users	$183,757	$162,479	
Patent Agents/dial-up	$81,585	$39,330	
Patent Agents/dedicated	$190,182	$406,598	
CPO Access	N/A	$127,680	
Total	$455,524	$736,087	$1,191,511

For estimating purposes, the projected holding time charges were included as substitute for fixed monthly access chrages on dial-up access arrangements.

Assuming the CPO adopts the policy of subsidizing external users for the duration of the automation project, the only cost component that one could audit to ensure that its incurrence was directly associated with CPO services is the usage. A user with a dedicated access arrangement could use that connection to the network to access any host computer on the network. Thus, assuming 100% subsidization of external usage, the projected cost incurred by the CPO would be the total usage costs in addition to the access charges for the CPO front end circuits.

Subsidized Costs	
CPO Access	$127,680
Usage	$455,524
Total	$583,204

BIBLIOGRAPHY

A. List of Tasks

Task 1: Work Flow Analysis
Internal CPO Study

Task 2: Clarifying Objectives
Internal CPO Study

Task 3: Trilateral Agreement
Internal CPO Study (incomplete)

Task 4: Market Demand Assessment
The Coopers & Lybrand Consulting Group

Task 4a: User Needs Analysis
Loecus Informatics Inc.

Task 5: General Requirements
Stevenson Kellogg Ernst & Whitney

Task 6: Technology Assessment
Idon Corporation

Task 7: Foreign Patent Office Studies & Evaluations
Internal CPO Study (incomplete)

Task 8: CPO Pilot Studies & Evaluations
The Institute of Technology Policy

Task 8a: Pilot Backfile Conversion
The Institute of Technology Policy

Task 8b: UBC Pilot Study
Internal CPO-UBC Study

Task 8c: Class Listing/Numeric Index
Canadian Technology Marketing Group
Indigo Software Limited
RES Policy Research Inc.

Task 9: Economic Impact Analysis
The Coopers & Lybrand Consulting Group

Task 10: Feasibility Study/Business Case Analysis
RES Policy Research Inc.

Task 11: Plan Promulgation
Internal CPO Activity

Task 12: Automated Translation
Odyssey Research Associates Inc.

B. Other Related Material

1. Canadian Patent Office Long Range Informatics Plan (LRIP)

2. IT-Gesamptkonzept 1988 des DPA (GPO - Germany)

3. WIPO PCPI/P 491/86 (on the Generic Coding of Patent Documents)

4. Australian Patent, Trade Marks and Designs Offices Computer Systems (Systems Management Section, 23 April 1987)

5. Trilateral Co-operation EPO-USPTO-JPO 1983-1985 and Associated Projects of the EPO (World Patent Information, Vol 8. No 2. pp 66-69, 1986)

6. The United States Patent and Trademark Office - Automation Master Plan (Edition 3, April 1987)