Information Society Technologies

european patent office

erpanet

www.erpanet.org

ERPANET – Electronic Resource Preservation and Access Network – is an activity funded by the European Commission under its IST programme (IST-2001-3.1.2). The Swiss Federal Government provides additional funding.

Further information on ERPANET and access to its other products is available at http://www.erpanet.org.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu.int).

ISSN 1741-8682

© ERPANET 2004

ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK

Table of Contents

© ERPANET 2004

Executive Summary

For conducting its core business the European Patent Office (EPO) manages huge amounts of information. EPO staff require daily access to information from around the globe which reaches as far back as the first patents from the 19th century. All this information is available in digital form (often today *only* in digital form). Consequently, digital preservation is not an additional policy measure but an inherent necessity that 'happens' as part of the active management of their information.

Cooperation with other patent offices around the world is integral to the EPO. This includes exchange of information, knowledge and even technology. Before a patent can be passed, other international patents need to be scrutinised. Indeed, patent applications are often on a multi-national scale and consequently cooperation between the patent offices enhances efficiency. Compliance to international standards and contribution to their definition is therefore an important business requirement.

Due to these demands the EPO invests heavily in acquiring and developing adequate technology, and in producing clear guidelines and procedures to ensure appropriate information management. Consequently, the EPO has a successful track record and gained invaluable experience in digital preservation as part of their active stewardship of information.

© ERPANET 2004

Chapter 1: The ERPANET Project

The European Commission and Swiss Confederation funded ERPANET Project[1] (Electronic Resource Preservation and Access Network) works to enhance the preservation of cultural and scientific digital objects through raising awareness, providing access to experience, sharing policies and strategies, and improving practices. To achieve these goals ERPANET is building an active community of members and actors, bringing together memory organisations (museums, libraries and archives), ICT and software industry, research institutions, government organisations, entertainment and creative industries, and commercial sectors. ERPANET constructs authoritative information resources on state-of-the-art developments in digital preservation, promotes training, and provides advice and tools.

ERPANET consists of four partners and is directed by a management committee, namely Seamus Ross (HATII, University of Glasgow; principal director), Niklaus Bütikofer (Schweizerisches Bundesarchiv), Hans Hofman (Nationaal Archief/National Archives of the Netherlands), and Maria Guercio (ISTBAL, University of Urbino). At each of these nodes a content editor supports their work, and Peter McKinney serves as a co-coordinator to the project. An Advisory Committee with experts from various organisations, institutions, and companies from all over Europe give advice and support to ERPANET.

[1] ERPANET is a European Commission funded project (IST-2001-32706). See www.erpanet.org for more details and available products.

© ERPANET 2004

Chapter 2: Scope of the Case Studies

While theoretical discussions on best practice call for urgent action to ensure the survival of digital information, it is organisations and institutions that are leading the drive to establish effective digital preservation strategies. In order to understand the processes these organisations are undertaking, ERPANET is conducting a series of case studies in the area of digital preservation. In total, sixty case studies, each of varying size, will investigate awareness, strategies, and technologies used in an array of organisations. The resulting corpus should make a substantial contribution to our knowledge of practice in digital preservation, and form the foundation for theory building and the development of methodological tools. The value of these case studies will come not only from the breadth of companies and institutions included, but also through the depth at which they will explore the issues.

ERPANET is deliberately and systematically approaching disparate companies and institutions from industry and business to facilitate discussion in areas that have traditionally been unconnected. With these case studies ERPANET will broaden the scope and understanding of digital preservation through research and discussion. The case studies will be published to improve the approaches and solutions being developed and to reduce the redundancy of effort. The interviews are identifying current practice not only in-depth within specific sectors, but also cross-sectorally: what can the publishing sector learn from the aeronautical sector? Eventually we aim to use this comparative data to produce intra-sectoral overviews.

This cross-sectoral fertilisation is a main focus of ERPANET as laid out in its Digital Preservation Charter.[2] It is of primary importance that disparate groups are given a mechanism through which to come together as best practices for digital preservation are established in each sector.

Aims

The principal aims of the study are to:

- build a picture of methods and match against context to produce best practices;
- accumulate and make accessible information about practices;
- identify issues for further research;
- enable cross-sectoral practice comparisons;
- enable the development of assessment tools;
- create material for training seminars and workshops; and,
- develop contacts.

Potential sectors have been selected to represent a wide scope of information production and digital preservation activity. Each sector may present a unique perspective on digital preservation. Organisational and sectoral requirements, awareness of digital preservation, resources available, and the nature of the digital

[2] The Charter is ERPANET's statement on the principles of digital preservation. It has been drafted in order to achieve a concerted and co-ordinated effort in the area of digital preservation by all organisations and individuals that have an interest and share these concerns. http://www.erpanet.org/charter.php.

© ERPANET 2004

object created place unique and specific demands on organisations. Each of the case studies is being balanced to ensure a range of institutional types, sizes, and locations.

The main areas of investigation included:

- perception and awareness of risk associated with information loss;
- understanding how digital preservation affects the organisation;
- identifying what actions have been taken to prevent data loss;
- the process of monitoring actions; and,
- mechanisms for determining future requirements.

Within each section, the questions were designed to bring organisational perceptions and practices into focus. Questions were aimed at understanding impressions held on digital preservation and the impact that it has had on the respective organisation, exploring the awareness in the sector of the issues and the importance that it was accorded, and how it affected organisational thinking. The participants were asked to describe, what in their views, were the main problems associated with digital preservation and what value information actually had in the sector. Through this the reasons for preserving information as well as the risks associated with not preserving it became clear.

The core of the questionnaire focused on the actions taken at corporate level and sectoral levels in order to uncover policies, strategies, and standards currently employed to tackle digital preservation concerns, including selection, preservation techniques, storage, access, and costs. Questions allowed participants to explore the future commitment from their organisation and sector to digital preservation activities, and where possible to relate their existing or planned activities to those being conducted in other organisations with which they might be familiar.

Three people within each organisation are targeted for each study. In reality this proved to be problematic. Even when organisations are identified and interviews timetabled, targets often withdrew just before we began the interview process. Some withdrew after seeing the data collection instrument, due in part to the time/effort involved, and others (we suspect) dropped out because they realised that the expertise was not available within their organisation to answer the questions. The perception of risks that might arise through contributing to these studies worried some organisations, particularly those from sectors where competitive advantage is imperative, or liability and litigation issues especially worrying. Non-disclosure agreements that stipulated that we would neither name an organisation nor disclose any information that would enable readers to identify them were used to reduce risks associated with contributing to this study. In some cases the risk was still deemed too great and organisations withdrew.

© ERPANET 2004

Chapter 3: Method of Working

Initial desk-based sectoral analysis provides ERPANET researchers with essential background knowledge. They then conduct the primary research by interview. In developing the interview instrument, the project directors and editors reviewed other projects that had used interviews to accumulate evidence on issues related to digital preservation. Among these the methodologies used in the Pittsburgh Project and InterPARES I for target selection and data collection were given special attention. The Pittsburgh approach was considered too narrow a focus and provided insufficient breadth to enable full sectoral comparisons. On the other hand, the InterPARES I data collection methodology proved much too detailed and lengthy, which we felt might become an obstacle at the point of interpretation of the data. Moreover, it focused closely on recordkeeping systems within organisations.

The ERPANET interview instrument takes account of the strengths and weaknesses from both, developing a more focused questionnaire designed to be targeted at a range of strategic points in the organisations under examination. The instrument[3] was created to explore three main areas of enquiry within an organisation: awareness of digital preservation and the issues surrounding it; digital preservation strategies (both in planning and in practice); and future requirements within the organisation for this field. Within these three themes, distinct layers of questions elicit a detailed discovery of the state of the entire digital preservation process within participants' institutions. Drawing on the experience that the partners of ERPANET have in this method of research, another important detail has been introduced. Within organisations, three categories of employee were identified for interview: an Information Systems or Technology Manager, Business Manager, and Archivist / Records Manager. In practice, this usually involved two members of staff with knowledge of the organisation's digital preservation activities, and a high level manager who provided an overview of business and organisational issues. This methodology has allowed us to discover the extent of knowledge and practice in organisations, to understand the roles of responsibility and problem ownership, and to appreciate where the drive towards digital preservation is initiated within organisations.

The task of selecting the sectors for the case studies and of identifying the respective companies to be studied is incumbent upon the management board. They compiled a first list of sectors at the very beginning of the project. But sector and company selection is an ongoing process, and the list is regularly updated and complemented. The Directors are assisted in this task by an advisory committee.[4]

[3] See http://www.erpanet.org/studies/index.php. We have posted the questionnaire to encourage comment and in the hope that other groups conducting similar research can use the ideas contained within it to foster comparability between different studies.
[4] See www.erpanet.org for the composition of this committee.

© ERPANET 2004

Chapter 4: the European Patent Office

In 1973 European countries endorsed the Convention on the Grant of European Patents (EPC) establishing a uniform patent system in Europe. The European Patent Office (EPO) took up work in 1978 in order to implement this agreement.[5] It is supervised by the European Patent Organisation, an intergovernmental body with delegates from the EPC contracting states. As of March 2003 there were 27 contracting states and bilateral agreements with a number of Central and Eastern European countries.[6] The headquarters of the EPO are in Munich, and it has a branch at The Hague and sub-offices in Berlin and Vienna. Currently 5,500 people are working at the EPO.

The mandate of the EPO is to grant patent protection for inventions. This protection lasts for twenty years in some or all of the contracting states. In 2002, the EPO received over 165,000 patent applications.[7] The EPO compares each patent application with inventions published world-wide since 1920 to see whether it is truly novel. On average the registration process for a European patent takes about 50 months from the moment the applicant files the application,[8] which is about twice as long as the procedure takes at the patent office in the United States of America. The reason for this considerably longer processing time is the organisational structure of the EPO. The EPO is an umbrella organisation to national patent offices in Europe. After a patent has been registered at and examined by the EPO, each EPO member country needs to endorse the patent individually, which is a lengthy process. Nevertheless, the EPO would like to reduce the average time for granting a European patent to 36 month in the future.[9]

In a global context, the World Intellectual Property Organization (WIPO)[10] concluded in 1970 the Patent Cooperation Treaty (PCT)[11]. This treaty allows the applicant to seek patent protection simultaneously at all PCT contracting states. While eventually each regional or national patent office individually examines and, if applicable, grants the patent, they cooperate and exchange information about the patentability of the claimed invention. Thus, the PCT simplifies a worldwide patent application for the applicant, and at the same time it makes the examination process for the patent offices more efficient. Over 100 countries participate in the PCT.[12]

Together with the Japan Patent Office (JPO)[13] and the United States Patent and Trademark Office (USPTO)[14] the EPO is one of the biggest patent granting bodies. Together they process about eighty percent of the world's patents. Since 1983 these three organisations are cooperating closely in a "Trilateral Cooperation" to improve efficiency by exchanging information regarding the information infrastructure and even sharing software. Infrequently, different requirements do present obstacles. These are

[5] European Patent Office (EPO); http://www.european-patent-office.org/.
[6] EPO website: general presentation of the EPO, 2003; http://www.european-patent-office.org/epo/pubs/brochure/general/pdf/general.pdf.
[7] EPO Annual Report 2002, Chapter "2002 in review".
[8] EPO Annual Report 2002, Chapter "Business report", subchapter "Grant procedure".
[9] EPO: Official Journal. July 2003. Subchapter "Extended European Search Report". http://www.european-patent-office.org/epo/pubs/oj003/07_03/07_3193.pdf.
[10] World Intellectual Property Organization (WIPO); http://www.wipo.org/.
[11] Patent Cooperation Treaty (PCT); http://www.wipo.int/pct/en/.
[12] WIPO, Patent Cooperation Treaty, Contracting States as of January 2004; webpage: http://www.wipo.int/pct/guide/en/gdvol1/annexes/annexa/ax_a.pdf.
[13] Japan Patent Office (JPO); http://www.jpo.go.jp/.
[14] United States Patent and Trademark Office (USPTO); http://www.uspto.gov/.

© ERPANET 2004

mostly due to legal and cultural variations as has been the case with the JPO. Additionally, the EPO cooperates with other patent offices such as the State Intellectual Property Office in China,[15] giving them access to knowledge, solutions and assisting them in setting up an adequate information infrastructure.

[15] State Intellectual Property Office of the People's Republic of China; http://www.cpo.cn.net/.

© ERPANET 2004

Chapter 5: Details and circumstances of the interviews

The EPO was very open and co-operative during this case study. Interviews were conducted at their Den Haag offices as the department of 'Information Systems' (IS) responsible for the management of digital information is located there. Staff interviewed included the Director of Procedures, Security and Inventory at the Department of Information Systems, the Director of Research and Development at the Documentation Department, and other members of staff that introduced us to the active use of the EPO's systems.[16] This report was written in November 2003.

[16] We are grateful to Mr. David Allin and Mr. Mark Krier for giving us their time and sharing their expertise.

© ERPANET 2004

Chapter 6: Analysis

This section presents an analysis of the data collected during the case study. It is organised to mirror the sequence of topics in the questionnaire.

- Perception and Awareness of Digital Preservation
- Preservation Activity
- Compliance Monitoring
- Digital Preservation Costs
- Future Outlook

Perception and Awareness of Digital Preservation

Asset value

All information at the EPO is available in digital form, and increasingly *only* in digital form.[17] The patents themselves are the most valuable asset at the Office. EPO examiners check the database of existing patents on a daily basis to ascertain whether a patent application is a novel invention and may therefore be granted. The management and preservation of patent information is a question of business survival to the EPO.

In order to judge whether an application is truly novel, it is not sufficient to check only the patents the EPO has granted in the past. All patents worldwide that have at any point in time been granted need to be examined. For this reason the EPO is working closely together with patent offices world-wide.

But more than that, EPO examiners also have to check 'Non-Patent Literature (NPL)'. If the invention has been introduced at a conference by a different person, the patent will not be granted either. For this reason the EPO requires access to all relevant scientific literature. However, engaging cooperation from scientific publishers is proving difficult. Not all of them allow the EPO to store their literature in the EPO's local database.

Regulatory Environment

The EPO is required by EPC Rule 95a[18] to retain patent files at least five years until after the patent grant has expired.[19] PCT Rule 93[20] demands the retention of files for at least 10 years after an applicant submitted a patent for an international registration, which is longer than that of the EPC Rule in case a patent application is rejected. Which of the two Rules applies, however, does not matter to the EPO, since it strives to preserve its information for an undetermined period.

[17] See 'Preservation Activity – Policies and Strategies' below for a deeper discussion.
[18] European Patent Convention (EPC); see more information on the EPC in the Introduction, or refer to the References Section for the EPC document.
[19] As mentioned in the Profile of the EPO above, a patent term runs 20 years from the date of application.
[20] Regulations Under the Patent Cooperation Treaty (PCT); see more information on the PCT in the Introduction, or refer to the References Section for the PCT Regulations.

© ERPANET 2004

Preservation Activity

Policies and Strategies

Digital preservation at the EPO is implemented as part of their stewardship of information. The Information Systems (IS) department's prime focus is improving efficiency and ease of use of patent information. For example, decisions to migrate digital objects to new formats are taken primarily to improve system efficiency; at the same time this action is designed in such a way that it supports preservation.

There is an intense urge towards the use of digital information in the patent world, foremost for its obvious advantages in search and retrieval functions, and for accelerating communication. In May 1998 the EPO decided the step-by-step implementation of a system called Phoenix[21] for "the creation, maintenance, preservation and inspection of files".[22] Since the year 2000 clients may submit patent applications to the EPO in digital form, which is called electronic filing.[23] Today digital versions are the primary reference. If any documents are received in paper format, they are digitised. While the paper is stored in a depot in Spain, it no longer serves any specific purpose.

This position is in line with international developments. The Patent Office in Japan was the first to introduce electronic filing in 1991. Also the WIPO has inaugurated a project to put electronic filing into place. This project called PCT-SAFE[24] is planned to finish in the end of 2003. This is of importance for the EPO due to possibly emerging WIPO standards and the prospect for closer cooperation between patent offices.

Close cooperation between the patent offices existed before the initiatives by WIPO as an umbrella organisation. Already in 1984 the creation of the "PCT minimum collection" was decided, underlining the commitment of international patent offices to move towards the use of digital documents. As part of this the patent offices of Japan, the USA, and Europe decided to digitise all their patents back to the year 1920. In a next step, some patent offices digitised even further back. In fact, the oldest patent was digitised in the US and dates back to 1836. This effort took well into the 1990s, and yielded 30 million digitised patents.[25]

The proliferation of patents more recently in addition to the rich history of patents has resulted in a huge and growing amount of information assets to be reliably preserved into the future. To achieve this, the EPO has high demands on the reliability of their systems and the quality of their information.[26] References to other scientific literature within a patent document establish the context and scientific background, so that patents can be understood in the future. Furthermore, each patent is classified into a topic category along with other metadata and documentation of the registration process.

---

[21] EPO, online Description of Phoenix; http://www.epoline.org/phoenix.html.
[22] Decision of the President of the European Patent Office dated 14 May 1998. http://www.european-patent-office.org/legal/anc_reg/e/pdf/1998s360.pdf.
[23] Decision of the President of the European Patent Office dated 7 December 2000. http://www.european-patent-office.org/epo/president/e/2000_12_07_e.htm.
[24] WIPO's PCT-SAFE (Secure Applications Filed Electronically); http://www.wipo.org/pct-safe/en/.
[25] In this collection the average size of a patent was ten pages; nowadays patents usually are considerably bigger.
[26] See 'Preservation methods and systems' below for a deeper discussion.

© ERPANET 2004

The responsibilities for all the necessary actions are clearly defined within the main departments: the Documentation department and the Information Systems (IS) Department. The former ensures the quality of the information assets and establishes their context. The latter has the responsibility for the information infrastructure and hence also for digital preservation. As there are no off-the-shelf patent systems, IS had to develop and implement most of the system themselves. A mix of outsourcing and home-made development proved most efficient to achieve this. In fact, the 400 staff of the IS department are divided equally between permanent and contracted staff; thereby the latter bring in vendor specific knowledge, whereas permanent staff cater for continuity in the overall system development.

The EPO offers comprehensive possibilities for training. Employment at the EPO starts with a six-week academy, and training continues for three years on a regular basis. After that further training is on a flexible basis. As part of this training, examiners are also taught how to use the systems for filing, searching and retrieving patent information as well as the Non-Patent Literature (NPL).

Selection

A file is established for each submitted patent application. This file is extended with relevant data throughout the examination process, and subsequently retained over the long-term. There are standard templates detailing the structure and appearance of a patent file. The applicant has to prepare the patent information accordingly prior to submission.

The preparation of the NPL follows guidelines as well; NPL is searched, annotated and accessioned by the EPO's Documentation department on a regular basis. Alternatively, where NPL cannot be imported into local systems, the external databases of the publishers need to be linked to the knowledge-base of the EPO. This is either done by technical assistance from the IS, or the examiners have to manually search the relevant literature that is external to the EPO systems.

For each patent there are about thirty to forty metadata fields. Metadata are created either by the applicant or by EPO staff. Those fields are mainly of a descriptive nature, and there are no specific metadata for the long-term preservation of the documents. Similarly, NPL has about the same volume of metadata, yet different fields, which are prepared by the Documentation department.

Standards for patents and NPL exist at an international level maintained by the WIPO, as well as an institutional level. Mostly those standards supplement each other. In some cases they demand dual tracks. For example, each patent is classified according to the International Patent Classification (IPC).[27] This international classification scheme, however, has difficulties in keeping pace with development, and does not have the granularity and dynamism for the requirements of the EPO. Additionally there are small differences in the approach of a classification scheme between the various patent offices. Specifically, in summer 2003 when this case study was conducted, the classification scheme of the EPO has 160,000 class entities, whereas the scheme by WIPO comprises almost 69,000 entities.

[27] WIPO, International Patent Classification (IPC); http://www.wipo.int/classifications/en/ipc/.

© ERPANET 2004

The EPO established a technical standard for electronic filing and other communication between applicants and the EPO.[28] This standard builds on encryption using a Public Key Infrastructure. By the mandatory use of encryption the EPO attempts to satisfy requirements for authentication, data integrity, non-repudiation as well as confidentiality. In the same vein PCT-SAFE[29] will rely on Public Key Infrastructure, employing digital certificates, digital signatures and data encryption.

While electronic filing has been in place at EPO since the year 2000, there are still paper applications coming in. These documents first need to be digitised and then converted to the form suitable for the EPO system. Digitisation is outsourced. The external contractor scans the paper, makes an automatic OCR[30], and subsequently converts the file to its required form.[31] Quality assurance is performed by the contractor, as well as by the EPO by means of a sample check according to ISO 2859[32]. (A statistically calculated number of samples are taken from a batch of digitised documents; the quality requirements for this sample as established by the EPO allow only 0.1 percent of defects, such as blurry images and cut edges.)

Preservation

The EPO applies a migration strategy for preserving its digital assets. All digital objects are in a few standard formats, forming a homogeneous and well-structured collection. Also for the encryption algorithms and digital signatures, a migration strategy will be applied. The huge holdings of the EPO necessitate the application of compression algorithms. Similarly, compression will be renewed before a compression algorithm becomes obsolete. For all their different formats and algorithms the EPO has an implicit technology watch: the objects in the homogeneous archive are used on a daily basis and only with tools provided by the EPO's IS department; therefore, staff at the IS department are always informed about the status of their holdings, and whether a specific format or algorithm needs to be migrated.

The EPO built a massive system to manage their information assets. This system is as much as possible based on standard components. The specific requirements, however, necessitate considerable input of in-house development work by the EPO. As of Summer 2003, EPO's holdings comprise 150 terabyte of data volume, and they expect to reach the petabyte[33] border by 2005. There are multiple backups of their digital holdings. Firstly, at the EPO branch in Rijswijk, near The Hague, are two mirrored data centres that are located in separate buildings in a distance of several hundred meters. Additionally, copies of the patent information are transferred to the patent offices in the US and Japan as part of their trilateral cooperation. Digital objects once captured into the system cannot be updated or deleted; only new versions can be added.

To allow for this rapid growth in data volume the EPO system is hugely extensible, not only concerning the pure storage space but also in performance since those data holdings need to be accessed by examiners on a daily basis. A three-tier system was

[28] EPO: Technical Standards for Online Exchange. 1 September 2000. http://www.european-patent-office.org/epoline/pdf/epo-olf-standard.pdf.
[29] an international standard for electronic filing currently being developed by WIPO; refer to Chapter "Management Strategies and Policy Building".
[30] OCR – Optical Character Recognition; http://www.wikipedia.org/wiki/Optical_character_recognition.
[31] The required form is discussed in detail in the next Chapter.
[32] ISO 2859 - Sampling procedures for inspection by attributes. ISO, International Organization for Standardization: http://www.iso.ch/.
[33] 1 petabyte = 1.000.000 gigabyte = 10^{15} bytes

© ERPANET 2004

developed together with IBM to accommodate the EPO's requirements.[34] The back-end is formed by a mainframe that houses the data. The mid-tier connects data storage to the client offering specialised services for search and retrieval. These services are employed by the client interface at the top layer. The client layer uses Java technology[35] for system independence. System independence was deemed important to allow the use of different systems at the client side at the same time; there are, in fact, currently workstations with the operating system OS/2 and others with Microsoft Windows that run the same Java application.

The different kinds of business information at the EPO, being foremost patent information and NPL, are all stored in the same system for optimal integration. Each object needs to be in a file format as required by the EPO. Patents in specific usually consist of written text and a number of pictures. The pictures have to be either in the format TIFF Group 4 or JPEG, which are linked together with the text in an SGML data container.[36] NPL is stored in SGML data containers, as well.

In 1996 the EPO turned to the storage of their data holdings in SGML,[37] and today the EPO is moving step by step to the use of XML. This is due to commercial pressure, since XML is now increasingly supported by the software industry. At the same time, WIPO now supports XML in its standards. Thus EPO is in the process of changing from SGML to XML.[38] The migration from SGML to XML is being undertaken in a phased manner. In this transitory phase the EPO is therefore employing SGML and XML in parallel.

Similar to the migration of SGML to XML, the EPO is now considering a change of image format. The currently employed image formats are compressed formats; TIFF Group 4 is a black-and-white TIFF derivate with a lossless compression algorithm as recommended by the CCITT.[39] Since this format was selected, image formats with better compression have become available. The EPO is therefore in the process of reviewing current image formats. A format that may be employed is the JPEG 2000 image format, which promises better compression; additionally it accommodates colour images and is to be an ISO international standard[40]. If, indeed, the EPO changes support of image formats, it has not yet been decided whether the existing images will be converted to current formats. Most possibly a transitional period can be expected, in which the old and the new image formats will be supported in parallel.

---

[34] See the IBM case study; see References.
[35] Java technology by Sun Microsystems; http://java.sun.com/
[36] EPO: Technical Standards for Online Exchange. 1 September 2000. http://www.european-patent-office.org/epoline/pdf/epo-olf-standard.pdf.
[37] P. Brewin et al.: EPO Publications I: Standards - WIPO ST.32. Recommendation for the Markup of Patent Documents using SGML. VERSION 3, 1996. (EPO Version 3.5, Nov. 1996). http://ebd.epoline.org/doc/st32.pdf.
[38] Paul Brewin (EPO) and Shiro Ankyu (WIPO): Filing and Processing Patent Data Using XML - A World Standard. XML Europe 2002, Conference. http://www.wipo.org/pct-safe/en/xml/previewTest.htm.
[39] CCITT - Consultative Committee on International Telegraphy and Telephony, Comité Consultatif International de la Télégraphie et du Téléphone; a body of the IUT (International Telecommunication Union, http://www.itu.int/.). The official definition of CCITT Group 4 is specified in recommendation T.6.
[40] in fact, JPEG 2000 Part 1 already is accepted as ISO standard ISO/IEC 15444-1; for closer information refer to http://www.jpeg.org/JPEG2000.html.

© ERPANET 2004

Access

The huge collections of patent information and NPL are used on a daily basis. Their size and the requirement to effectively search them puts enormous stress on the system. To cope with this, systems for search and retrieval have been geared towards the idiosyncrasies of patent information (as have been described above).

Three main user groups have to be served by the EPO: (1) applicants for patents; (2) the EPO examiners; and, (3) the general public. The EPO offers a suite of products to satisfy these requirements.

(1) The Epoline[41] gateway is a web accessible interface designed for communication with applicants, providing a secure environment that enables online filing, fee payment, file inspection, and communication between applicant and EPO. Communication and document exchange are well secured. A Public Key Infrastructure (PKI) for encryption was established to ensure the integrity and confidentiality of documents. Users anywhere in the world need to authenticate using Smart Cards. These Smart Cards contain the certificates and encryption keys at the basis of the PKI.[42]

(2) Efficient access to the patent information and the NPL is essential for the work of EPO examiners. While this is largely possible for the information maintained in-house, NPL from some publishers needs to be accessed via external databases.[43] The integration of this information from its various sources is a huge challenge, politically (in negotiating with the publishers of NPL) as well as technically. This results in complex interfaces for search and retrieval. The EPO's International Academy gives specific courses for examiners on how to use the information systems when searching and examining patent applications.[44]

(3) For the general public, access to approximately 400 million pages of published patents from all over the world is possible online via the "esp@cenet network".[45] About 80 percent of the world's technical information is published in patent information and, hence, esp@cenet provides a service that is very sought after. Apart from esp@cenet the EPO provides a number of other services for the public, such as patent information on CDs, delivering printed patent documents, and others.

Compliance Monitoring

The actions at the EPO are monitored along various lines. Firstly, every three years external consultants are appointed by the EPO. While these consultants mostly scrutinise financial aspects, their focus changes with each inspection. Internal monitoring is conducted by the department for Quality and Planning. Furthermore, employees have the possibility among other things, to report any problems they may have and make suggestions to adapt or streamline procedures. Overall monitoring activities appear to be successful along these lines. Especially communication between the departments as well as between staff and senior management is important to ensure the adequacy and efficiency of work at the EPO.

[41] The online *epoline*® gateway: http://www.epoline.org/.
[42] epoline Presentation. Helsinki, 9 June 2000. http://patent.prh.fi/patentline/2000-06/Epoline%20pres1.ppt.
[43] European Patent Office: Access to Non-Patent Literature Resources in the EPO. http://www.wipo.org/scit/en/meeting/6/presentations/npl_ep.pdf.
[44] EPO International Academy: http://www.european-patent-office.org/intcop/intl_academy/.
[45] esp@cenet network: http://ep.espacenet.com/.

© ERPANET 2004

Digital Preservation Costs

Generally speaking, the EPO depends on its information stocks to conduct its business in the first, and it does therefore not take any risks to lose any of this information. As a matter of consequence, the EPO will cover the costs necessary for digital preservation now and in the future.

It is virtually impossible to separate costs for digital preservation from other costs for the information infrastructure at the EPO, since all information remains in active use. Some insight can be gained, however, by viewing the budget of the department for Information Systems, which covers overall expenditures for the information infrastructure. This budget amounts to seventy to eighty million Euro per year, excluding the salaries for the 200 permanent members of staff. For comparison, the total expenditures of the EPO in 2002 were 1,308.4 million Euro; this results in an almost balanced operating result, as income from patent fees and the like amounted to 1,194.5 million Euro.[46]

Future Outlook

While the EPO is quite happy with what it has achieved to this point, improvement is always possible. Even with 200 permanent staff at the Information Systems department, technology and procedures are being designed and implemented gradually following a pragmatic agenda; especially considering the magnitude of the task to be tackled at the EPO.

The EPO recognises that NPL needs better integration with the other EPO systems. Currently it demands quite an effort to search the numerous databases that are mostly external to the EPO system. The optimal solution to this rather political and organisational issue would be a model contract that all publishers agreed to. Not all publishers, however, will be willing to transfer their scientific literature to the EPO systems. Instead, the EPO may need to extract the required information from the various external systems of the publishers. To this end, a meta-search facility that adapts to the diverse systems may be required. Cooperation of the publishers and, as part of this, the interoperability of systems would, of course, reduce technological complexity and expenses considerably.

Strengthening cooperation with other patent offices is an ongoing effort. This is done via the WIPO as well as directly. Most importantly, systems at the patent offices are not fully interoperable at this point of time. This means, for example, that patent descriptions need to be generated more than once when an application for an international patent is filed. Currently, a core patent description is being worked on as part of the PCT-SAFE initiative[47], which is the same for all patent offices and which may be supplemented with local information if necessary.

Work processes can also be further streamlined by electronic registration of patent applications. Currently patent applications are converted manually to XML if applicants submit them to EPO in another format. Recognising the inefficiency of this, the EPO is developing software tools that enable the applicant to automatically generate the appropriate XML. Those tools need to be provided for authoring programmes such as

[46] EPO: Operating result, balance sheet; Annual Report 2002. http://annual-report.european-patent-office.org/report/download/EPA_JB02_balance_sheet.pdf.
[47] WIPO's PCT-SAFE (Secure Applications Filed Electronically); http://www.wipo.org/pct-safe/en/.

© ERPANET 2004

Microsoft Word. At the same time the EPO aims to further promote electronic filing, so that digitising of paper documents can be avoided altogether.

Each of the above mentioned activities contributes to the long-term accessibility of the EPO's assets, including the interoperability of systems and electronic filing, which ensure the quality of the digital objects. In the same vein and triggered by the trilateral cooperation and newly issued WIPO requirements, the migration of the EPO's assets from SGML to XML is being furthered.[48]

[48] For this and other migration activities, see 'Preservation Activity', and 'Preservation methods and systems'.

© ERPANET 2004

Chapter 7: Conclusions

The EPO built an impressive system that incorporates long-term digital preservation as one of their core business needs. In focussing on the permanent accessibility of their digital assets and the efficiency of retrieval, the EPO takes a pragmatic stance towards preservation actions as exemplified in the following aspects of the EPO's approach:

A successful digital preservation programme does not require each employee to understand the corporate approach towards digital preservation in full detail. Employees at EPO acknowledge the pivotal role of the management of digital information, and they implement the necessary procedures as a natural part of their work.

Cooperation between patent offices offers huge potential for streamlining work processes and eventually saving costs. The trilateral cooperation with the USA and Japan allows faster work processes and more quality while saving significant costs. Also the transfer of EPO's technology to the Chinese Patent Office underlines this: While the EPO loses possible income by giving away technology that has cost them considerable time and money to develop, close cooperation with the Chinese Patent Office is deemed more important and more beneficial in the long term. After all, Chinese patents need to be included in the EPO's examination process. With a growing production of Chinese patents the interoperability of systems gains importance.

The patent system at the EPO is geared towards patents and has advanced functionality that cannot be bought in standard solutions at this point of time. Despite this the EPO system is not built from scratch, but it effectively uses existing solutions and combines them with own development work. This shows that even hugely dedicated systems can be built based on off-the-shelf components.

The EPO do not let encryption and compression obstruct digital preservation, incorporating them into their solution and devising a migration strategy for them. Additionally documentation about the applied algorithms and migration activities is retained.

The dynamic organisational structure of the EPO allows them to easily keep pace with technology development. The Information Systems department found their mix of permanent and contracted engineering staff to support efficient and advanced solution development. Contracted staff bring in new knowledge and ensures that implementation projects are implemented swiftly – engineers from a technology supplier surely have the most experience in implementing it – and permanent engineering staff put in the local knowledge and ensure coherence with the big picture of the EPO information infrastructure.

© ERPANET 2004

Appendix 1: References

The website of the *European Patent Office, EPO*, can be found at
http://www.european-patent-office.org/

European Patent Office, Annual Report 2002.
http://annual-report.european-patent-office.org/

European Patent Office, Official Journal; published yearly, available back to the year 1997.
http://www.european-patent-office.org/epo/pubs/oj_index_e.htm

Convention on the Grant of European Patents, *European Patent Convention, EPC*, of 5 October 1973, as last amended 10 December 1998.
http://www.european-patent-office.org/legal/epc/

The website of the *World Intellectual Property Organisation, WIPO*, can be found at
http://www.wipo.int/

WIPO, the Patent Cooperation Treaty, PCT.
http://www.wipo.int/pct/en/index.html

WIPO, Regulations Under the Patent Cooperation Treaty, PCT; as from January 1, 2003. http://www.wipo.org/pct/en/texts/rules/rtoc_short.htm

WIPO, Standards Relating to Patent Documents.
http://www.wipo.org/scit/en/standards/standards.htm#group3

WIPO, Electronic Filing Project: PCT-SAFE (Secure Applications Filed Electronically)
http://www.wipo.org/pct-safe/en/

Paul Brewin (EPO) and Shiro Ankyu (WIPO): Filing and Processing Patent Data Using XML - A World Standard. paper presented at XML Europe 2002 conference in Barcelona.
http://www.wipo.org/pct-safe/en/xml/previewTest.htm

OECD: Compendium of Patent Statistics. 2003.
http://www.oecd.org/dataoecd/60/24/8208325.pdf

IBM, An IDC e-business Case Study: European Patent Office - Moving to an Open Platform via IBM Technology and Java ....; 15 Jun 2001.
ftp://ftp.software.ibm.com/software/solutions/pdfs/g325-1846-00.pdf

Catherine Lavallée: Maîtriser les données - L'European Patent Office améliore ses performances en automatisant l'administration des données.
In: Infotel Actualités N° 33 mars 2003.
http://www.infotel.com/obj/IA-33.pdf

© ERPANET 2004

Appendix 2: List of acronyms used

CCITT	International Telegraph and Telephone Consultative Committee
EPC	European Patent Convention
EPO	European Patent Office
IPC	International Patent Classification
IS	Department of Information Systems at the European Patent Office
ISO	International Standards Organisation
JPO	Japan Patent Office
JPEG	Joint Photographic Experts Group; an image file format employing lossy compression
NPL	Non-Patent Literature
OCR	Optical Character Recognition
PCT	Patent Cooperation Treaty
PCT-SAFE	Secure Applications Filed Electronically
PKI	Public Key Infrastructure
SGML	Standard Generalized Markup Language (ISO 8879:1985)
TIFF	Tagged Image File Format
USPTO	United States Patent and Trademark Office
WIPO	World Intellectual Property Organisation
XML	Extensible Markup Language

© ERPANET 2004

erpanet

CONTACT DETAILS

ERPANET Coordinator
George Service House
11 University Gardens,
University of Glasgow
Glasgow, G12 8QQ,
Scotland

Tel: +44 141 330 4568
Fax: +44 141 330 3788
Coordinator@erpanet.org

ERPANET STAFF

directors
Seamus Ross, Principal Director
Niklaus Bütikofer, Co-Director
Mariella Guercio, Co-Director
Hans Hofman, Co-Director

coordinator
Peter McKinney

editors
Andreas Aschenbrenner
Georg Büchler
Joy Davidson
Prisca Giordani
Francesca Marini
Maureen Potter

www.erpanet.org