



Information Society  
Technologies

# erpastudies

**centraal bureau  
voor de statistiek  
(CBS)**

---

[www.erpanet.org](http://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 [Hhttp://www.erpanet.org](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 ([Hhttp://europa.eu.int](http://europa.eu.int)).

ISSN 1741-8682  
© ERPANET 2004

## **Table of Contents**

<b>Executive Summary .....</b>	<b>2</b>
<b>Chapter 1: The ERPANET Project.....</b>	<b>3</b>
<b>Chapter 2: Scope of the Case Studies .....</b>	<b>4</b>
<b>Chapter 3: Method of Working .....</b>	<b>6</b>
<b>Chapter 4: the Centraal Bureau voor de Statistiek (CBS) .....</b>	<b>7</b>
<b>Chapter 5: Details and circumstances of the interviews .....</b>	<b>9</b>
<b>Chapter 6: Analysis .....</b>	<b>10</b>
Perception and Awareness of Digital Preservation.....	10
Preservation Activity.....	13
Compliance Monitoring .....	16
Digital Preservation Costs.....	16
Future Outlook .....	17
<b>Chapter 7: Conclusions .....</b>	<b>18</b>
<b>Appendix 1 References.....</b>	<b>19</b>

## **Executive Summary**

The Centraal Bureau voor de Statistiek (CBS, also known in English as *Statistics Netherlands*) is the Dutch agency of national statistics, responsible for collecting, processing, and publishing statistics about the Netherlands for use by policymakers and in scientific research. Since the end of 2001, the CBS has been preparing to migrate from paper-based records and archives into an electronic record-keeping environment and to transfer the legal status of their official records from a paper-based format to digital. This will result ultimately in full electronic archives and records management, with completion expected by the end of 2004. The CBS is not legally required to preserve raw individual statistical data, but must keep published items that contain statistical data, as well as administrative records, for twenty years before permanent transfer into longer-term preservation by a national body. This study therefore focuses on the CBS's own efforts to preserve their administrative data and published documents containing statistical information, and *not* their statistical databanks.

Preparations for migration to a digital record-keeping environment and archive have been extensive, incorporating research into Records Management Applications (RMAs), technical and archival issues, a new archival policy, and pilot RMA implementation projects. Rather than simply transferring the traditional paper paradigm, the CBS has redesigned its documentary working processes for a new digital reality, successfully incorporating aspects of the electronic environment that were unsuited towards paper-based preservation (such as email preservation) directly into their approach. Other aspects, including preservation of electronic signatures and websites, have been postponed until the solutions or technical infrastructure are sufficiently developed. The careful and methodical approach followed at the CBS has resulted in a clearly defined project that considers not only the record-keeping department and archives but also the needs and wishes of the rest of the organisation, all of which must be clearly understood if the project is to be a success.

The CBS are beginning to preserve archival records in the context of a clear and well-defined record-keeping environment. They have drawn upon Dutch research programmes and information sources to develop their approach, complemented by international sources where necessary. Despite the fact that this is not yet reflected in the rest of the organisation, responsible staff show good levels of awareness about digital preservation – likely due to the extensive preparations that have been taking place. It is these preparations that form the bulk of this study, providing a solid example of an organisation that intends to manage preservation by explicitly incorporating it into active records management.

## **Chapter 1: The ERPANET Project**

The European Commission and Swiss Confederation funded ERPANET Project<sup>1</sup> (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.

---

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

## **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.<sup>2</sup> 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

---

<sup>2</sup> 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>.

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.

### **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<sup>3</sup> 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.<sup>4</sup>

---

<sup>3</sup> 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.

<sup>4</sup> See [www.erpanet.org](http://www.erpanet.org) for the composition of this committee.

## **Chapter 4: the Centraal Bureau voor de Statistiek (CBS)**

<http://www.cbs.nl/>

The Centraal Bureau voor de Statistiek (CBS, also known in English as *Statistics Netherlands*) is the Dutch agency of national statistics, responsible for collecting, processing, and publishing statistics about the Netherlands for use by policymakers and in scientific research. Founded over one hundred years ago, the CBS was up until recently an agency of the Ministry of Economic Affairs. In January 2004 it became an autonomous agency, although the Minister for Economic Affairs is still politically responsible for legislation and budget. The Ministry of Economic Affairs is represented at the CBS via an independent Central Commission for Statistics (CCS), which consists of 11 representatives from the Ministry who are responsible for a number of tasks in relation to the provision of statistical information for central government<sup>5</sup>. The CBS itself is structured into three statistical divisions, one technical division, and four central departments for organisational management that also incorporate the records management section and the Archives. In total, the CBS employs over 2500 people at two separate locations in the Netherlands: Voorburg (near Den Haag, in the province South Holland) and Heerlen (in the province of Limburg). Seventeen people are employed in the Archive.

The information the CBS collects and publishes covers many social and economic issues, ranging from macro-economic indicators such as economic growth and consumer prices, to the incomes of individual people and households. Publications and statistical data are hosted and maintained in the Statline<sup>6</sup> databank, which is used by all three statistical divisions. Statistical background and working material does not have to be preserved<sup>7</sup>, but preservation is required for any type of documents derived from and related to statistical data<sup>8</sup>. The CBS proposes which of the statistics shall be made into publications; after approval of this by the CCS, the CBS Director General oversees implementation of the selected statistics into publications, which are then disseminated in hard copy and via the website. Other information created and stored at the CBS includes policy documents, financial documents, contracts and correspondence.

In late 2001, the CBS established the VIDIV<sup>9</sup> records management project and began to prepare for transition from a paper record-keeping system and archive, to a digital one. This was considered necessary as the existing paper system was proving insufficient to meet the demands of an electronic production environment, and information was being lost. Much research has been carried out since then, and although the organisation is (at the time of writing) still relying officially on its paper archives, they expect to have their chosen Records Management Application (RMA) in place by the end of this year (2004) and to be relying on the digital records as the official versions shortly after.

---

<sup>5</sup> See <http://www.cbs.nl/nl/organisatie/ccs/ccs.htm> for more information on the activities and responsibilities of the CCS.

<sup>6</sup> Publications from the Statline databank are freely accessible over the Internet. See: <http://statline.cbs.nl/StatWeb/start.asp?LA=en&DM=SLEN&lp=Search/Search>.

<sup>7</sup> Although the statistical data does not have to be preserved, the CBS has recently shown interest in doing so. So, although such data is not a focus of the case study, it is likely that it will assume greater preservation significance in the future.

<sup>8</sup> As noted in the CBS Archiefbeleid/Archives policy (draft V10.2), following on from the Archives Law and Archives Decision of 1995.

<sup>9</sup> VIDIV stands for *Voortzetting Implementatie Documentaire Informatievoorziening*, which roughly translates in English as 'continuation of implementing records management'.

As the project is ongoing, the records management situation at the CBS is rapidly changing. This case study should therefore be considered as a snapshot, representative of the organisation during a transitional period and not as a permanent overview. Their records management and archival practices have developed even since the interviews, on which this case study is based, were held.

## **Chapter 5: Details and circumstances of the interviews**

ERPANET first approached the CBS' RMA project leader, Mr Jan van Leeuwen, to enquire about their participation in a case study in late 2003. Although initially under pressure with their RMA project, the CBS were interested and agreed to participate. The interviews took place at the CBS offices in Voorburg on March 16th 2004, attended by Mr Jan van Leeuwen, production co-ordinator of the CBS central archive Mr Jeffrey Pootjes, and the Head of Documentation Flow at the CBS, Mr Walter Grenda.

In addition to the interviews, Mr van Leeuwen kindly provided ERPANET with several relevant background documents concerning development of the RMA project, including archival policy, RMA metadata, and administration rules.

## **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

The CBS is *not* legally required to preserve raw statistical data (i.e. individual data fields that are independent from published data); only publications that are derived from or related to statistical data must be preserved.<sup>10</sup> It is also legally required to preserve administrative records. As will be seen below, the CBS has embarked on an digital archives and records management<sup>11</sup> project to ensure legal archiving obligations can still be met after transfer to a digital records and archiving environment. This study therefore focuses on the CBS's efforts to preserve their internal administrative records and end-product documents (i.e. publications). Preservation techniques for electronic signatures and websites are not yet considered developed enough by the CBS for inclusion in its project at this stage.

### **Perception and Awareness of Digital Preservation**

Respondents showed good awareness of the basic issues and problems surrounding records management and the preservation of digital information, but indicated that this was not representative of the rest of the organisation. A cited and notable exception to this was the Director of the Automation section, located within the Division of Technology and Facilities, who was reported to be in strong favour of the RMA project.

Respondents noted that according to their sources, there is little digital preservation activity in the international field of statistical offices upon which they can draw. There is some recognition of the problems involved, but activity is concentrated in the hands of just a few countries such as the UK, Canada, and the Netherlands. Most other countries do not display any advanced recognition of the problems, nor are there any off-the-shelf market solutions upon which the CBS feels it can rely. The records management (RM) staff at the CBS are keen to gather more experience with preserving and maintaining digital information, as they recognise that the more experience they have, the more their knowledge of the issues will increase. Staff are guided and educated on digital preservation issues from Dutch laws and regulations,<sup>12</sup> advice and information originating from the National Archives (NA)<sup>13</sup> and ICTU,<sup>14</sup> and technical documents released by the CBS Automation section.

---

<sup>10</sup> 'End product published documents' includes websites, see later in this section for comments on CBS website preservation.

<sup>11</sup> For a comprehensive definition of records management, refer to the ISO 15489 standard, available from the ISO website. See <http://www.iso.ch/iso/en/commcentre/pressreleases/2002/Ref814.html> for further information.

<sup>12</sup> See *Regulatory Environment* for further details.

<sup>13</sup> See <http://www.en.nationaalarchief.nl/archiefbeheer/archiefzorg/>.

<sup>14</sup> ICTU is the Dutch organisation for information and communication technology in the public sector. Its aim is to contribute to the development of the structure of Dutch electronic

### *The main problems*

The CBS is facing two main problems in the transfer to a digital records management program. The first concerns organisational change. The RM department and the VIDIV project leader know that collaboration across the different departments and divisions of the CBS is essential if the project is to be a success. One of the biggest challenges they face is the sheer scale of their operation – with over 2500 employees, migration to a digital environment will have a significant impact on many employees' working practices, despite the fact that not all members of the organisation will ultimately use the RMA and archive, but only about 20 to 25 percent. The traditional role of the RM staff will also change, not only in terms of responsibilities and ways of working, but also regarding the level of support and advice they must provide to their colleagues for whom archiving and records management will become a new aspect of their job.<sup>15</sup> Awareness of this can be seen in the extensive preparations that began back in 2001, that incorporate not only research into the CBS's RMA requirements but also pilot projects, reports by external consultants, and scheduled training sessions. Proper preparation is considered key to the projects success.

Secondly, once the RMA has been implemented and accepted, there is the issue of technical solutions. Dutch archival regulations already specify required formats for documents to be transferred to the NA.<sup>16</sup> Additionally, the Dutch government established a research project<sup>17</sup> in 2000 to identify long-term preservation approaches for four types of electronic records. However, although extremely useful, the CBS does not consider this to be fully crystallised as it addresses only four different record-types, and is of the opinion that not all of the required technical developments are yet complete.<sup>18</sup> This puts it, from its perspective, in an intermediate stage, with only part of the problem sufficiently addressed. The CBS is particularly concerned about the wide range of original file formats and believes this could cause problems in the future if not addressed early on.

### *Asset value and risk exposure*

The CBS collects a great deal of data, but is not legally required to preserve it all. Statistical information in published documents must be preserved, such as that found in paper and digital publications, publications originating from the Statline database, and the website of the CBS. These document types form the majority of their published records. Furthermore, their administrative records must be preserved, but statistical background and working material does not have to be preserved. These decisions are in accordance with the definition of an archive file in the Archives law of 1995 and the Archival decree of the same year.<sup>19</sup> The retention schedule for the CBS is thus

---

government, and it houses and supports several projects and programmes in support of this. See <http://www.ictu.nl/>.

<sup>15</sup> As noted in the Introduction of the CBS internal document 'Administration rules for the documentary provision of information at the CBS' ('*beheersregels*') (document undated; see *References*).

<sup>16</sup> These are specified in the Regulation on the 'Arrangement and Accessibility of Archival Records' (see below).

<sup>17</sup> Digital Preservation Testbed/Testbed Digitale Bewaring. Research covered e-mails, text documents, spreadsheets and databases, but not images or multi-media data. See <http://www.digitaleduurzaamheid.nl/>.

<sup>18</sup> The gaps are particularly in the widespread availability of XML products suited towards the CBS's requirements.

<sup>19</sup> See '*Regulatory Environment*' for further details of these regulations. For a full description of an archival file, refer to the Archiefwet/Archives law, accessible at [http://www.en.nationaalarchief.nl/images/4\\_2554.pdf](http://www.en.nationaalarchief.nl/images/4_2554.pdf).

comprised of administrative documents, which must be retained for purposes of business accountability<sup>20</sup> and archival value, and end product documents containing or derived from statistical data, i.e. publications. Publications are disseminated in hard copy and via the website; however, preserving the website is out of scope of the project until the CBS has installed a website Content Management System (CMS).<sup>21</sup>

The risks of not preserving CBS information over the long-term are legal, financial, and historical. None of the information has been officially assessed in light of a business needs or risk analysis; however, an internal and informal investigation was carried out at the commencement of the RMA project in 2001 on the consequences of losing digital information. This resulted in the continuation of the project and selection of an RMA for implementation at a later date.

More significantly, the CBS is also aware of the risks it runs if it does not migrate to a digital records management environment. Staff realised some years ago that the paper-based system they were using was no longer compatible with digital working practices, and that information was being lost. Change was therefore essential in order to minimise the risks of losing digital information, which are of a different nature for the CBS than the primarily operational business risks faced by commercial companies.

### *Regulatory Environment*

As a government agency, the CBS must comply with archival regulations for the public sector, as well as more general national legislation. The legal archival context of the CBS is comprised of:

1. The Archives Law<sup>22</sup> and the Archival Decree<sup>23</sup> of 1995; regulation on the Arrangement and Accessibility of Archival Records<sup>24</sup> (2002), and; regulation on the Longevity of Archival Records<sup>25</sup> (1999, focused on the medium/record carrier). These pieces of legislation concern the preservation of archival records by government bodies.

Furthermore, they must comply with other laws for government bodies, pertaining not specifically to archival records but oriented around security and data protection:

2. The Personal Data Protection Act and the Freedom of Information Act.<sup>26</sup>
3. The Information Security regulation for Central Government.<sup>27</sup>
4. The General Act on Public Administration.<sup>28</sup>

These are all clearly identified in the draft administration rules for the forthcoming RMA. In addition to this, the CBS has based these administration rules on the principles of the ISO 15489 standard for records management.<sup>29</sup>

---

<sup>20</sup> Accountability is very important to the CBS, which only recently gained autonomous status.

<sup>21</sup> As indicated in the CBS internal document, *VIDIV: Besluiten en hun argumentatie*, February 2004.

<sup>22</sup> In Dutch, this is the *Archiefwet*. See [http://www.en.nationaalarchief.nl/images/4\\_2554.pdf](http://www.en.nationaalarchief.nl/images/4_2554.pdf).

<sup>23</sup> In Dutch, this is the *Archiefbesluit*. See [http://www.en.nationaalarchief.nl/images/4\\_2559.pdf](http://www.en.nationaalarchief.nl/images/4_2559.pdf).

<sup>24</sup> In Dutch, this is the *Regeling geordende en toegankelijke staat archiefbescheiden*. See [http://www.en.nationaalarchief.nl/images/4\\_2563.pdf](http://www.en.nationaalarchief.nl/images/4_2563.pdf).

<sup>25</sup> In Dutch, this is the *Regeling duurzaamheid archiefbescheiden*. See: [http://www.en.nationaalarchief.nl/images/4\\_2562.pdf](http://www.en.nationaalarchief.nl/images/4_2562.pdf).

<sup>26</sup> In Dutch, *de wet bescherming persoonsgegevens en de wet openbaarheid van bestuur*.

<sup>27</sup> In Dutch, *het voorschrift informatiebeveiliging rijksoverheid*.

<sup>28</sup> In Dutch, the *Algemene wet bestuursrecht*.

## **Preservation Activity**

Preservation activity at the CBS is still oriented around paper records. However, the RMA project has been in place since 2001 and preparations for the migration to a reliable digital environment have been extensive. This section will therefore concentrate not on actual preservation activity at the CBS, but on the preparations that have been made and the intended implementation of the digital records management system. This is the place in which records will be preserved for as long as they are under the curation of the CBS.

The CBS has drawn on research from the Dutch public sector in designing their approach. Specifically, this incorporates the Digital Longevity Programme (comprising the Longevity Taskforce and the Preservation Testbed), and the Dutch electronic government knowledge centre based at ICTU. These are all collaborative efforts funded and supported by several Dutch ministries.<sup>30</sup> The CBS also carried out an Internet search on records management and records management applications at the beginning of the project, covering government, archives, and suppliers. The issue of digital preservation was raised as a specific issue in the search results, via the Digital Preservation Testbed and the legislation on the longevity of archival records.<sup>31</sup>

### *Policies and Strategies*

Policies and strategies for preserving digital records are being formulated within the context of the RMA project, VIDIV. There are not yet any approaches in place that relate specifically to the preservation of digital information, but the policy and strategy to be used with the RMA is in preparation. This is taking place with great consideration for the rest of the organisation, as staff in the records management section are aware that they require the support of the users, particularly so-called 'super-users', who will support their ideas and encourage the other employees. Learning and encouragement are considered essential to the successful take-up of the records management system, which will be used by between four to six hundred people.

Responsibility for establishing policies and strategies lies ultimately with the Head of Facilities Management, the department in which the RM section is located. The section intends to define an archival policy that incorporates specific directives on email use and maintenance, and is preparing the necessary policies for implementation and procedures concerning the new system. When this has undergone departmental consultation and is approved by a steering committee, then it will be put to the Director General for ultimate approval. Staff recognise the importance of creating a policy that is acceptable and understandable across the whole organisation, so careful creation of these policies will play a large role in ensuring their implementation. Their new archival policy caters for both paper and digital documents, identifies archiving responsibilities and powers, and provides further advice on identifying an archival file (as defined in the Archival law of 1995).<sup>32</sup> It also includes guidelines on simultaneous retention of paper records during the proposed trial period of relying on digital records for legal purposes (scheduled for the first year of operations), discusses non-archival material and how to deal with it, and relates this to retention periods specified in related documents (see below). Additionally, the administration rules ('beheersregels') for records management at the CBS contain extensive guidance and information on actual use of the RM system. Staff are aware that these rules will need to be periodically updated in order to

---

<sup>29</sup> See note 11 above for further information.

<sup>30</sup> Interlab is now supported by all Dutch Ministries, having been originally founded by just six.

<sup>31</sup> Regulation on the Longevity of Archival records, op.cit.

<sup>32</sup> Draft Archival policy/Archiefbeleid, v10.2, February 2004 (see *References*).

adjust to new developments and advancing insights of the CBS, and this is also noted in the administration rules themselves.

The strategy the CBS intends to employ is relatively straightforward. Given that the raw statistical data do not require preservation, the types of digital information that the CBS must preserve consist mainly of office files, particularly text documents and email messages, and end-product documents containing or derived from statistical data. It has defined an approach that is consistent with Dutch legislation and, although oriented primarily around active management of records, is simultaneously catering for their preservation over a period of up to twenty years before transfer to a national establishment for permanent preservation.

### *Selection*

The CBS has a retention schedule (known within Dutch agencies as a Basis Selectie Document or BSD) in place that defines which records are to be kept and for how long, and which are to be destroyed and when. The retention schedules used by Dutch government bodies have since 1991 been devised using an institutional research method developed in the PIVOT<sup>33</sup> project.

The PIVOT project developed a method of functional appraisal. The method involves a description of the actions of government bodies in a particular area. This makes the context of records clear, and the results are recorded in an Institutional Research Report (RIO). The RIO forms the basis of a Basic Selection Document (BSD), which is a retention schedule based on the identified actions in a particular policy field. In the BSD, actions from the RIO are assigned either a B (Bewaar – Keep) or V (Vernietigen – Destroy) status code. Actions with status code B are accompanied by the relevant appraisal criteria; actions with status code V are allocated destruction times.

The retention schedules for the CBS will be implemented in the chosen records management application and will apply across the whole organisation. As with the archive and RM policy, responsibility for maintenance and implementation of these schedules lies with the Head of Facilities Management.

### *Preservation*

Records will be preserved within the RMA. On the basis of the research carried out in 2001/2, including definition of the administration rules and the new archives policy, the CBS selected the Canadian system 'Hummingbird'<sup>34</sup> as its chosen RMA, and a private consultancy firm to assist in implementing it. Hummingbird was selected particularly for its compatibility with the DOD 5015.2 STD<sup>35</sup> and the MOREQ<sup>36</sup> requirements for

---

<sup>33</sup> PIVOT stands for Transfer Delay Reduction project. In the late 80's and early 90's, the Dutch government decreased the amount of time between organisational creation of records and their later transfer to the National Archives from fifty to twenty years. This required a new method of appraisal, as the existing method was too slow to deal with the massive increase in transferred records that would result from the new legislation.

<sup>34</sup> See the Hummingbird Enterprise website:

<http://www.hummingbird.com/products/enterprise/index.html>.

<sup>35</sup> DOD 5015.2 is a records management application (RMA) design criteria standard from the US Department of Defence (1997). It defines the basic requirements based on operational, legislative, and legal needs that must be met by RMA's acquired by the Department of Defence (DoD) and is widely accepted by record keepers outside of the DoD. See <http://jitic.fhu.disa.mil/recmgt/standards.htm> for further information.

<sup>36</sup> MOREQ is a set of MOdel REquirements for the management of electronic records (2001). It was developed from funding from the Interchange of Data between Administrations (IDA)

records management applications. Hummingbird has components for document management and records management, and also has a built-in workflow process. This was of particular interest to the statistical divisions as it offered significantly more control over the records than the previous paper-based approach. A pilot project using this RMA took place in 2003, lasting six months (three months in preparation, and three months in practice). This involved thirty-five people, half of whom were from the archive and the other half from the rest of the organisation. These people represented the real-time users of the archive in the future and were selected from three different units – statistics, personnel, and policy. The results of the pilot project led to redefinition of some aspects of the process – in particular the way in which keywords were used – and played an important part in developing the implementation plan.

Implementation of the RMA will be staggered. A number of Management Assistants have been identified, who will act as 'super-users' during the first term of the project. During this first term, other users must go through the Management Assistants to obtain or store documents in the system.

Text documents in the RMA will be preserved in Adobe's Portable Document Format (PDF<sup>37</sup>). PDF is one of the formats recommended by the Digital Preservation Testbed for the long-term preservation of text documents, and is also identified as a suitable format in the Ministerial regulation on the Arrangement and Accessibility of Archival Records.<sup>38</sup> Guidelines on migration settings and conversion procedures are also identified in the Testbed recommendations, based on those designed for the National Archives' 20th century digital depot.<sup>39</sup> The preservation strategy for emails has not yet been defined, although it is likely that it will involve XML. Paper correspondence will be scanned in TIF format, then converted to PDF with Optical Character Recognition (OCR) and entered into the RMA.

Metadata for the RMA is already defined. The CBS metadata set is an amalgamation of elements from the UK e-government metadata standard, items arising from the CBS administration rules, comments from the pilot project team meeting, the final evaluation of the pilot project, and some elements intrinsic to the Hummingbird system. Most of the metadata concern active management of the resource, although some elements contain technical details that can also be used for preservation (i.e., application and record-type). The CBS intend to automate approximately one-third of the fields to begin, with the majority of the remainder selected from drop-down boxes. Only a small number of the fields will require manual textual entries, keeping the potential for human error (for example, through spelling mistakes) down to a minimum.

---

programme of the European Commission. The MOREQ specification has been adapted in Dutch to form the 'Software Specifications for Records Management Applications for the Dutch Government', known as REMANO. For further information on MOREQ, see <http://www.cornwell.co.uk/moreq/>. For the REMANO specification, see: [http://www.archiefschool.nl/docs/ReMANO\\_2004.pdf](http://www.archiefschool.nl/docs/ReMANO_2004.pdf) (Dutch language only).

<sup>37</sup> For further information on Adobe and PDF, see

<http://www.adobe.com/products/acrobat/adobe.pdf.html>. Although this format is a standard, it is also under continuous development. Seven versions have been developed, or are in the development process, during the last twelve years. Adobe offers backward compatibility, but this will become more difficult to maintain the more format changes are experienced.

<sup>38</sup> Regulation on the Arrangement and Accessibility of Archival Records, op. cit.

<sup>39</sup> See

[http://www.en.nationaalarchief.nl/archiefbeheer/digitalisering/projecten/digitaal\\_depot.asp?ComponentID=2297&SourcePageID=3260#1](http://www.en.nationaalarchief.nl/archiefbeheer/digitalisering/projecten/digitaal_depot.asp?ComponentID=2297&SourcePageID=3260#1).

Technical storage of the archived information has not yet been discussed in great detail but will be the responsibility of the automation department, sub-department Infrastructure and Communication (TIC).

The project has been purposefully restricted for its first year of operation. Digital signatures will not be archived or preserved during the first year of using the RMA as they are not used extensively at the CBS. It is too large a financial investment for such an infrequently used technology at the moment, but will be re-considered in the following year. The website is also being considered for archiving as it contains unique publications, but this must wait until the CBS's Content Management System (CMS) project has been finalised and the web-pages have been standardised.

For the first year of use, paper originals will be kept in storage. This is to ensure the persistent accountability of the organisation should the project cause or encounter any unexpected problems. During this period, the paper versions will be the official records insofar as legal obligations go, and the digital ones used only for access. Once the transitional period is over, an investigation will take place on the project so far, and on exchanging the official status of the paper and digital documents. The results of this will define the next steps.

### **Access**

As implementation is to be staggered, the ways in which the records can be accessed will vary. During the first term, management assistants will co-ordinate much of the work streaming through the RMA, minimising the extent to which other users must access the system. This will enable the super-users to assume much of the initial responsibility, increasing their own familiarity and skills with the system, encouraging other users to accept the system, and minimising the initial scale of change that users would otherwise experience. Training is a vital aspect of using the system, and all users will receive training. The levels of training will vary, depending on the requirements of the different users.

Access protocols in the Hummingbird system will be enabled either as group-based permissions, for individual persons, or on individual documents. The exact method to be implemented has yet to be decided, but it is likely that group-based permissions will be used, and this will undergo a period of evaluation.

The system will be linked across the two locations (Voorburg and Heerlen), with both parties able to access the same set of data. The remainder of the access issues, such as whether to give users direct access to the stored digital information, are yet to be worked out.

### **Compliance Monitoring**

The CBS has not planned for compliance monitoring of their preservation actions. They will be audited for compliance to the data protection regulation for central government, as are all government organisations.

### **Digital Preservation Costs**

The CBS has identified the cost benefits of digital records management and archiving as threefold: firstly, the records management can become an integral part of the automated working processes of the organisation; secondly, a decrease in the use of paper and increase in the management of digital records enables better sharing of documents and fewer localised collections of records; and thirdly, digital records

management and archiving allows for organised maintenance of the organisational historical memory. There is no separate funding available for digital preservation activities, and the budget of the IT department is expected to cater for ongoing maintenance of the records. The VIDIV project budget, on the other hand, is part of the digitisation budget, which is allocated from the Infrastructure and Communication budget.

Respondents indicated a low priority amongst management to consider subjects for and over the long-term, and that this was consistent throughout the organisation. Of our interviewees, the management representative was fairly satisfied with the amount of money currently available for records management. The project leader was also satisfied, although he expected the availability of finances to increase as the service proves itself. The archives representative would prefer to see more money invested at the beginning of the project as the limited availability of funds means that the number of licences available at the beginning of the project is limited, which restricts the number of people gaining experience with the system.

### **Future Outlook**

The CBS has invested a great deal of time and effort in preparing their approach. The final stage will take place in late 2004. Consultants will be employed to prepare for implementation of the system, which is expected to begin in July. As noted earlier, implementation will be staggered, and they hope to have everything in place by the end of the year. Implementation preparation will include a number of training courses. RM staff will embark on a four-day training course, and other staff will receive training appropriate to their expected use of the system. A training procedure for new employees will also be established. The department is aware that measures will need to be continued after implementation has finished, but knows that this will require further financial support.

Respondents were satisfied with the direction of their preservation approach so far but are unsure of how long it will last, especially as elements of the national approach remain to be solved (as noted in *Main Problems*). They are aware that they still have issues to tackle, but are taking it one step at a time. They are fairly confident that investment in the records management programme will increase, but not necessarily explicitly in preservation. All in all, they are fairly satisfied with the availability of information about digital records management and preservation for government agencies but are aware that facilities for longer-term preservation of the records after their curation at the CBS have yet to be established.

## **Chapter 7: Conclusions**

The CBS have embarked on a records management project incorporating the entire records continuum, from creation to destruction or transfer to the Archives. They have drawn on a variety of sources, most of which are inside the Netherlands and related to either Archives or the public sector, to develop an extensive programme to ensure that implementation will be successful. The programme involves staff from across the organisation to ensure that it is suitable for the organisation as a whole, and stands a very high chance of success if it continues in the same careful and logical manner by which it has so far been characterised.

Preservation is explicitly considered in conjunction with active records management, and is not viewed as a separate function. Such explicit consideration is a positive aspect of their approach. Many organisations consider preservation as implicit in records management, and are unaware of the challenges preservation can pose. This can lead to problems in the future as unexpected issues concerning formats, metadata, and obsolescence become apparent. The CBS, however, is explicitly considering preservation, and not simply taking it for granted. This means that staff have given thought to many of the issues they are likely to experience in the future and are thus unlikely to be taken by surprise. The absence of explicit financial support for preservation may, however, prove an issue at a later date, and relying solely on Management Assistants (MA's) to manipulate the records for other users also incorporates some elements of risk. Users may not bother to liaise with the MA's unless they see a direct benefit from using them and the RMA. The benefits of doing so will therefore have to be emphasised with high visibility for all users, to minimise the chances of this happening.

Although respondents indicated that there was little digital preservation awareness within the organisation as a whole, the awareness they themselves showed was extensive, and they are confident that they grasp a good deal of the practical aspects they must deal with. They are aware that compliance to metadata standards is important and therefore took pains to ensure that the relevant standards were accounted for when designing their metadata schema. The Interlab project is currently attempting to develop a national metadata set for public bodies that the CBS will also be able to draw upon. In addition, future versions of PDF should be monitored to ensure that when conversion takes place, the CBS convert to the most up-to-date version of the format. This is particularly so with regards to PDF-A, which is being developed specifically for the purpose of archiving.<sup>40</sup>

All in all, the CBS approach is a fine example of an organisation who is effectively preparing for digital preservation of archival records.

---

<sup>40</sup> PDF-A is a collaborative effort in which Adobe are involved. The intention is to make it an ISO standard, number 19005. See <http://www.iso.ch/iso/en/stdsdevelopment/techprog/workprog/TechnicalProgrammeProjectDetailPage.TechnicalProgrammeProjectDetail?csnumber=38920> for further details. The PDF/A format is discussed in the recommendations of the Digital Preservation Testbed, op. cit.; see also the proceedings of the ERPANET Vienna seminar at <http://www.erpanet.org/events/2004/vienna/>.

## **Appendix 1 References**

### **Publicly available websites and documents:**

Digital Preservation Testbed, Digital Longevity Taskforce, and Interlab website:

<http://www.digitaleduurzaamheid.nl/>

National Archives of the Netherlands website:

<http://www.nationaalarchief.nl/>

(See

[http://www.en.nationaalarchief.nl/archiefbeheer/archiefzorg/archiefwet/wet\\_en\\_regelgeving/default.asp](http://www.en.nationaalarchief.nl/archiefbeheer/archiefzorg/archiefwet/wet_en_regelgeving/default.asp) for links to documents on archival legislation)

Centraal Bureau voor de Statistiek website:

<http://www.cbs.nl/>

E-Government Knowledge Centre website:

<http://www.elo.nl/>

ICTU website:

<http://www.ictu.nl/>

Hummingbird website:

<http://www.hummingbird.com/>

DOD 5015.2 Design criteria:

[http://www.interpares.org/display\\_file.cfm?doc=DoD\\_50152.pdf](http://www.interpares.org/display_file.cfm?doc=DoD_50152.pdf)

MOREQ website:

<http://www.cornwell.co.uk/moreq/>

REMANO specifications:

<http://europa.eu.int/ISPO/ida/export/files/nl/1492.pdf>

PDF-A Standard development websites:

<http://www.aiim.org/standards.asp?ID=25013> and

<http://www.iso.ch/iso/en/stdsdevelopment/techprog/workprog/TechnicalProgrammeProjectDetailPage.TechnicalProgrammeProjectDetail?csnumber=38920>

### **Internal documents received from the CBS:**

Archives Policy of the CBS (*'Archiefbeleid van het centraal Bureau voor de Statistiek'*), V10.2, draft. 207th February 2004.

VIDIV Decision and Argument (*'Besluiten en hun argumentatie'*), V1.0, draft. 27th February 2004.

Administration rules for records management at the CBS (*'Beheersregels voor de documentaire informatievoorziening bij het CBS'*), V1.0, draft (undated)

Metadata for VIDIV , V1.0 (undated)

**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](http://www.erpanet.org)