erpa

Information Society Technologies

erpastudies

koninklijk nederlands meteorologisch instituut

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

ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK

Table of Contents

Executive Summary

The Royal Dutch Meteorological Institute (KNMI) is the Dutch national centre for weather forecasting and climate research. The Meteorological sector thrives in a culture of international cooperation and open exchange of information since its formation in 1853, which provides a fertile ground for preservation activities. Various cooperative activities - including international standards and codes of practice, discussion groups specifically on data management and digital preservation, as well as international data archives - project a positive future for preserving observational data in the meteorological sector and at the KNMI.

Meteorological services create a vast amount of observational data. This is true not only of digital data collected today, but also of data collected in the sector's 150 years of history. The enormous value of these data are far from lost on the sector and the KNMI. The KNMI operates a network of observation stations in the Netherlands and is responsible for maintaining the data created by them. Data reliability is of great concern and therefore the KNMI transfers the original scientific data together with exhaustive metadata directly into its archives for preservation. Their advanced position with regard to digital preservation is most notably exhibited through the KNMI archiving policy, which synthesizes their experience in digital preservation issues.

Even before the creation of the KNMI archiving policy, the KNMI implemented many measures conducive to preservation in order to satisfy their high data management requirements.[1] Since the introduction of the policy, preservation actions have become tightly integrated with data management and the KNMI's business environment. Indeed, their actions are implemented from the point of data creation. Another outstanding feature of the KNMI preservation approach is its emphasis on organisational aspects, with their coordination group ensuring involvement of the entire organisation.

[1] Such high standards are indicative for the entire meteorological sector.

Chapter 1: The ERPANET Project

The European Commission and Swiss Confederation funded ERPANET Project[2] (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.

[2] ERPANET is a European Commission funded project (IST-2001-32706). See 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.[3] 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

[3] 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/www/content/documents/Digitalpreservationcharterv4_1.pdf.

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[4] 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.[5]

[4] See www.erpanet.org. 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.
[5] See www.erpanet.org for the composition of this committee.

Chapter 4: Introduction to Sector

National Meteorological Services such as the KNMI have an important and leading role in the international exchange of meteorological data. The Meteorological sector produces a wide range of meteorological and climatological products, services and advice to meet the needs of government, businesses and citizens. In addition, meteorological services advance the technical competence and quality of applied research in meteorology and climatology.

The roots of meteorological activities world-wide date back to 1853, when a 'Maritime Conference' was convened in Brussels.[6] This key conference was the first international meteorological conference and triggered a number of activities fundamental for the meteorological sector and international cooperation in this area: national meteorological services were founded; measurement stations were built; voluntary observing ships and other observing agents were involved; a uniform format for observational data was internationally accepted; and the International Meteorological Organisation, the precursor of the World Meteorological Organisation (WMO)[7], was formed in 1873.

The WMO coordinates global scientific activity to allow increasingly prompt and accurate weather information and other services for public, private and commercial use, such as international airline and shipping industries. The WMO's major scientific and technical programmes include the World Weather Watch (WWW).[8] This is the backbone of the WMO's activities and offers real-time world-wide weather information through member operated observation systems. Part of this are: telecommunication links with four polar-orbiting and five geostationary satellites, about 10,000 land observation stations, 7,000 ship stations and 300 moored and drifting buoys carrying automatic weather stations.[9] The collection of this wealth of observational data, which also includes data received from voluntary observing ships[10] and other contributors is only possible through the close cooperation of meteorological centres world-wide. The constant supply of timely data is pooled in various international data centres[11], such as the Global Observing System[12] for the WWW. It is through the WMO that the tight international cooperation in the meteorological and climatological sector with its complex agreements on standards, codes, measurements and communications are established internationally.

[6] Godwin Obasi, 'International Cooperation in Meteorology', Statement presented at the First Student Conference of the American Meteorological Society (AMS) (January 2002). http://www.wmo.ch/files/sg_statements/english/SG177E.pdf.
[7] World Meteorological Organisation: http://www.wmo.ch.
[8] World Weather Watch, WWW. http://www.wmo.ch/web/www/www.html.
[9] World Meteorological Organisation: http://www.wmo.ch/index-en.html.
[10] See the joint WMO/IOC Commission for Oceanography and Marine Meteorology (JCOMM), 'Voluntary Observing Ships (VOS)'. http://www.bom.gov.au/jcomm/vos/vos.html.
[11] The World Data Center for Meteorology in Asheville (http://lwf.ncdc.noaa.gov/oa/wmo/wdcamet.html), for example, allows an impression of the wealth of data in its listing of datasets.
[12] Global Observing System, GOS. http://www.wmo.ch/web/www/OSY/GOS.html.

Chapter 5: the Koninklijk Nederlands Meteorologisch Instituut (KNMI)

The "Royal Dutch Meteorological Institute" was founded in 1854 and is an Agency of the Ministry of Transport, Public Works and Water Management[13]. Its tasks are twofold; it is both an operational centre and the national centre for climate research. As an operational centre the KNMI collects meteorological data, forecasts the weather and issues warnings in case of dangerous weather situations. KNMI's various services are provided by 500 staff. Its broad activities attract a range of clients from both the public and privates sectors: maritime traffic and aviation depend on the KNMI's forecasts to ensure the safety of goods and travellers; the Netherlands government uses KNMI's expertise for their policy making regarding climate; insurance companies refer to KNMI's records of past storms for possible compensations; even the beer industry bases its production on KNMI's weather data, as the overall beer consumption was found to be significantly higher in good weather conditions.

There are various kinds of information in the meteorological sector and at the KNMI. The KNMI's operational branch collects observational data via a network of weather stations, which gather various ground measurements including temperature, humidity, and global radiation. These kind of observational data are homogeneous scientific data.[14]

Other types of meteorological information include remote observational data such as radar scans and satellite pictures. Specialised bodies such as the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)[15] create these more complex data. It is also a feature of the specialised bodies that they are responsible for the preservation of the remote observation data they create, not the KNMI.

The observational data stemming from both KNMI's own observations and from partners, are processed to create weather forecasts[16] and they are used in a range of research activities by the KNMI as well as other meteorological and climatological services world-wide. Processed information can take a great variety of forms. In addition to this information inhomogeneity, the organisations creating them are for the most part relatively disconnected from an organisational perspective. The KNMI has not yet embedded these information products and research information in a comprehensive preservation strategy.

This case study focuses on the preservation of the observational data that the KNMI collects in its network of weather stations of which they are responsible.

[13] Dutch Ministry of Transport, Public Works and Water Management – Ministerie van Verkeer en Waterstaat. http://www.verkeerenwaterstaat.nl/.
[14] Scientific data are columns of measurement numbers collected to support research. A closer description of the structure of the KNMI's scientific data is provided in the body of the report.
[15] EUMETSAT – European Organisation for the Exploitation of Meteorological Satellites; http://www.eumetsat.de/.
[16] The ERPANET case study on the Meteorological Office in the UK highlights their effort to preserve weather charts; this case study report is available on the ERPANET website at www.erpanet.org.

Chapter 6: Details and circumstances of the interviews

Gerrit Burgers (Head Oceanography Department), Aryan van Engelen (Head Climatological Service), and Piet Helfrich (senior consultant Information Technology) offered their valuable time for comprehensive interviews. Interviews were conducted via telephone in short succession in January 2004. Interviewees provided background information including a copy of the KNMI's digital archiving policy. The web proved a valuable resource with comprehensive background information.[17]

[17] See the References in the Appendix of this report.

Chapter 7: 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

The KNMI had to face the issue of digital preservation very early on. In the 1950s, ship observations were imprinted on punch-cards. Back on the mainland they were disposed off once read and saved digitally as they were considered of no further use. However, the ship stamp on the back of the cards was not read. This failure was recognised too late and the metadata identifying the creator of the data was lost. Altogether about 6 million punch-cards were processed that way. This and similar incidents led to the KNMI's high awareness of digital preservation issues. The KNMI's level of awareness is indicative of this sector, and digital preservation is addressed at data management meetings and is informally discussed within other activity areas within the sector.

The main problems

The KNMI is confronted with a number of challenges in the context of digital preservation. Their network of observation stations accrues a huge amount of data with a measurement frequency of up to once a minute. In addition to handling this data torrent, the reliability of the data must be assured. Formats of these observational data are relatively stable and homogeneous. However, variations in the local set-ups of measurement stations worldwide and the sheer scale of available data demand comprehensive metadata and documentation to allow for global data exchange. Standardisation is therefore a prerequisite for the tight international cooperation in the meteorological sector. With so many stakeholders, however, consensus is often difficult to establish and standardisation processes can take a long time.

Asset value and risk exposure

KNMI's observational data are a building block to its tasks in national weather forecasting and climate research. All past and current data are needed for these tasks, so preservation is a primary business need. Moreover, KNMI's information and expertise are essential to government, industry, as well as the general public. In the international network of meteorological services the KNMI is responsible for the preservation of the observational data it creates as agreed with the other national meteorological services and stakeholders.

Regulatory Environment

The KNMI is a governmental institution and has the obligation to preserve its data according to the rules of the National Archives of the Netherlands.[18] However, the KNMI is exempted from the obligation to transfer its historical data to the National Archives. As part of the agreement between the KNMI and the National Archives, the KNMI has the mandate to steward and to preserve its observational data for an indefinite time period as the national archive of meteorological observations. Also as a WMO member the KNMI is obliged to administer, preserve, and provide access to its climatological data.

Preservation Activity

Policies and Strategies

Observational data is the basis of the KNMI's primary business, and hence their requirements for data management are very high. Many measures they take to support data management, interoperability and data reuse are also conducive to data preservation, and may be qualified as preservation actions. With the clear mandate to preserve its data indefinitely, the KNMI took additional steps to include long-term preservation in their existing data management. Building on its various experiences with digital preservation challenges, the KNMI gradually developed its policy-guided preservation approach now in place.

The first version of a formal archiving policy was created in summer 2001. Initially the policy was designed for the KNMI's Historical Climate project (HISKLIM)[19], which has the goal to digitise historical observational data and integrate them into KNMI's current systems. However, the staff responsible for the creation of the policy soon came to the conclusion that only a policy that spans the entire organisation and is embedded from the time of data creation can be effective. Therefore the policy now includes the entire operational branch of the organisation[20] and all stakeholders involved in the creation and processing of the KNMI's observational data.

Three senior members of staff were responsible for the creation of the archiving policy.[21] They had little external input. The KNMI was one of the first meteorological services to have such a policy and could not fall back on existing experiences in the sector. However, when the KNMI embarked on creating its policy, awareness about digital preservation started to emerge among information professionals in the meteorological sector, and corresponding issues were raised; most notably discussions with the expert team of the World Climate Data Management Programme[22] were helpful. Expert knowledge from other sectors was also used.

The KNMI archiving policy clearly identifies the origin and scope of the data in question as well as the required metadata and organisational issues (including responsibilities). It also calls for adequate redundant data storage facilities to avoid data loss. In its

[18] National Archives of the Netherlands, Nationaal Archief van Nederland. http://www.nationaalarchief.nl/.
[19] Historical Climate - HIStorisch KLIMaat, HISKLIM; http://www.knmi.nl/onderzk/hisklim/.
[20] The KNMI research branch is excluded from the preservation policy discussed in this report; also see the Introduction to the KNMI.
[21] Two of the initially three members of staff who created the archiving policy (Gerrit Burgers and Aryan van Engelen) also served as interviewees in this case study.
[22] World Climate Data Management Programme, WCDMP. http://www.wmo.ch/web/wcp/wcdmp/html/wcdmp.html.

current form the policy addresses purely KNMI internal issues (it excludes for example cooperation with external partners or influences through the KNMI's business context). It does not address monitoring of preservation actions nor does it call for regular updating of the preservation policy. However, all policies at KNMI are updated when necessary and the archiving policy will be updated likewise. Updating procedures are triggered when changes in the daily practice demand this. Moreover, interviewees expect the archiving policy to be adequate for several more years in its current form.

The policy also laid out the need for a coordination group. The group is composed of stakeholders form various departments and is tasked with ensuring the implementation of the policy and keeping an open dialogue with staff on the efficacy of the policy. The experiences with this coordination group are positive up to now.

However, not all those involved in the creation and processing of observational data are KNMI staff, and thus some effort is afforded to disseminating guidelines beyond the organisation. These efforts fall on fertile ground in the meteorological sector with its existing structures for international cooperation and where adherence to codes of practice are customary.[23]

Selection

The KNMI preserves all its observational data created by or on behalf of them. While the technology for these kinds of measurements (e.g. temperature, humidity, flashes) are relatively stable compared to other scientific areas such as chemical and genetic analyses[24], some development occurred over the long history of the KNMI. When the KNMI was founded in 1854, data was predominantly recorded on a daily basis. Around 1900 the resolution was already hourly, and in the 90s the frequency became to be 'ultrahigh' with observations at a rate of one per ten minutes or even one per minute depending on the unit. With changes in frequency and accuracy of the measurements, also the data format changed slightly. The KNMI's climate data system combines all these data, the continuous accumulation of born digital data today with historical data, the digitisation of which is in progress in the framework of the HISKLIM project.

Preservation

KNMI's observational data are taken from measurement instruments at the weather stations and directly transferred to storage for preservation to ensure data reliability. The KNMI's experience led to this policy of preserving data in its original raw form. In the past they had problems with conversions when errors were introduced in the data. Once, for example, the maximum wind-speed was wrongly set for a batch of data due to a conversion fault. This may appear like a minor problem as only one number in the array of incoming data was wrong, but in fact this number may be decisive, for example, for operations at the Dutch airport Schiphol.

Metadata are key for the identification of measurements so that meaningful access to the data remains now and into the future. Original observational data are therefore put in a container together with two basic types of metadata. Firstly, each measurement number is assigned attributes that describe what the number actually refers to (e.g. temperature, humidity), and when the data was recorded at what location. The WMO created standards and recommendations for this type of metadata that are

[23] As early as 1853, the Brussels conference (as described in the introduction to the sector) established a standard format for ship observations in order to facilitate data exchange.
[24] See the case study on the Pharmaceutical Sector, available on the ERPANET website.

internationally recognised. The second type of metadata is a more precise description of the environment of the measurement station that created a number (e.g. the thermometer reading is fixed on the eastern wall of the observation station). This is mainly a textual description that is not covered by any of WMO's guidelines up to now, although some efforts in this direction have been underway for quite some time.

To facilitate international interoperability the WMO defined the SYNOP format[25] that is widely used for exchange of observational data. This format has already been superseded by the BUFR format[26], which introduced a self-descriptive header for metadata. While the BUFR format was defined several years ago it has not yet taken SYNOP's position due to the slow pace of the standardisation process. Embedding metadata directly in the data format was primarily done to support interoperability and data reuse; however, this step is important for digital preservation as well and will allow interpretation of the data in the future.

The KNMI plans to adopt the BUFR format. This change will be relatively easy to conduct as far as the technology environment is concerned. As outlined above the KNMI stores the data together with metadata in its original format rather than in the SYNOP format. This basic data is converted to any other form by software modules, including SYNOP and any other desirable format. When adopting the BUFR format the KNMI will continue to store its data in its original form, and only develop a converter module that generates the BUFR format when required. Thereby, the reliability of data is ensured. The basic form of the original data together with comprehensive metadata will allow access to the KNMI's data stocks far into the future.

The KNMI has accrued a huge data repository. This repository serves the entire organisation and is currently comprised of 35 terabytes of data. All this data has at least one redundant copy for safety reasons. The digitised historical observational data takes a relatively small share of these data stocks. The KNMI estimates that their repository will grow at more than 35 terabyte per year in the future, which includes pilot projects on tapping new forms of observational data including satellite pictures.[27] Their mass storage system can accommodate 600 terabyte (stored redundantly) in its current set-up and is designed to be scalable. This system is a hierarchical storage management (HSM) system, which is based on a robot mechanism that administers up to 6,000 magnetic tapes.

Refreshing storage media is estimated to be necessary about once every five to ten years. There is no formal policy on this, but the KNMI has not encountered any problems to date. In 2003 they took the decision to transfer all their data to new media. This decision was mainly taken for capacity reasons rather than for refreshing, as the new high density devices allow the storage of 10 times more data than the former ones without any modifications necessary to the robot-based system. It is this and similar moves that boosted their storage system to its current size, which is a thousand times more capacious compared to their storage capacity in 1992.

---

[25] SYNOP FM-12, common code for surface observations;
Refer to: Recommendation on Standards for the Representation of Atmospheric Observations in Flood and Fire Disaster Management Projects. A CLIFF (IST-1999-14104) deliverable (May 2002). http://styx.esrin.esa.it/cliff/WP500/docs/D5-4_1.0_15-05-02.pdf .

[26] BUFR - Binary Universal Form for the Representation of Meteorological data;
For the official definition of the code refer to: World Meteorological Organization (WMO) Manual on Codes, Publication 306, Volume 1, Part B, 1988 Edition, Supplement No. 3 (VIII.1991).
A description is also available at: John D. Stackpole: GRIB and BUFR. In: NOST Formats Workshop; April, 1993. http://ssdoo.gsfc.nasa.gov/nost/formats/bufr.html.

[27] see the Chapter "Future Outlook" for more details.

Most of the data is stored redundantly and at least twice, whereby backup tapes are brought to a location about 25 kilometres outside of the KNMI premises. It is up to the discretion of KNMI staff responsible for data creation whether or not their data is further duplicated for additional safety. It is they who can best determine the value of their data, and the KNMI trusts in their decision. They can choose the number of redundant duplications they consider appropriate at data ingest; data duplication and storage is fully automated. Assisting them are 'data managers', who serve as a contact point on questions surrounding data management. These data managers are part of the various KNMI departments rather than staff of the information technology department, but they are trained on their tasks as data managers.

Access

In principle everybody has access to KNMI data, provided the data is not copied and distributed to external parties. Practically most users are KNMI staff who access the data via the internal system. Some of KNMI's information products are available on the KNMI website, but this does not include observational data.

External users need to pay a small fee for data access. After the request of a specific dataset and payment of the fee, the KNMI transfers the data to a designated server. The external party can then access the respective batch of data via the designated server, as the KNMI internal systems are detached from outside for security reasons. This and other security measures are taken in compliance with the information-security instructions 'Voorschrift Informatiebeveiliging Rijksdienst'[28], a compendium by the Dutch government providing guidelines and best practices for ensuring data security.

Compliance Monitoring

As yet, there are no formal and explicit measures for internal monitoring or external auditing of digital preservation. While regular reviews of KNMI's business processes are being conducted, they do not include digital preservation. However, as preservation actions are tightly integrated with data management, the data is monitored on various levels as part of daily practice. Already at data creation the KNMI has internal standards for the quality of observational data that need to be satisfied. The coverage of a temperature measurement in both time and space must be higher than 97.5 percent, with a measurement error that is less than a tenth of the measurement unit, in this case degree Celsius. Concerning data storage, the data managers are together with the staff of their respective departments responsible for the departmental data. And when archived data is accessed and analysed within the scope of research its quality is indirectly assessed as well. These measures on various levels can be regarded as informal internal monitoring, which works fine as all KNMI staff are well aware of the value of their data.

Digital Preservation Costs

Digital preservation is intrinsically linked with other business processes and thus there is no separate figure on digital preservation costs. Some figures can, however, give an impression of the scale of investments taken at the KNMI. Digitisation work, which is a more isolated action, is conducted by five full time equivalents and costs about 500,000 Euro per year. Operational costs for the database system amount to 500,000 Euro per

[28] Voorschrift Informatiebeveiliging Rijksdienst, VIR – instructions for information security management in government departments and agencies.
http://www.rijksarchiefinspectie.nl/wetgeving/overige_VIR1994.html.

year as well. Adding another terabyte of storage to the current mass storage system costs about 3,000 Euro. Generally interviewees are confident that senior management will provide a budget for preservation actions if necessary; this is underlined by last year's conversion of storage media, which was carefully prepared by IT staff and approved of by senior management without further detraction.

Future Outlook

Probably the biggest worry of the KNMI currently is the second type of metadata[29] that describes the conditions under which a measurement is done. This metadata type is currently only an informal textual description without further structure. The lack of a WMO standard is particularly worrying, as the metadata created at the moment may need to be revised in the future. This could cause additional effort in the future, and in the worst case insufficient documentation will only be discovered when it is too late. A WMO task group with KNMI participation is working on a standard, but their current results are not yet sufficient and it may still take time until an agreement is found.

Further development of the technical equipment at measurement stations are expected in the near future. For example the scanning modes of rain measurements, which is in two dimensions now, will become three-dimensional. This leads to a new type of data, the adequate preservation of which remains to be addressed. The biggest challenge may be just approaching as the KNMI ventures into remote-sensing observational data. These data are currently created and preserved by specialised organisations; centres creating satellite data including the European Centre EUMETSAT[30] are responsible for their preservation. However, the KNMI embarked on an Ozone Monitoring project[31] together with NASA that involves satellite data, and it may have the mandate to preserve them as well. This will pose a particular challenge as satellite data is inherently different to the other observational data currently overseen by the KNMI.

The continually rising storage requirements and the close cooperation within the sector almost naturally lead to the interest of the sector in grid technology[32], which allows establishing huge integrated global data networks. Grids will enhance the availability and accessibility of data, and they may also foster synergies and division of labour. They facilitate, for example, the outsourcing of data storage and data processing if this is beneficial. One of the projects in this area is the UNIDART[33] project that will employ grid technology to establish uniform access to meteorological data and products; KNMI will provide the pilot data for this project. The implementation of grid technology may radically transform not only the technological infrastructure but also organisational structures in the meteorological sector facilitating even closer global cooperation on a practical level.

---

[29] See the Chapter 'Preservation Activity – preservation' for details on the second type of metadata.
[30] EUMETSAT – European Organisation for the Exploitation of Meteorological Satellites; http://www.eumetsat.de/.
[31] The KNMI participates with an Ozone Monitoring Instrument (OMI) in the NASA mission "Earth Observing System (EOS) Aura". For closer information see the KNMI's OMI website at http://www.knmi.nl/onderzk/atmosam/English/WorkingGroups/OMI/, or NASA's EOS Aura website at http://eos-aura.gsfc.nasa.gov/.
[32] Reagan W. Moore, Andre Merzky: Persistent Archive Basic Components. Persistent Archive Research Group, Global Grid Forum. Draft 1.0; July 27, 2002.
http://www.sdsc.edu/NARA/Publications/Web/PA-cap.doc .
Also refer to the Wikipedia entry on "Grid Computing",
http://en.wikipedia.org/wiki/Grid_computing .
[33] UNIDART - Uniform Data Request Interface; http://www.dwd.de/UNIDART/ .

Chapter 8: Conclusions

The KNMI has developed a preservation approach with an archiving policy that has an emphasis on organisational issues, identifying responsibilities and involving the entire observational branch of the organisation in preservation actions. The archiving coordination group and the role of the data managers are two measures aimed at enhancing communication with staff, who are after all the pillars of any digital preservation approach.

The KNMI archiving policy builds on comprehensive technology standards and codes of practice from the meteorological sector for their active management of their data assets, which are geared at facilitating data reuse and interoperability with their global partners. The KNMI excelled in embedding preservation measures into their existing structures and their business environment. This is particular to the KNMI and to the meteorological sector. Organisations in other sectors may have 'softer' requirements for their active information management, and they may face more complex data types that demand specialised preservation actions, and they may have to implement a preservation approach that is more clearly separated from their active information management.

Preservation at the KNMI's research branch is not as advanced as the preservation of observational data that is described in this report. The various research projects differ drastically from the operational branch regarding their work as well as their information products[34] (though the research projects build on the observational data created and preserved by the operational branch, of course). This leads to the clearly separate preservation approaches in the operational and the research branch of the KNMI. The KNMI found the smallest common denominator between them to be data storage; storage facilities are equally available to the whole organisation including redundant storage and refreshing of data, but preserving accessibility to the information needs to be tackled by the various departments separately by providing exhaustive documentation and perhaps even applying regular format migration. This layered preservation approach is a valuable model for any other organisation that has similarly independent departments with diverse information products to preserve.

The long history of standardisation in the sector, and their willingness to cooperate closely and to also make concessions may lead to global solutions for preserving operational data in the future. Despite these good prerequisites to achieve a preservation strategy for the whole sector, this will no doubt take time. First steps in that direction have, however, already been taken. On the lowest level, data storage, projects are under way to combine efforts in the sector and to establish a grid network that serves the whole sector.

Despite the positive situation today, the KNMI cannot rest on its laurels. With the prospect of inherently different and more complex forms of observational data, the KNMI faces new preservation challenges, and needs to continually adapt their preservation approach to the changing technological and organisational landscape. The institute is, however, in a good position to tackle these challenges and preserve its assets reliably into the future.

[34] This is also underlined by the ERPANET case study on the Meteorological Service UK, which focuses on the preservation of weather charts; an inherently different type of information to the observational data focused on in this case study.

Appendix 1: References

Koninklijk Nederlands Meteorologisch Instituut (KNMI)
http://www.knmi.nl/

Koninklijke Bibliotheek: Literatuur Dossier 150 jaar KNMI
– a literature collection of the National Library of the Netherlands at the KNMI's 150 years existence jubilee
http://www.kb.nl/kb/dossiers/knmi/knmi-lit1.html

archived (climatological) data at the KNMI

Historical Climate - HIStorisch KLIMaat, HISKLIM
http://www.knmi.nl/onderzk/hisklim/index-en.html

Climatological data and metadata of the Netherlands
http://www.knmi.nl/voorl/kd/index_eng.html

KNMI Climate Explorer, global climatological datasets
http://climexp.knmi.nl/

European Climate Assessment & Dataset (high resolution datasets)
http://www.knmi.nl/samenw/eca/

KNMI Hydra project, Dutch wind data set
http://www.knmi.nl/samenw/hydra/

Organisations

EUMETNET – The Network of European Meteorological Services;
http://www.eumetnet.eu.org/

European Meteorological Society
http://www.emetsoc.org/

World Meteorological Organization
http://www.wmo.ch/index-en.html

WMO World Weather Watch, WWW; http://www.wmo.ch/web/www/www.html

some collaborative initiatives

CLIWOC project - CLImatological database for the World's OCeans 1750-1850
http://www.knmi.nl/onderzk/hisklim/CLIWOC/CLIWOC.html

DAAC - NASA Distributed Active Archive Center
http://daac.gsfc.nasa.gov/

UNIDART - Uniform Data Request Interface
http://www.dwd.de/UNIDART

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

ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK