



Information Society  
Technologies

# erpastudies

**deutsche  
presse-agentur  
(dpa)**

---

[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 <http://www.erpanet.org>.

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

ISSN 1741-8682  
© ERPANET 2004

## **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 Deutsche Presse-Agentur (dpa) .....</b>	<b>7</b>
<b>Chapter 5: Circumstances of the Interviews .....</b>	<b>8</b>
<b>Chapter 6: Analysis .....</b>	<b>9</b>
Perception and Awareness of Digital Preservation.....	9
Preservation Activity.....	10
Compliance Monitoring .....	13
Digital Preservation Costs.....	13
Future Outlook .....	13
<b>Chapter 7: Conclusions .....</b>	<b>15</b>
<b>Appendix 1: References .....</b>	<b>16</b>

## **Executive Summary**

The Deutsche Presse-Agentur (dpa), the leading German news agency, currently produces some 4000 news items per day and distributes them through its different news services worldwide. All of these are preserved indefinitely and make up the agency's news archive, currently consisting of some 14 million textual news items. The archive is one of the dpa's main business assets, since it offers real-time access to reliable information back to 1983 when digital message delivery was first introduced.

Format standardisation of news messages is high throughout the news agency sector, which facilitates preservation. The dpa uses a format based on the IPTC Recommendation 7901 for all of their text news items. These are converted into an XML based preservation format for archiving purposes and stored in an offsite database. All other preservation and value-adding activities are the internal responsibility of the dpa's documentation service. The current solution fully meets the dpa's needs, although the agency envisages that migration of the system will be required in 2 to 3 years and have already begun to make appropriate preparations.

Every year, around 200,000 archived news items are requested by the dpa's editors or by external customers. Items are delivered either in the preservation format or re-converted to the agency's standard news format. The dpa news archive is a highly active and current archive. Quick access is key and providing value-adding services is a must. Access requirements thus drive preservation activities.

## **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 [Hwww.erpamet.org](http://www.erpamet.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.  
Hhttp://www.erpanet.org/charter.phpH.

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 [Hhttp://www.erpanet.org/studies/index.php](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 [Hwww.erpanet.org](http://www.erpanet.org) for the composition of this committee.

## **Chapter 4: The Deutsche Presse-Agentur (dpa)**

Deutsche Presse-Agentur GmbH (dpa) – German Press Agency Ltd.  
<http://www.dpa.de/>

The dpa is the leading German news agency and counts among the leading four worldwide. It is the only German-language news agency that covers the whole spectrum of international and national affairs, featuring text, photo, audio, graphics, and multimedia services. Following a merger of three national news agencies, the dpa started business in 1949 and international activity soon followed. Today, the dpa has editorial production departments in Washington D.C., Cork, Buenos Aires, Madrid, Nicosia, and Bangkok, and produces news services in German, English, Spanish, and Arabic. The dpa has not only German but also English, Spanish, and Arabic speaking journalists in many regions of the world to cover the needs of its different services. News is not translated, but scrutinised by the various desks in the originating language for possible inclusion in the respective service. The company's headquarters are located in Hamburg, while the photo services are centred in Frankfurt a/M and the audio services in Berlin.<sup>5</sup> This case study focuses exclusively on the text news services.

The dpa is a news wholesaler. This means that the agency does not publish news itself, but provides it to newspapers, journals, radio stations, broadcasting companies, and online providers. In addition, the dpa also provides tailor-made news services for parliaments, governments, political parties, organisations, and industry.

Through its various news services, the dpa produces and releases some 4,000 news messages per day. The agency currently has about 900 employees (of which 550 are editors), some 100 of which are foreign correspondents. Several thousand freelance journalists work for the dpa as well. The company had a turnover of 106 million Euro in 2001. The dpa umbrella group comprises a dozen further press and image agencies in Germany and abroad. A documentation department is responsible for the news repositories and manages news structures and taxonomies. It reports directly to the editor-in-chief but also plays an important role when it comes to market the assets in archives and the dpa's news-knowledge.

The dpa is a limited company owned by some 200 shareholder firms from the publishing and broadcasting sectors. An executive board manages the financial matters, while the editorial management is responsible for editorial matters. Both are accountable to the company's supervisory board.

---

<sup>5</sup> For further information of the dpa group of companies please refer to <http://www.dpa.de/en/unternehmenswelt/unternehmen/unternehmensgruppe.html>.

## **Chapter 5: Circumstances of the Interviews**

ERPANET first contacted the dpa in September 2003. The press office provided contacts to the documentation section, and Ms Barbara Bliefert, head of documentation, agreed to participate in an interview. This was conducted by telephone on February 26, 2004, and lasted for sixty minutes. The interview was supplemented with additional information by the head of technical development, Mr Klaus Herwig, and the Senior Vice President Technology, Mr Klaus Sprick.

ERPANET wishes to thank all people involved for their very valuable participation.

## **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

ERPANET's interviews with the dpa focused on the preservation of text news items, since these are the most characteristic kind of digital information as well as the main information asset. The dpa's photo, audio, graphics, and multimedia services are the responsibility of separate subsidiaries and are outside of the scope of this study, as are digital business records. Consequently, this report focuses on the news items only, although a similar solution to the one presented here exists for the dpa's digital images.

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

The importance of preserving the dpa's news items is beyond doubt among management and staff. As detailed below, the archive of news items is one of the dpa's main assets, and everybody is aware of its value. Both the dpa's staff and external clients intensely re-use archived information, and the news archive is a key resource in the daily editorial business. Collaboration on preservation is sought with other news agencies and publishing houses, but not beyond the sector.

#### *The main problems*

News always has a history. Nothing happens without relationship to other events. To be able to put current events into context, news editors at the dpa rely on the agency's huge database of past news items. The requirements for their reliability and availability are high. In fact, real-time access to the archive database is a key factor in the dpa's work, 24 hours a day, 7 days a week – all the more so since the agency's main offices are spread around the globe. To be able to deliver news faster than competitors is one of the key factors for the success of every news agency.

However, time is not everything. Quality is the other key factor for success in the news business. Whilst reliability is established during the editorial process, measures must be taken to guarantee it for the long term.

Since in the international news business a high degree of automation characterises the information exchange, interoperability and automatic format transformation are a must. This is also important with view to the dpa's user base. Since the agency provides its news to press houses and broadcasting stations, bulk handling is important.

#### *Asset value and risk exposure*

The dpa produces some 4,000 news items per day through its various services. The first system that incorporated digital workflow processes was put into use as early as 1973. A second generation followed in 1980, with some 400 video editing terminals for automated workflow and on-line editing. Since 1983 news items are produced and

stored in digital form. These are being retained indefinitely, with over fourteen million archived news items accrued over the past twenty years.<sup>6</sup>

The value of the archived news items, besides the sheer wealth of information they represent, rests on two main characteristics for both for the agency's editors and for external clients: their high quality, and; their neutrality. The dpa follows internal guidelines for their news items. Information can only be delivered if it has been checked for accuracy by at least one other editor. The dpa also claims that the information it distributes is politically neutral and sticks to the facts, using and naming its reliable sources. This makes the dpa's news archives a particularly reliable source for historical research, and complementary to alternative information distributed by other media such as newspapers and broadcasting stations.

Consequently, the main risks associated with the news archive are of a business nature. The dpa editors access some 200,000 archived news items per year. The instant access to this basic and background information is therefore of paramount value for daily business. Loss of this information would decrease the daily amount of work that editors can accomplish. In particular, double-checking of news would become much more tedious. Furthermore, financial risks are also evident. The dpa generates some revenue through the sale of access to their news archive to external customers. This income would be threatened by information loss.

Legal risks are on the other hand negligible. The dpa is not legally bound to preserve their information, and loss of information would thus not have legal consequences.

The dpa has not conducted a formal risk assessment. According to ERPANET's interviewees, the risks at stake are evident for everybody within the agency, and in particular for management.

### *Regulatory Environment*

There are no legal prescriptions as to how long news items have to be preserved, and the agency is free to keep them as long as it sees fit. The dpa decided to preserve its news items indefinitely.

International cooperation among news agencies has led to considerable levels of standardisation. In particular, the current news format is defined as an international standard to facilitate information interchange and use. For more information on this, see the *Preservation* chapter below.

## **Preservation Activity**

### *Policies and Strategies*

There is no explicit policy for digital preservation, but a framework of general rules and guidelines applies throughout the entire dpa. These mainly specify technical and organisational details.<sup>7</sup> They do not give detailed guidance for preservation but the preservation work is adapted to these general guidelines.

---

<sup>6</sup> News items that predate this move to the digital realm have been appraised, and relevant items are stored in a paper archive. Some of them are also stored on microfilm.

<sup>7</sup> Two of the most important guidelines are the dpa style guide and the news format. The style guide (<http://www.dpa.de/en/stylebook/home.htm>) serves to unify the way news items are presented, the spelling, punctuation, transliteration, etc. On the news format see below.

Digital preservation of news is the responsibility of the dpa's documentation service. This unit comprises some twenty staff with different backgrounds (for example, documentalists<sup>8</sup>, historians, IT specialists, economists, editors). All receive on-the-job training and have highly specialised tasks. The unit also offers a trainee position. Apart from preservation, other tasks of the documentation service include the aggregation and final editing of news, as described below, and a reference service for the dpa editors. This service includes researching and delivering information from the dpa news archive and external news providers. The documentation service also produces ready-to-use text blocks. Some 100 reference requests are treated per day.

It becomes clear from the above that curation is the key activity related to the preservation of news items.

### *Selection*

It is a basic policy of the dpa that all news items released through one of its services are preserved indefinitely. Therefore, no further selection is applied. It is possible, however, to see a kind of implicit selection in the editing and publishing process: Only what is deemed noteworthy will finally be released as news and consequently preserved.

Under certain circumstances the documentation service undertakes appraisal of news items. Whenever an event is treated by a series of news items (as is frequently the case) in many cases one of them can serve as a reference article because it contains all of the available information. It is the responsibility of the documentation service to identify the most valuable news item and to flag it specifically. Consequently, users will be referred to this item first when they seek information about the specific event.

In other cases, documentation staff (usually one of the editors working with documentation) write the reference item themselves, based on previous items. For these items, the same rules apply as for all the dpa's news. This implies that one or two editors check the news item before it goes into the archive database.

These procedures are accord with the dpa's objective to provide timely access. They support quick access to archived information by reducing the number of items through which users must trawl to access the most pertinent information about an event, as users will not have to go through a series of news items, but are directly referred to the reference item. Quick access to the essential and condensed information is also perceived as more important than the more tedious access to the whole, less structured set of original news items, although these remain accessible at all times.

### *Preservation*

The dpa news archive consists of an active database that stores around 14 million news items. Currently, the dpa is running the second generation of its archive database, which is hosted by an external company. The decision not to host the database in-house has been influenced by financial considerations related to organisations demands for reliability and quick response times. The external providers were able to offer a stable platform, while building up an in-house hosting has been judged too costly. The database used is named Trip. This is an old database system considered to be extremely reliable that is also used by other, large knowledge portals.<sup>9</sup>

---

<sup>8</sup> Akin to information managers.

<sup>9</sup> Other usages include the Genios economic databases ([Hhttp://www.genios.de/H](http://www.genios.de/H)) and the Trip Plus healthcare database ([Hhttp://www.update-software.com/trip/about.htmH](http://www.update-software.com/trip/about.htmH)).

While the future of the Trip database format is somewhat uncertain due to frequent ownership changes during the last years, its considerable user base is considered by some users as sufficient indication that the format is not going to disappear soon. However, it is obvious to the dpa that the life span of the current database is limited to another two to three years. Consequently the agency has started to look into migration choices. Staff are evaluating potential for migration to an Oracle platform and also evaluating the possibility of hosting the database in-house.

The format in which news items are released and distributed is an international standard. The International Press Telecommunications Council (IPTC)<sup>10</sup>, founded in 1965 and of which the dpa is an active member, maintains these standards. The dpa has been following them since 1980. The main format that the dpa uses for text services is the "dpa-Diensteformat". This is based on the IPTC Recommendation 7901, "the text transmission format", but partially adapted to the German language.<sup>11</sup> Adaptation has been made in collaboration with other German-language news agencies.

The preservation format used for the news archive is based on XML. News items that are ready for preservation are automatically converted into the preservation format and transferred into the archive database via a dedicated and secured line. The database is read-only. Logs are recorded whenever a news item is transferred to the database and for every change to the metadata in the database. In the occasional case of disagreement on or challenge of any of the dpa's news items these logs serve to prove the original content. When an archived item is requested, it is delivered either in the preservation format or converted back into the "Diensteformat".

Documentation for both delivery and preservation formats is open and freely available. Therefore, the dpa estimates that successful conversion into other formats will not represent any particular challenge.

Metadata recorded with the news items are restricted to descriptive metadata. The dpa uses the IPTC subject code standard SRS (Subject Reference System) to assign keywords to its news items.<sup>12</sup> The dpa is also involved in the standardisation process for the SRS. Descriptive keywords are maintained for the preserved items. Metadata are assigned manually, but automatic metadata extraction is currently being researched. No specific preservation metadata are recorded, except for the news format specifications.

The archive database is backed up daily by the host. The dpa itself regularly copies a backup of the database on CD-ROMs as a second layer of security. Some years ago, at the time of the first generation of the database, the then hosting company mistakenly overwrote their backup disks, but the dpa succeeded in restoring their data from the CD-ROM backup.

---

<sup>10</sup> International Press Telecommunications Council: [Hhttp://www.iptc.org/H](http://www.iptc.org/H).

<sup>11</sup> See [Hhttp://www.iptc.org/IPTC7901/H](http://www.iptc.org/IPTC7901/H) for information about IPTC 7901 and a downloadable version of the format. The dpa-Diensteformat specification is available at [Hhttp://www.dpa.de/de/produkte/service/pdf/TH\\_50200-01.pdf](http://www.dpa.de/de/produkte/service/pdf/TH_50200-01.pdf).

The development of IPTC 7901 has been frozen in 1995 to make way for new standards, but the format continues to be in widespread use throughout news agencies worldwide.

<sup>12</sup> More precisely, IPTC maintains a number of sets of terms or "Topic sets". The Subject Reference system is made up of the most basic of these sets of term. See [Hhttp://www.iptc.org/metadata/H](http://www.iptc.org/metadata/H) for more information about IPTC sets of terms and the SRS.

## Access

A powerful rights management system controls access to the archive database. This differentiates between users according to the type of news service they access. Through these measures the dpa guarantees the integrity of information in their news archive.

The dpa editors have real-time access to the archive database. For access management purposes, every editor has his own personal account. External clients have delayed access to the archive, whereby the delay times depend on their contract and the fees they pay. This is to guarantee that the news archive is not misused as a substitute for the actual news service.

As has been detailed above, items are delivered either in the preservation format or converted back to the "Diensteformat"; and the dpa's editors can additionally benefit from further editing services performed on the archived news by documentation staff. Access is web-based, either through the company's intranet or through the Internet.

Around 200,000 archived news items are requested annually. Given this demand and the importance of quick access, the database is available round-the-clock, with only small maintenance windows.

## **Compliance Monitoring**

There are no specific monitoring procedures in place regarding compliance to the dpa's rules and guidelines. However, format compliance checks are performed on the messages when they enter the archive system. For these purposes the dpa uses some routine programs that scan the data transfer to the archive database and alert the responsible persons if required fields are missing or if the hourly message flow falls below or exceeds normal boundaries. Furthermore, the number of recorded daily news items is checked the following day against the number of items that were stored in the system.

Monitoring of the physical storage media and of the database is the responsibility of the database host.

## **Digital Preservation Costs**

Once a year, the dpa conducts a cost benefit analysis for digital preservation. The documentation section is obliged to assess and possibly reduce their costs on an annual basis. In addition, the dpa analyses the internal cost structure of its service, and documents the number of requests each business unit makes to the documentation service. However, these figures are not used for amortisation, but solely serve information purposes.

The documentation service has a part of its budget specifically allocated for long-term preservation of news items. The dpa was not in a position to reveal detailed figures, but the long-term preservation costs are roughly in the dimension of one percent of the company's turnover. This budget covers the present preservation needs.

## **Future Outlook**

Interviewees consider the dpa's current news archive solution to be robust and stable, and there are no issues to be addressed immediately. It is clear, however, that the dpa must keep up to date with information and developments in the fields of news formats

and database solutions. As for the volume of information to be archived, the dpa's business habits make it highly predictable. Safe predictions for future growth can be built upon the figure of 4,000 news items per day. This may change, however, should the company close one of their services, or open or acquire a new one.

One major field of action for the near future has already been identified: as mentioned above, the current database solution will likely meet the end of its commercial lifetime in some 2 to 3 years, which will make migration necessary. The responsible people at the dpa have already begun to assess this challenge. They are focussing in particular on which platform to choose and on whether the dpa will host the new database in-house or again outsource the hosting.

Concerning the news format, IPTC currently works heavily on developing the XML based NewsML<sup>13</sup> standard. As of today, this has not yet been adopted by the dpa, but staff are monitoring its development and expect to adopt this standard some time in the future.

In a general way, the head of documentation underlined the importance of providing added-value services to the news archive. Since the archive is heavily used in-house and generates revenue through licensed users, constantly improving its usability and access is indispensable. There are two projects that align with this demand, namely the integration of the news archive with the dpa's image archive, and the construction of an archives portal that integrates different services. While at present the market demand for these innovations is not high enough to warrant pursuing them further, they reveal some of the directions the dpa will take in the more distant future.

---

<sup>13</sup> See [Hhttp://www.newsml.org](http://www.newsml.org). See also the ERPANET case study on the sda, which already uses this format for archiving textual news items, available at [Hhttp://www.erpanet.org/studies/H](http://www.erpanet.org/studies/H).

## **Chapter 7: Conclusions**

Access to past news items is one of the key elements of daily business at the dpa. First of all, it is indispensable for editorial work, and additionally, it is a source of revenue. As a consequence of this, digital preservation receives high attention among the dpa management and staff. Unlike other organisations analysed through ERPANET case studies<sup>14</sup>, those responsible for preservation at the dpa do not need to convince their co-workers and superiors of the importance of their work. The present study once again underlines the oft perceived notion that financial considerations are the most powerful, if not the only, factor that drives digital preservation efforts. Since the dpa would face considerable inconvenience and risks from information loss, preservation receives constant attention.<sup>15</sup>

While digital preservation is strongly supported by the dpa's focus on access at organisational level, there are at least two technical characteristics that enable the agency to implement a solid preservation solution. Firstly, the level of standardisation is high among news agencies. The IPTC as the international rule-setting body promotes robust and widespread standards for agency news. The dpa only uses a small number of standards, all of which are highly structured, well documented, and openly specified. These general characteristics of the news business are ideal conditions for implementing a digital preservation solution and offer also good perspectives for future format changes or migrations.

Secondly, the use of standard formats has further preservation implications as the news items are saved in what is essentially structured text. This has obvious benefits for long-term preservation in that structured text is easier to understand and maintain through time than unstructured text. Clearly the preservation challenge for the dpa's documentation section is much smaller than that of for example, the National Library of Wales.<sup>16</sup>

---

<sup>14</sup> See, e.g. the case studies on the Council of Europe or the International Labour Organization, available from [Hwww.erpanet.org/studies/H](http://www.erpanet.org/studies/H).

<sup>15</sup> Note that therefore the archive is actually an active database, not an archive in the strict sense of the word. This is very similar to the situation at the European Patent Office (see the respective ERPANET case study), where the main information asset, the patent database, is both an archive and an actively used database.

<sup>16</sup> See the respective ERPANET case study. Of course, this comparison only holds true for both organisations' main information assets. Business records have been disregarded in this report; it is foreseeable that their preservation would pose similar challenges to both organisations.

## **Appendix 1: References**

The dpa  
<http://www.dpa.de/>

International Press Telecommunications Council  
<http://www.iptc.org/>

The IPTC 7901 Text Transmission Format  
<http://www.iptc.org/download/download.php?fn=IPTC7901.zip>

The dpa's "Diensteformat" based on IPTC 7901  
[http://www.dpa.de/de/produkte/service/pdf/TH\\_50200-01.pdf](http://www.dpa.de/de/produkte/service/pdf/TH_50200-01.pdf)

The IPTC Subject Reference System  
<http://www.iptc.org/metadata/>

Gruener+Jahr  
<http://www.guj.de/>

ERPANET  
<http://www.erpanet.org/>

**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)