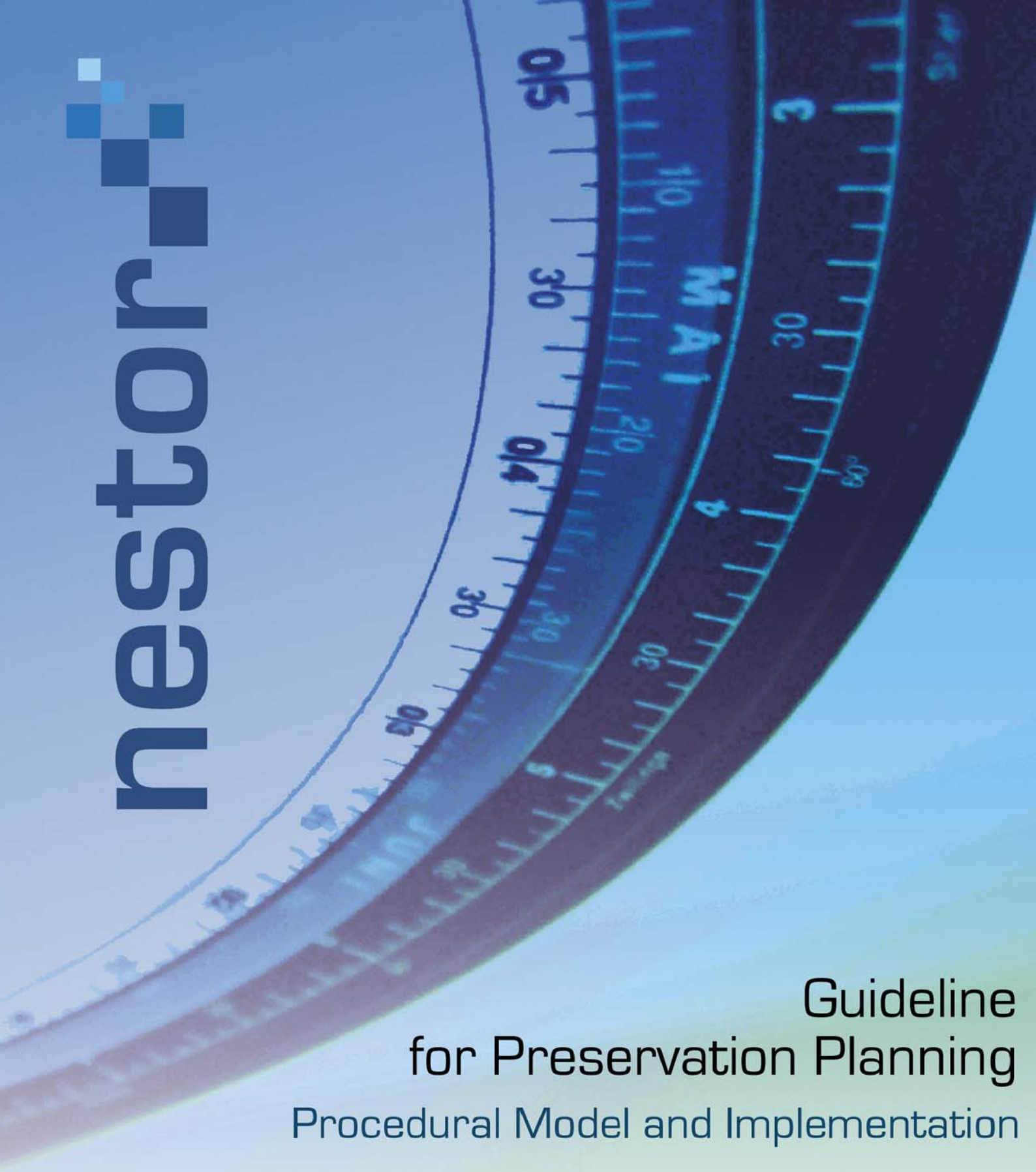

Guideline for Preservation Planning

Procedural Model and Implementation

authored and published by the
nestor working group Preservation Planning

nestor materials 15

Guideline for Preservation Planning

Procedural Model and Implementation

authored and edited by the
nestor working group Preservation Planning

nestor-materials 15

nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland

nestor - Network of Expertise in Long-Term Storage of Digital Resources

http://www.langzeitarchivierung.de

nestor partners

- Bayerische Staatsbibliothek
- Bibliotheksservice-Zentrum Baden-Württemberg
- Bundesarchiv
- Computerspiele Museum Berlin
- Deutsche Kinemathek – Museum für Film und Fernsehen
- Deutsche Nationalbibliothek
- FernUniversität Hagen
- GESIS - Leibniz-Institut für Sozialwissenschaften
- Goportis - Leibniz-Bibliotheksverbund Forschungsinformation
- Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
- Georg-August-Universität Göttingen / Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Humboldt-Universität zu Berlin
- Institut für Deutsche Sprache
- Institut für Museumsforschung (Stiftung Preußischer Kulturbesitz)
- Konrad-Zuse-Zentrum für Informationstechnik Berlin
- Landesarchiv Baden-Württemberg
- Landesarchiv Nordrhein-Westfalen
- PDF/A Competence Center
- Rechenzentrum der Universität Freiburg
- Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden

URN: urn:nbn:de:0008-2014102709
[http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2014102709

Guideline for Preservation Planning.
Procedural Model and Implementation

Version 2.0

Authored and published by the nestor working group Preservation Planning

nestor materials - 15

nestor materials - 15

Guideline for Preservation Planning. Procedural Model and Implementation

Version 2.0

Authored and published by the nestor working group
Preservation Planning

urn:nbn:de:0008-2014102709

Frankfurt am Main, October 2014

Imprint

nestor "Preservation Planning" working group

c/o Landesarchiv Baden-Württemberg

Dr. Christian Keitel

Eugenstraße 7

D-70182 Stuttgart

Tel.: +49 -711 -212-4276

Fax: +49 -711 -212-4283

e-mail: christian.keitel@la-bw.de

nestor Competence Network

c/o Deutsche Nationalbibliothek

Adickesallee 1

D-60322 Frankfurt am Main

Tel.: +49 -69 -1525-1141

Fax: +49 -69 -1525-1799

Web: www.langzeitarchivierung.de

Authors of the Guideline:

Reinhard Altenhöner, Deutsche Nationalbibliothek

Georg Büchler, Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen

Jörg Homberg, Brandenburgisches Landeshauptarchiv

Karsten Huth, Sächsisches Staatsarchiv

Dr. Mathias Jehn, Universitätsbibliothek J. C. Senckenberg

Dr. Christian Keitel, Landesarchiv Baden-Württemberg

Jens Ludwig, Staats- und Universitätsbibliothek Göttingen

Heike Maier, Kulturamt-Stadtarchiv Stuttgart

Dr. Stefan Rohde-Enslin, Institut für Museumsforschung, Staatliche Museen zu Berlin

Dr. Peter Sandner, Hessisches Hauptstaatsarchiv

Kathrin Schroeder, INFORA GmbH

Sabine Schrimpf, Deutsche Nationalbibliothek

The working group was established on 10 November 2009 in Stuttgart. Decisions were subsequently taken on content-related and organisational aspects and version 1.0 of the guideline was then prepared in six further sessions before eventually being published in autumn 2011. Public comments received during the period from November 2011 to April 2012 have been incorporated in version 2.0. This English translation was published in October 2014.

We would also like to thank the following for their input:

Dr. Henrike Berthold, Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden

Miriam Eberlein, Stadtarchiv Heilbronn

Dr. Kai Naumann, Landesarchiv Baden-Württemberg

Dr. Christoph Schmidt, Landesarchiv Nordrhein-Westfalen

Dr. Astrid Schoger, Bayerische Staatsbibliothek

Dr. Werner Schweibenz, Bibliotheksservice-Zentrum Baden-Württemberg

Content

Management Summary

The guideline for preservation planning describes a procedural model for the long-term archiving of digital objects and provides information on possible forms of implementation. It serves above all as a theoretical and practical implementation of the "Preservation Planning" functional unit of the OAIS reference model. Other key concepts introduced in the last 15 years have been included and brought together.

Established terminology has been used where possible. Concepts and related terms associated with the topics of *designated use*, *information type* and *preservation group* have been introduced. The purpose of allocating information objects to these categories is to facilitate the necessary joint and automated processing of the same kind of objects. File type is inadequate as a category in this context. Sample implementations are provided in an appendix.

The guideline takes into consideration the fact that ideal solutions are rarely attainable and focuses instead on the four conditions of financial viability, authenticity, adequacy and potential for automation.

The guidelines is targeted at the traditional memory institutions - libraries, archives and museums - and at other organisations tasked with preserving information. The members of the nestor working group which authored the guideline all work in such institutions.

1. Introduction

SUBJECT MATTER OF GUIDELINE

The present guideline describes a procedural model for developing a set of preservation planning regulations for a specific institution and provides a suitable framework for this. It is based on theoretical concepts from the last 15 years which are being brought together in this specific context for the first time.

The guideline describes the categories, characteristics and processes which are essential for preserving digital objects. In particular it determines relevant pre- and framework conditions and lays down technical specifications in the form of a systematic and documented needs analysis. Digital preservation is already well established in various communities and is currently being implemented in numerous projects. Conformity to a globally recognised OAIS (Open Archival Information System) reference model, ISO standard 14721, is a basic requirement here, albeit one which is seldom fully met. "Preservation Planning" is a functional unit in the OAIS model which is often postponed. It consists of activities which serve the systematic preparation of preservation planning measures and therefore safeguard the accessibility of archived objects. The purpose of the present guideline is to fill this gap - both in theoretical and practical terms.

The long-term preservation of digital objects starts from the initial contact between the information producer and the preservation repository, even before any objects have been ingested into the digital archive. Parts of the "Ingest" and "Administration" OAIS functional units have therefore also been incorporated in the guideline.

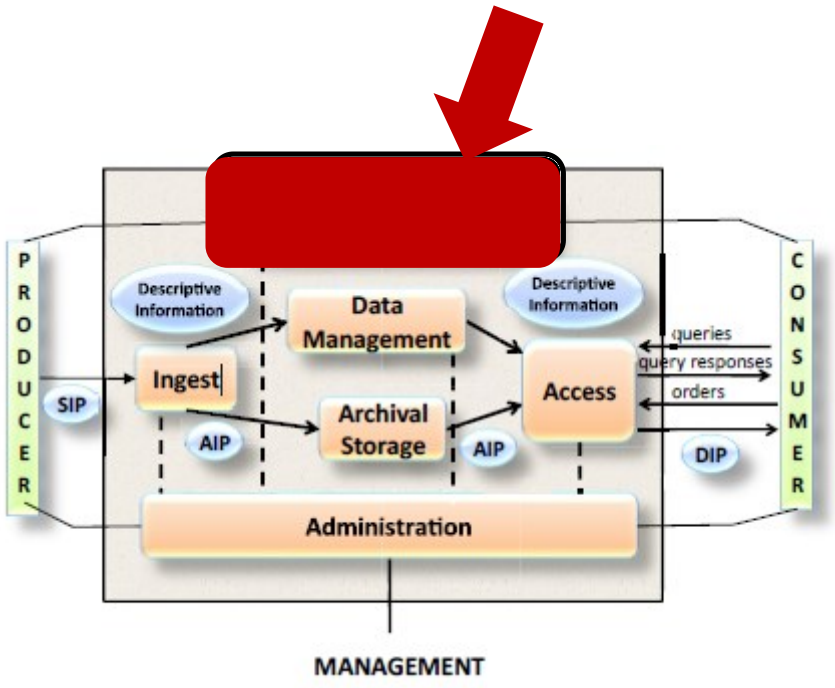

Figure 1: Functional units of the OAIS reference model

BASIC PRINCIPLES

According to the OAIS standard, the main task of digital preservation is to preserve the usability of digitally stored information, regardless of whether any changes occur to hard- or software. This is based on the (assumed) wishes and knowledge of future users (designated community). Which preservation planning measures are necessary to preserve information stored in digital form in line with this standard? In the search for suitable answers, various concepts and approaches have been devised over the last 15 years:

- 1999: Concept of significant properties (Cedars project)
- 2002: Concept of designated community (OAIS)
- 2002: Performance Model (National Archives of Australia)
- 2005: Concept of intellectual entity and representation (PREMIS)
- 2006 ff.: Preservation planning-concept and implementation in PLATO (Planets Project)
- 2007 ff.: Format selection catalogues (e.g. AKEA, KOST)

From the current perspective, many of these models have also brought forth aspects which have proved crucial for answering the initial question.

The guideline makes use of these approaches and describes them in a process overview. What is new about this guideline is this overview of the different aspects and their incorporation and allocation within an overall system: on the one hand, competing views at the conceptual level, which often arise as "counter models", reference each other; on the other hand this means that they are applied to practical considerations and matters, allowing the essential functions and characteristics to be crystallised and their significance for practical requirements to be assessed accordingly.

Besides OAIS and the concepts mentioned, the guideline also builds upon the "Into the Archive" guideline also generated by the nestor project. A further key point of reference is the PREMIS standard. This applies in particular for the terminology used which - except in cases where new terms are introduced - is based on the DIN standards 31644 and 31645.

AUDIENCE

The primary audience of the guideline are *memory institutions* such as libraries, archives and museums - i.e. institutions tasked with preserving the digital material entrusted to them and keeping it accessible for use for an indefinite period. The guideline also provides information to other institutions on preserving digitally stored information in general, on developing concrete implementation solutions and on establishing processing rules. It is also intended for institutions which offer services to the aforementioned institutions.

USING THE GUIDELINE

Besides modelling, which is crucial for understanding, the main purpose of preservation planning is the preservation of concrete information objects. The large number of these requires the joint and automatic processing of similar objects, and therefore the definition of such jointly processable groups. Here, the working group has introduced the concepts of *designated uses* on the one hand and *information type* or *preservation group* on the other hand, as file type does not represent an effective category. These concepts are integrated into a model with existing terms such as *Significant Properties, Designated Community,*

Performance Model or *Intellectual Entity/Representation* in the **main part** of the guideline (sections 2 to 5). The entire process of digital archiving is described on the basis of the four premises of financial viability, authenticity, adequacy and potential for automation described below. This renders the resulting individual requirements practicable.

The description must of necessity remain abstract in order to ensure the transfer of the entire process into the widely varying contexts of the individual preservation repositories. The working group has compiled a list of possible implementation methods for the information types of text, image, audio recording, moving image, structured information, geographic information system and software in **Appendix A**. These implementations should be regarded as examples of solutions which can be used directly or fundamentally re-adapted to suit the situation of a particular preservation repository.
The appendix should therefore be viewed as a proposal which digital repositories can make use of for their own purposes, or as a source of other implementations.

Some processes in the guideline are described discretely, whereas in daily practice they are interwoven as the result of various interactions. For instance, the recording of the individual information objects and the definition of the information types developed by the repository generally have a mutual impact upon each other. The discrete descriptions given here are intended to aid understanding of the processes, but they should not suggest artificial distinctions in the practical implementation.
Individual processes are described in ideal-typical terms. Here, too, the priority is upon understanding the individual steps. At this level of abstraction, processes describe activities rather than technical implementations. These are not system-dependent and can therefore be implemented in different ways. Just because the initial entry of an information object is described in the following section in a particular way for process modelling reasons, it cannot be concluded, unless other previous knowledge dictates otherwise, that every single information object in a repository must be analysed in the same way. Instead, most preservation repositories have generally amassed experience with different information objects which they can then draw upon.
The guideline does not cover issues related to bitstream preservation. Digitisation issues are not covered either, although individual results may also be applied to digitisation.

PREREQUISITES AND CONDITIONS

In preparing the guideline, the working group took into account a number of conditions which determine the practical implementability of a procedural model. These were used to derive four premises which can serve as a measure of the practicability of the measures presented.

- **Financial viability:** A maximum solution for archiving electronic information could be the transfer of the original object in a planned migration into all possible target formats in maximum quality. The same applies for the selection, creation and configuration of suitable emulators. Such a solution is of course neither financially quantifiable and therefore plannable, nor is it suitable, by its very nature, for being funded. In practice, a decision has to be made between two or more options. This can also mean deliberately limiting the scope of actions for the future, thereby risking losing full accessibility of the digital objects. This basic situation, however, necessitates a conscious and fully documented approach; the guideline aims to provide substantial assistance here.
  The working group therefore believes that preservation planning needs to find a sound and viable balance between the material requirements and the financial viability of corresponding measures. The same applies for evaluation of the objects to be archived. Here, too, selection is inevitable. It must also be understood that the basic premises also mean that not all wishes of the designated communities can be fulfilled.

- **Authenticity:** The costs of preserving objects in libraries, archives and museums can only be justified if users can trust the information offered, i.e. the objects are recognised as authentic. The authenticity of digitally stored information cannot be judged from the carrier material of the objects in the same way that paper-based or parchment documents can. Other criteria and characteristics must be found instead. With regard to authenticity, the guideline uses the concepts described in the nestor "Trusted Digital Repositories" criteria catalogue and DIN standard 31644 "Criteria for trusted digital long-term repositories".

- **Adequacy:** The adequacy dimension describes the relationship between the objectives of the preservation repository and their specific implementation. The

objectives, designated community and the objects to be preserved can differ from repository to repository. Meanwhile, the range of technical solutions is continually expanding. It is not therefore possible to describe the ideal preservation repository or the optimum technical solution in abstract terms. Rather, each preservation repository must establish which properties it regards as significant, and which not. The guideline provides support and a framework which can be filled in different ways. In doing so it is itself adhering to the principle of adequacy described above.

- **Potential for automation:** The number of objects to be ingested requires the automation of as many processes as possible. By describing suitable processes and orienting itself to the overriding goal of automation, the guideline also makes a contribution to the development of suitable software products. The established principles and criteria can serve as the starting point for formulating special requirements which serve in turn as the preconditions for software development.

Terminology

Archive

The starting point here is a digital repository operated by a memory institution. According to the OAIS, this digital repository is a clearly defined unit from a conceptual and organisational-operational perspective and is also discriminated as such by the *information producers* and also by the *users*.

Object

Objects can only be preserved if they are discrete, i.e. their scope is clearly defined. For digital preservation it is therefore necessary to create such objects both in the physical and logical domains. The physical units (*data*) are combined to form a *representation*; the counterpart to this is an *information object*. However it is possible that the stored data don’t have a direct one-to-one correlation to the form of an information object.

Performance

The ingested data are processed by hard- and software and transmitted to a reproduction device. This is used to generate a performance, e.g. on a monitor or via loudspeakers, i.e. the information is rendered intelligible to the human senses.

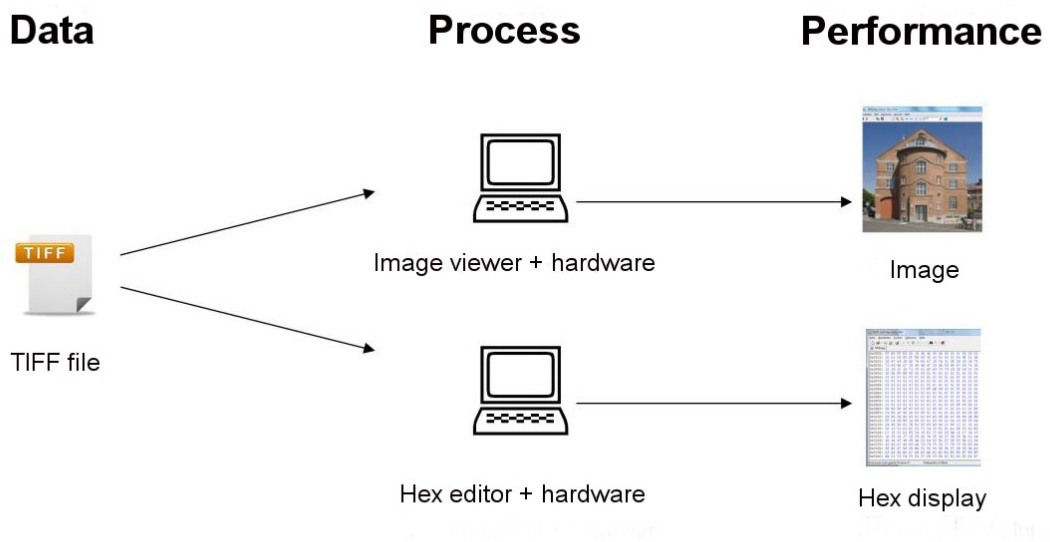

Figure 2: Performance Model

Information object

An information object is a logically discrete unit of information. It is created when a person intellectually processes the sensory information resulting from a performance. It is thus generated in his/her mind.

Information type

In order to preserve millions of digital archival documents, institutions first need to divide them into groups of the same kind of object (e.g. text, images, ...). The objects then belong to a particular information type and have more or less the same properties (e.g. for text: linear arrangement of alphanumeric characters), although the actual values differ (e.g. "This is where book 1 begins..." in contrast to "Article 2 starts as follows ....").

Preservation group

Subset of an information type. This contains information objects, the same properties of which are to be preserved (=significant properties). The information objects of a preservation group can therefore be processed together and preserved using the same processes. This represents a precondition for preserving large quantities of digital objects.

Significant properties

A selection has to be made because, over time, it will not be possible to preserve all of the properties in the ingested representations. This subset of all properties is the significant properties. In some cases this selection needs to be made before ingest. The significant properties should be preserved throughout the archiving process. Up to a certain extent, the authenticity of the performances obtained from the migrated or emulated representations can be ascertained from these.

Further terms are explained in Appendix B Glossary.

2. Procedural model

2.1. Initial entry

The first stage of the initial entry process is based on how the archivist perceives a hitherto unknown object. The second stage also includes the subsequent evaluation process. This process is described below in ideal-typical form to facilitate better understanding. In practice, knowledge is already available in most cases on the information object, information type and preservation group.

Activities

- A performance is created by the interaction of data, hardware and software.
- The performance is perceived via the senses of the archivist, and is then intellectually processed. S/he forms an impression of the information object.
- The archivist determines the preservation group and the significant properties. The values are documented as metadata.
- Simultaneously, the archivist also evaluates the information object; the object is partly or wholly ingested, or rejected entirely.
- The preservation repository ingests the information object, the metadata and possibly also the hard- and software.

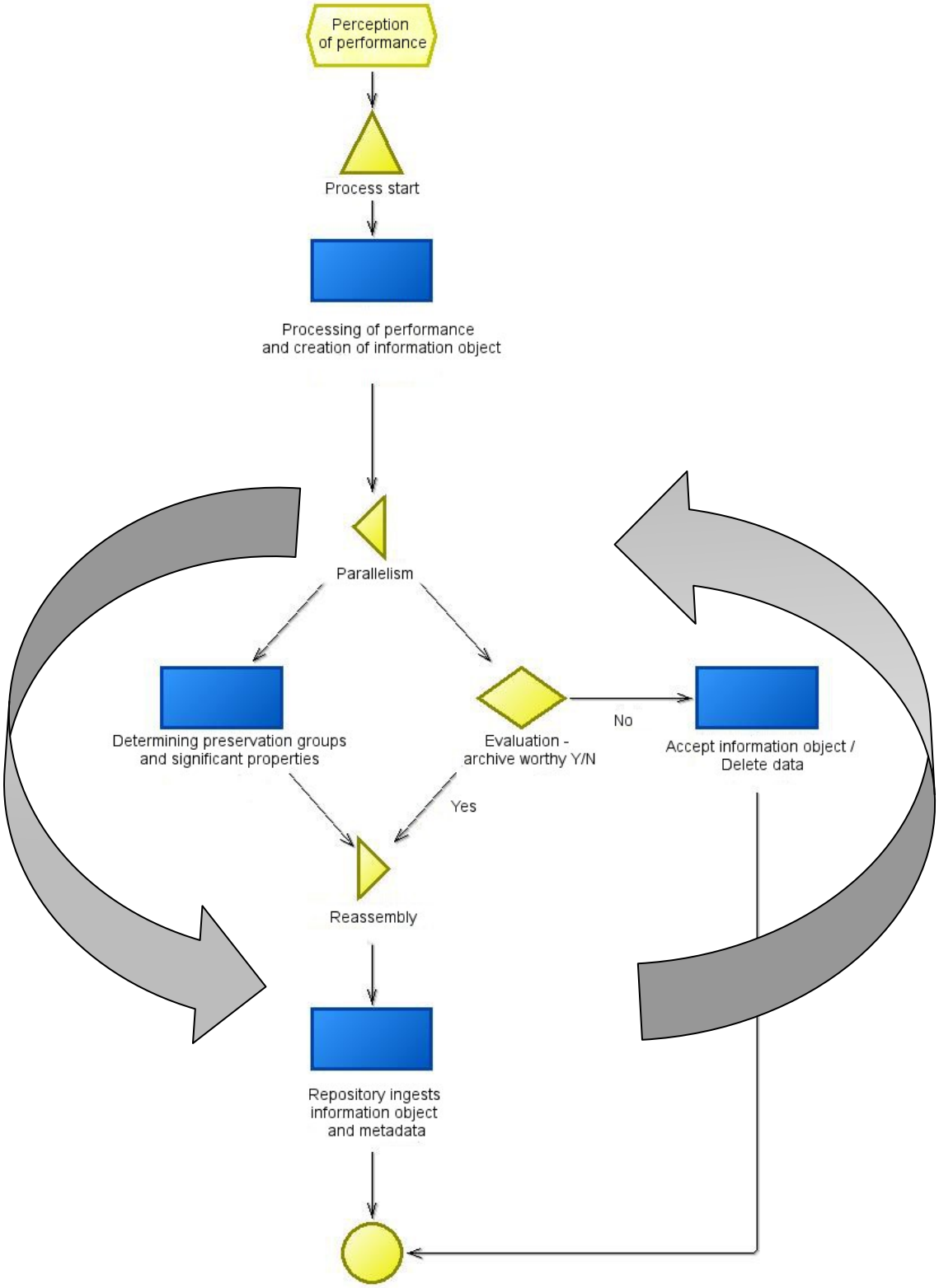

Figure 3: Initial entry

Descriptions of the individual activities

The starting point for digital archiving is the data offered to, or to be ingested by, the preservation repository. The purpose of preservation planning, however, is not to preserve the data itself, but rather the long-term preservation of the stored information. This information is referred to by the term intellectual entity (information object) in the PREMIS system. Its material manifestation is called a representation.

A complex perception and creation process takes place between the data being offered and the markup and archiving of an information object. The steps involved can be taken in a different order to that described below.

Creation of the performance

People cannot perceive or understand digital data without the use of additional equipment. Only through the interaction of the data with hard- and software can the stored information be reproduced on an output device (monitor, loudspeaker, printer) in such a form that humans can perceive it using their senses. Data plus hard- and software therefore constitutes a performance. The performance is the starting point of the considerations and decisions surrounding preservation planning.

Purely in terms of preservation planning it would suffice to preserve the performance indefinitely. The performance represents the yardstick by which future use must be measured. It indicates the essential features. For example, if a text is to be archived, it is important that it can be displayed correctly on a monitor. It is only of secondary importance whether the underlying data are stored in ASCII or in Unicode 8 character format. The same applies for the software used for the performance.

However, because the current technological state-of-the-art makes it impossible to preserve the combination of data and hard-and software indefinitely, digital preservation planning measures need to be taken which can help yield a similar performance. This is commonly achieved by converting the data, the goal of which is to reproduce it using new hard-and software (migration strategy). An alternative is to preserve both the data and software so that they can be played back using additional programs (emulators) on hardware in an

altered interpretative environment. In both cases the new performance must be compared to the first performance.

Processing of the performance and creation of the information object

A performance does not, however, directly define a discrete information object. Instead, this is the result of an interpretative process. First, the archivist performs a conceptual and intellectual act in order to combine the perceived sensory impressions in his or her mind. This creates the impression of one or more information objects.

Only discrete digital objects can be preserved indefinitely. If the data are not organised in a discrete form and are provided e.g. as a data stream, the archivist is also responsible for delimiting the information object. For example, the output of a radio station broadcasting 24 hours a day can only be archived if the archivist defines clear starting and finishing points for the object being archived.

Determining the preservation group and significant properties

The information object possesses a number of properties. A subset of these must be preserved for future users. These are the significant properties of the information object which the archivist must expressly determine. S/he can make use of the concept of preservation group (see section 2.2) in order to handle large quantities of digital objects.

Evaluation of the information objects

The information objects are not generated in a vacuum in the archivist's mind. Their evaluation depends on the options which these objects would offer future users. The range of options is limited by the significant properties of the objects under evaluation. Designated communities, designated uses and significant properties are first identified for the selection of the objects being archived. The selection of objects to be preserved and suppositions about, or definitions of, designated communities and designated uses are inseparably linked and are mutually dependent.

Further information

In most cases an archivist cannot preserve all the data and information offered. The same applies for the properties of the information object. Properties which did not count as significant properties may however be preserved (possibly by chance). Their preservation is not, however, a necessary precondition for migration or the selection of a suitable emulator.

Digital data can often be accessed using different programs and different hardware. Different performances may be created by processing the data in different ways. If required, the archivist can also develop his/her impression of an information object on the basis of different performances.

In some cases the information object is reproduced in full by the performance (e.g. photos). In other cases the performances only offer a selection of the usage possibilities. Databases provide an example of this (structured information). In the case of large databases, the archivist can only perceive a small proportion of the many usage possibilities and displayable values. Instead, what is displayed on the monitor merely provides an indication of how the database can be used. His or her impression of the information object to be archived is correspondingly abstract as a consequence. Yet the starting point even of this information object lies in the sensory perception of the performance.

Over time, the possible performances may alter due to changes in the data (representations) or the hard-and software. At a later juncture a future archivist may change the allocation to a preservation group on the basis of the performances now available (thereby determining further significant properties).

2.2. Determining the preservation groups

Section 2.1 pointed out that the archivist assigns information objects to preservation groups. These preservation groups are not stipulated externally, rather they, too, have to be defined by the archivist. An ideal-typical version of the process is described here. The archive can base the initial definition of preservation groups on its existing holdings or on holdings which it will ingest in the future. Within these holdings the preservation repository determines groups of similar objects which are to be preserved in the same way (preservation groups). Once defined, preservation groups are not static; they can be refined ad infinitum.

Activities

- The preservation repository names the information types it is to preserve (e.g. images, text, sound recordings).
- The possible designated communities of an information type are named.
- Conceivable designated uses of the information type are specified.
- The grouping of information objects with the same designated communities and designated uses results in preservation groups which are subsets of the information type. The designated communities for this group select the designated uses.
- The subset of significant properties is determined for a particular preservation group on the basis of the designated uses.
- The level of compliance of the individual significant properties is defined.

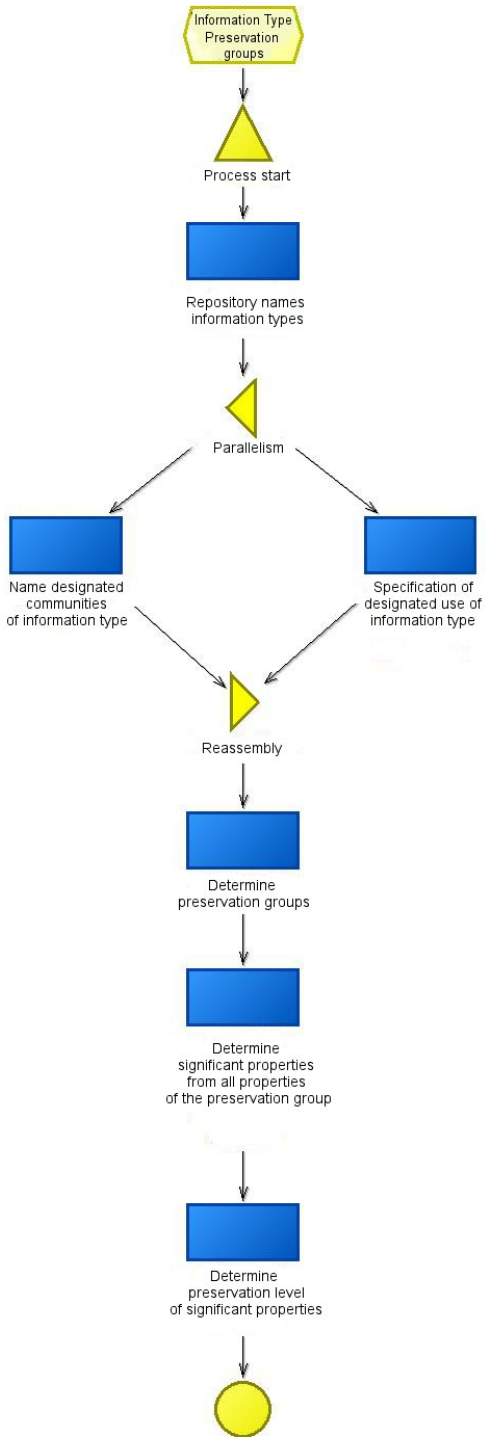

Figure 4: Determining the preservation groups

Descriptions of the individual activities

Most archives have to preserve large quantities of digital objects. In such cases it is not practicable to create a separate preservation plan (see the premises described in section 1). Similar kinds of objects should therefore be grouped together where possible and preserved using the same processes and tools. Such grouping permits efficient processing of the individual objects and is a necessary prerequisite for automated preservation planning procedures (e.g. migrations of large quantities of digital objects). In this process a category (information type) based on the archive documents is specified based on two categories which are independent of the objects (designated communities and designated uses). A subset of the information type and therefore a preservation group can be determined as a result.

The purpose of archiving is to preserve the information so that it can be perceived and processed by people. The starting point of the preservation therefore lies neither in the data nor in the performance. The performance as perceived by humans, or more precisely the information object represented by it, must serve as the starting point for the preservation work rather than a technical phenomenon. This must serve as the basis for the grouping.

Why are file types insufficient?

Certain preservation planning methods are based on individual file types. This approach is inadequate for various reasons. First of all, a file cannot be permanently preserved in its original hard- and software environment. The central point of reference of a preservation planning method should, however, lie in what can and must be preserved. Secondly, the archival evaluation refers not to file properties, but to what people perceive as information, i.e. it concerns the performance. According to the OAIS definition of digital archiving, the information itself should be preserved beyond the life time of the hardware and software. Under this definition, the physically stored files can be counted as hardware, and their file formats as software. Thirdly, files and their formats play an important role in creating the performance, yet this also needs to be interpreted by the hard- and software before a performance is created which can be perceived by people. The performance can change if the hard- or software is replaced. Here, too, the file offers no clear point of reference for what is to be preserved in the long-term preservation process.

Naming information types

Which information types an archive defines for its documents represents a fundamental decision. This divides the totality of the objects into more manageable subsets. The following information types are suggested as a possible structure:

- Text
- Images (2 or 3 dimensional)
- Audio
- Moving images
- Structured information (tables ...)
- Software

The information type "Moving images" can be regarded as a container which is based on the information types of image and audio but which also contains further properties (e.g. the running time of the film). Other conceivable containers include e.g. files, e-mails or websites. These containers can represent information types of their own into which other information types can be inserted.

Information types are too intangible as to represent a group of the same kind of archival documents, i.e. a group of the same kind of archival documents for preservation. Instead, subsets (preservation groups) need to be created by defining designated communities and designated uses.

Naming the designated communities of the information type

The designated communities provide a basic framework. These designated communities should not, however, be confused with the current users of the archive. The designated community describes the future users which the preservation repository assumes will use its services. Users, on the other hand, are those who actually come into the preservation repository at present. Each preservation repository must establish the level of detail in which it describes its own designated communities (e.g. "historians" or "social historians" or "social historians of the late 20th century"). The description of the designated communities can be given e.g. in a policy. The following aspects can be included:

- Expectations of the archive
- Prior specialist and technical knowledge of the designated communities
- Technical equipment of the designated community
- Legal restrictions (e.g. archive laws, copyright)
- Size of designated community
- Frequency of use

The properties of a designated community may change over time as the result of the actual use. The preservation repository must ensure that it registers and documents these changes.

Specifying the designated uses of the information type

The designated uses are defined in more concrete terms than the designated communities. They include the functional possibilities for using information objects. They are not related to content-based issues meaning that it is better to give e.g. "Statistical utilisation" as the designated use rather than "Search for distribution of women and men". It is, however, up to each individual archive to decide how detailed it wishes to describe its designated uses. The list of possible concrete designated uses can be derived from four abstract designated uses:

- Perceiving the complete object, e.g. viewing a picture or film, listening to an audio stream
- Evaluating/obtaining information, e.g. searching for an individual piece of information or a complex of information within an information object.
- Further processing after extracting a use package, e.g. describing a text or using a film clip for a documentary
- Running the object, e.g. playing a computer game

Different designated uses can be assigned to the objects of an information type, e.g.: research data for new analysis or for proof.

Determining preservation groups

The preservation groups are identified on the basis of a group of the same kind of information objects. Following their initial description, their designated uses and significant

properties can be refined ad infinitum and specified through the first entry of further information objects (see examples in Appendix A).

A preservation group contains all conceivable designated uses for a group of the same kind of objects and the object properties (=significant properties) needed to achieve these. The individual information objects are only ever assigned to one preservation group. If the same information is to be assigned to different preservation groups for different designated communities or designated uses, then each individual preservation group should have its own information objects. Because of their being assigned to different preservation groups, the information objects of such a representation are also treated differently; this enables them to be handled as distinct administrative units. The information objects of a preservation group share basic similarities as a result of their common information type, and they also serve the same designated uses. They therefore constitute a group and are preserved in the same manner. It is conceivable for a preservation repository to form preservation groups for different information types which have the same significant properties. For example, the decision could be made only to preserve the existing layout of certain forms of abstract poetry. Although these objects can still be counted as the text information type, they could also count as a preservation group with the same properties as an image information type preservation group. In this case, the preservation repository can decide whether it wishes to combine the different preservation groups to form a new preservation group.

Determining significant properties

The significant properties are a subset of all the potential properties of a particular information type. They relate to the first performance (e.g. the length of the first performance of an audio object) or the information derived from it (e.g. database). The significant properties determined for a specific archival document should be preserved at all costs during the preservation process.

The function of the significant properties becomes clear during the course of the preservation process. If a performance changes as the result of a migration or an emulator, the significant properties of the old performance can be compared to those of the new performance. For example, if an audio object saved in file format A is 11 minutes long, it should also be 11 minutes long in file format B. Determining the individual values can either

be carried out automatically or by an employee and can either include all objects of a preservation group or merely a sample of them. Significant properties therefore permit evaluation of a performance in a certain environment. If the significant properties are preserved to the extent stipulated above, then the information objects will also be authentic in their new performances.

The selection of significant properties can be changed (any such changes should be documented). Newly defined designated uses can render the significance of certain properties obsolete. If the preservation process is extended, this can result in new properties being created during the preservation process. The preservation repository must decide whether a new information object has been created or whether it counts as the successor object of the "original object". This also involves comparing the authenticity and usability of the demands. For instance, the scanning of a book and the resulting storage as a TIFF file largely preserves the original designated use of reading. Subsequent OCR processing would give rise to possibilities which the paper-based version did not possess.

The significant properties should be preserved in full in the archiving packages. They do not, however, need to be automatically contained in all use packages in which individual significant properties can be dispensed with, depending on the requirements. In this regard, archiving packages represent the sum of all possible use packages.

Determining the level of compliance

The significant properties can sometimes be fulfilled entirely (e.g. the number of data records) and sometimes only partially (e.g. preservation of colour space). The preservation repository must therefore also define the desired level of compliance.

3. Definition of representation information

Assuming the data of an information object are present, different types of additional information are needed to generate and understand the performance and all its inherent significant properties. This additional information is termed "representation information" in the OAIS. It is mostly saved as metadata. Determining the representation information and its ongoing maintenance is one preservation element.

The representation information is always related to a specific time (today, tomorrow, in one hundred years) at which the understanding of the user must be obtained. However, the significant properties to be distinguished between cover an open period in the future; they provide arguments for why it is advisable to go from state A to state B but not to state C. The representation information is also based on representations, significant properties and information objects.

Two basic types of representation information can be distinguished between. Additional information is required to create the performance from the data in a technical environment. The simplest examples of this include naming the format to allow the right technical reproduction environment to be selected, or providing code lists for translating the bit sequences into a character format (e.g. ASCII). This type of representation information is called Structure Information in OAIS. In practice it is well protected by metadata standards such as the PREMIS metadata elements for object characteristics and environment.

Also, information is required to help interpret the performance itself, such as information on the context and the parameters of a scientific experiment which allows the measured data obtained to be evaluated. Without this additional information, an otherwise authentic performance of an information object may be unusable. In OAIS this type of representation information is called semantic information. Part of this semantic information can be covered by descriptive metadata standards. However, descriptive metadata are often only used for finding information objects, whereas the use and interpretation can require additional information which normal search criteria cannot encompass (e.g. measurement units of a set of research data).

Both types of representation information must be determined relative to the designated community, its given specialist and technical knowledge and its technical equipment. Different technical and material aspects are necessary or require clarification for different

designated communities, contexts or times. Representation information can also become obsolete if the designated community or the assumed knowledge changes to such an extent that the information object can no longer be used.

The necessity of providing representation information gives rise to three preservation planning tasks:

- Understanding and documenting the designated communities in a way which identifies which assumed knowledge and which equipment they possess (see also section 2.2 "Determining the preservation groups");
- Determining the difference which can exist between the requirements of the designated community and the information required for use. If a knowledge gap exists which cannot be bridged by applying a reasonable amount of effort or by the designated community itself, then the additional required information must be stored along with the information object or externally referenced. It is not, however, always easy to determine what other people take for granted and which assumed knowledge and equipment they possess. If the archivist him or herself is not a member of the designated community, one possible solution is to collaborate with people from the designated community on this task (see also section 5.1 "Community watch"):
- Continuously monitoring the changes within the designated community to see whether additional representation information must be saved along with the information objects. (This task is described in further detail in section 5 "Monitoring tasks".)

4. Preservation strategies

Different strategies have been developed in recent decades for preserving digital information objects. Currently most widespread are migration and emulation strategies which can be used independently of, or in combination with, each other.

4.1. Preserving information using the migration strategy

The migration strategy aims to preserve the information by migrating files to new file formats. This means that the performance is now based on the new migrated files. There is a risk that this can change both the performance and also the information object.
Migration is a response to the realisation that the files currently used for the performance either have file formats which are unsuitable for archiving or that these file formats are at risk of becoming obsolete, i.e. are not supported by the latest software.

This necessitates the following activities:

- Assessment of the initial entry and decision regarding which significant properties should be preserved after the migration.
- Selection of suitable and long-lasting file formats for the preservation group. The specific characteristics of the potential file formats need to be taken into consideration (migratability, open documentation, no licence, widespread distribution, stability etc.).
- Migration of the files to new file formats. The new files are combined to create a new performance.
- Assessment of the new performance.
- Validation through comparison of both assessments.

The complexity of the migration depends on the file format used and on the scope and variety of the defined preservation groups; this has a direct impact on costs, compatibility and interoperability. The authenticity of the migrated data can be ascertained by comparing the new performance with the original one. It is deemed to be authentic if the significant properties of both performances concur to the stipulated level of compliance. The preservation repository should keep the results of this validation (e.g. in the form of a

report). At present, most archives still store older representations, especially the ingested representations.

4.2. Preserving information using the emulation strategy

The emulation strategy should ensure that the information files remain readable through the timely programming of an emulator, i.e. the technical environment of the computer changes after the emulation, whereas the performance remains largely preserved within the conditions of the emulator. It is assumed that the ability of the new computers arriving on the market to open and process the software currently used for the performance will gradually decline. An emulator running within a modern hard- and software environment should be able to generate a new performance from the original files and programs, the significant properties of which match those of the original performance. "Encapsulation" is an especially suitable preparation method for emulation. The file or the information object to be preserved, the software which is used to visualise and reproduce it, plus related metadata are also saved in an archival package.

This necessitates the following activities:

- Assessment of the significant properties of the first performance
- Selection, test and use of a suitable emulator for the preservation group.
- Assessment of the new performance.
- Validation through comparison of both assessments.

The complexity of the emulation can increase with each new hardware environment because the distance to the "old" software being emulated will increase along with the accelerating technological development of future generations of computers. The authenticity of the new performance can be ascertained by comparing it with the original performance. It is deemed to be authentic if the significant properties of both performances concur to the stipulated level of compliance. The preservation repository should keep the results of this validation (e.g. in the form of a report).

5. Monitoring tasks

In order to apply the preservation strategies it is necessary to record when a preservation task or a representation information update must be carried out. The time depends on the changes in the designated community which must be monitored by a "community watch" and by the general technological developments which need to be monitored by a "technology watch".

5.1. Community Watch

The designated communities with their specialist and technical requirements and possibilities are the main point of reference in determining whether information objects can still be used or whether they risk becoming unusable. Even if a format is very widespread and is supported by a large number of programs which authentically reproduce the significant properties, preservation measures may be needed because the designated community does not or cannot use this technology. An example of this would be a designated community with very strict security requirements which does not use common word processing software and formats due to the corresponding document formats being susceptible to viruses. And conversely, preservation measures such as the updating of representation information or format migrations do not need to be effected, even in the case of uncommon or complex technical and content-related requirements, if they are fulfilled by the designated community using well-established instruments. This may occur in specialist academic fields which develop and maintain their own formats and the corresponding application programs.

It is not sufficient, however, to determine on the basis of a single analysis whether a designated community can use the information objects authentically. Digital repositories must take into account the fact that further designated communities may be added and others may disappear, and even the designated communities themselves may change their requirements and possibilities. It is therefore necessary to check at regular intervals whether such changes have occurred in the designated communities. This task is called "Monitor Designated Community" in OAIS. Possible implementation methods include carrying out annual interviews, surveys or workshops with representatives of the designated communities, and also receptive procedures such as participating in events of the designated communities or the targeted evaluation of user inquiries and requests.

5.2. Technology Watch

Besides the requirements and abilities of the designated communities, general technical developments and possibilities also need to be monitored and evaluated on a regular basis. This is referred to as "Monitor Technology" in OAIS. Monitoring technologies and standards helps to provide advance warning of information objects which are no longer usable and also to highlight possible new technical uses. Especially in the case of designated communities which are highly diffuse and difficult to define it may be important to monitor the general distribution of hard- and software environments which can be used together with the data in order to generate an authentic performance. A large number of criteria have been developed for evaluating file formats, however these are only effective for formats which provide the same use possibilities. The PLANETS project has described a more straightforward technology evaluation procedure for long-term preservation which is of relevance here in terms of systematic preservation strategy planning.

Appendix A: Possible properties, designated communities and designated uses of information types

Appendix A contains several examples of how preservation groups can be obtained from the information types. The exemplary character of these needs to be emphasised for various reasons: Firstly, each individual preservation repository determines which information types it uses. Secondly, the designated communities and designated uses can differ from repository to repository. Thirdly, the working group is assuming that the proposed level of compliance may be more or less differentiated depending on the preservation repository. Finally it would also be conceivable to draw up a list of priorities instead of the level of compliance.

The following information types are given as examples:

I. Text

II. Image (two dimensional)

III. Audio

IV. Moving images

V. Structured information

VI. GIS data

VII. Software

I. Text

Definition

"Texts" are discrete content in written language, expressed by a linear sequence of alphanumeric characters. The term text (see 5 below) also includes other possible text elements (pictures, tables).

Possible designated communities (sample list)

Z1: **Readers**
- Z1.1: Those interested in literature
- Z1.2: Those interested in (book) art

Z2: **Academics**

- Z2.1: Historians
- Z2.2: Social scientists, statisticians
- Z2.3: Linguists
- Z2.4: Those interested in the history of paper, writing, printing or typography

Z3: **Those interested in history**

- Z3.1: Local historians
- Z3.2: Genealogists

Possible concrete designated uses

N1: **To perceive the entire object**

The information object should be perceived by the senses as a work of art in the broadest sense, but one with no intention of conveying information.

- N1.1: Direct perception of the visual impression of the text, text document (or parts thereof) as a work of art
- N1.2: Reading the text as a literary work

N2: **To extract and analyse information**

Individual or multiple pieces of information are to be extracted from the information object.

- N2.1: Search for certain information units which can be defined and described in advance (e.g. a specific date)
- N2.2: Search for a new realisation which is obtained by analysing and combining a number of different information units
- N2.3: Search for evidence (e.g. way in which an author or an official body works)

N3: **To process information**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

- N3.1: Calculation from the information of an information object (e.g. statistical analysis of rows of figures in a text document)
- N3.2: Compilation of information from multiple information objects (e.g. writing a chapter on the biography of an author on the basis of his or her correspondence)
- N3.3: Further use of part of or the entire information object for illustrative purposes outside the context of the archive (e.g. use of a certificate as illustration, see N1)

Possible properties of texts in narrower sense

Texts in the narrower sense contain a linear sequence of alphanumerical characters.

E1: **Order and completeness of characters**

Potential level of compliance Full / not to be preserved

Test method Conceivably automatic (character count, content analysis), otherwise manual checks (samples)

E2: **Automatic processability**

The individual parts of the text (sections, words, characters) should be extracted by a computer and be capable of being changed as required (i.e. as alphanumeric characters and not only as graphic). Processability also includes full text searches.

Potential level of compliance Preserved / not preserved

Test method Human assessment, can be derived in some cases from file formats.

E3: **Markup and representation of certain characters or strings**

The specific function of certain characters or strings in the text is highlighted by different markups. Markup of different structural levels (2-3 title levels), markup of different semantic levels (bold, italic, different font sizes). These all help to preserve the layout.

Potential level of compliance Structural / semantic levels can be logically distinguished between (and can therefore be automatically processed); the levels can only be distinguished between visually; markup is not preserved.

Test method Conceivably automatic

E4: **Sufficient referenceability or addressability of certain characters or strings**

A reference to a certain point in the previous version of a text must identify the same point in the new version. The same applies for internal references within the text (especially footnotes)

Potential level of compliance Preserved / not preserved

Test method Manual check (page breaks, footnote numbering, illustration numbering)

E5: **Distinguishability of other (in some cases hidden) text layers**

Hidden text, comments, revision markings are distinguished between. It is assumed that the principal layer is accessible.

Potential level of compliance Preserved in each case as separate layer / preserved as a separate layer to the principal layer / integrate in the principal layer / not preserved

Test method Manual check

Possible properties of texts in broader sense

Texts in the broader sense include the linear sequence of alphanumeric characters, but also further elements such as pictures or tables.

E6 **Preservation of additional elements**

Additional elements such as images or tables are preserved.

Potential level of compliance Preserved / not preserved
Test method Manual check

E6 **Location of additional elements**

Additional elements such as images or tables are reproduced at the correct point in the text.

Potential level of compliance Preserved / not preserved
Test method Manual check

E8: **Separate processability of additional elements**

Additional elements such as images or tables can be extracted and processed separately to the rest of the text.

Potential level of compliance Preserved / not preserved
Test method Manual check

Note: If images are accorded the test level "Preserve", it may be necessary to take properties from the image category into consideration. The same applies for tables and other elements.

Possible matrix of texts in narrower sense

	E1 Sequence	E2 Process-ability	E3 Markup	E4 Address-ability	E5 Other text layers
N1.1 Visual perception	X		X		
N1.2 Literary enjoyment	X		X		
N2.1 Search for fragmented items of information	X	X		X	X
N2.2 Realisations obtained by combining information	X	X	X	X	X
N2.3 Search for evidence	X		X		X
N3.1 Calculation from information	X	X			
N3.2 Compilation of information	X	X	X	X	
N3.3 Re-use for illustrative purposes		X	X	X	

X = The property should be fulfilled in full or in part for this designated use.

Two examples of determining a preservation group

Example 1

The information objects of the preservation group are to be used by historians and those interested in literature (designated communities Z1.1 and Z2.1). Probable designated uses are: N1.2 (enjoying the text as a literary work) and N2.1-2.3 (gleaning and analysing information).

	E1 Sequence	E2 Processability	E3 Markup	E4 Addressability	E5 Other text layers
N1.2 Literary enjoyment	X		X		
N2.1 Search for fragmented items of information	X	X		X	X
N2.2 Realisations obtained by combining information	X	X	X	X	X
N2.3 Search for evidence	X		X		X
Significant properties	X	X	X	X	X
Degree of implementation	Full	Preserved	Automatic process-ability	Preserved	Preserved as separate layer

Example 2

The information objects are to be preserved for designated community Z1.2, those interested in (book) art

	E1 Sequence	E2 Processability	E3 Markup	E4 Addressability	E5 Other text layers
N1.1 Visual perception	X		X		
N1.2 Literary enjoyment	X		X		
Significant properties	X		X		
Degree of implementation	Full	Not preserved	Visual representation	Not preserved	Not preserved

II. Image (two dimensional)

Definition

An image is a two-dimensional, static representation of visual information. An image is determined by the fixed composition of colours.

Possible designated communities (sample list)

Z1: **Viewers**
- Z1.1: Those interested in arts

Z2: **Academics**
- Z2.1: Historians
- Z2.2: Social scientists
- Z2.3: Art historians / those interested in photographic history
- Z2.4: Scientists

Z3: **Those interested in history**
- Z3.1: Local historians
- Z3.2: Genealogists

Z4: **Professional users**
- Z4.1: Journalists/publicists
- Z4.2: Commercial companies (e.g. advertising)
- Z4.3: Artists/fashion designers
- Z4.4: Educationalists

Possible concrete designated uses

N1: **To perceive the entire object**

The information object should be perceived by the senses as a work (of art) in the broadest sense, with no intention of providing specific information.

- N1.1: Direct perception of the visual impression of the image (e.g. photo or painting) (or of parts thereof) as a work (of art)
- N1.2: Use of the picture as a decorative object (e.g. wall decoration)

N2: **To extract and analyse information**

Individual or multiple pieces of information are to be extracted from the information object.

- N2.1: Search for a specific piece of information which can be defined and described in advance (e.g. the presence of a particular building on a landscape photo, the extraction of information from a chart or map, e.g. representation of terrain on aerial photographs)
- N2.2: Search for a new realisation obtained by analysing and combining a number of different information objects.
- N2.3: Search for evidence (e.g. research into the artistic techniques of a painter or photographer within history of painting or photography, working methods of a cartographic workshop)

N3: **To process and forward information**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

N3.1: Further use of a part of or the entire information object for visual or illustrative purposes in a new context (e.g. pictures to accompany a magazine article, background picture in a collage, artistic processing, commercial use for advertising or design purposes)

N3.2: Use of the information object as a means of conveying information in a new context (e.g. photo as source in an historical exhibition or a text book)

N3.3: Further processing of an image using technical tools (e.g. changing the picture using an image processing program, selection of individual layers)

N3.4: Issue of a physical representation of the information unit on a specific medium (print, hard copy, photocopy)

Possible properties of images

Content

E1: **Content integrity (e.g. no retouching, montages, cropping etc.)**

No deliberate changes have been made to the image (e.g. retouching, montages); no unintentional changes have been made (e.g. artefacts). No section of the image has been selected.

Potential level of compliance Preserved / largely preserved (not discernible to the human eye) / not preserved

Test method Automatic method is problematic (image recognition software is not yet powerful enough; characters can only be counted if file format is identical), otherwise manual checking (samples - manual here means visually in the sense of visual inspection)

E2: **Preservation of the metadata added by the copyright holder, processor or system**

Label, title, file name, date of storage etc. are preserved.

Potential level of compliance Preserved / partially preserved / not preserved

Test method Automatic, manual (samples)

Form, size, orientation

E3: **Form of image**

The form (rectangular, irregular edges) is preserved; most digital images are rectangular in form (in certain formats the visual information can, however, take on other forms)

Potential level of compliance Preserved / not preserved

Test method Manual check (inspection of digital images in an image processing program)

E4: **Image size and resolution**

Of prime relevance are the image size intended or specified by the producer (e.g. in cm) and the resolution (the dpi or number of pixels per cm match)

Potential level of compliance Preserved / not preserved
Test method Automatic

E5 **Orientation of picture**

This property states where the top of the picture is.

Potential level of compliance Preserved / not preserved
Test method Automatic method is problematic (image recognition software is not yet powerful enough); otherwise manual checking (sampling)

Colour and representation

E6: **Colour, accuracy of shade, including brightness**

Colour or black and white picture? A change in the colour space (e.g. RGB, CMYK) or a reduction in the bit depth of the channel causes loss of information.

Potential level of compliance Preserved / not preserved
Test method Automatic

E7 **Composition of colours in two-dimensional space**

The property is reflected in the sequence of the pixels.

Potential level of compliance Preserved / not preserved
Test method Automatic method is problematic (image recognition software is not yet powerful enough); otherwise manual checking (sampling)

E8 **Transparency of image areas**

Certain image formats permit transparent or translucent sections.

Potential level of compliance Preserved / not preserved
Test method Automatic method is problematic (image recognition software is not yet powerful enough); otherwise manual checking (sampling)

E9 **Scalability of the image**

Even the smallest image details can be displayed distinctly. Whether or not the image is available as a raster or vector file may be relevant for preserving this property.

Potential level of compliance Preserved / not preserved
Test method Automatic

Layers

E10 **Separation of the image into layers**

The layers are preserved in certain file formats (e.g. Photoshop or Gimp)

Potential level of compliance Preserved / not preserved
Test method Automatic for digital images, otherwise manual checking (sampling)

E11 **Order or visibility of the layers**

The order of the layers and the information on whether the layer is deactivated or not are preserved.

Potential level of compliance Preserved / not preserved
Test method Automatic

E12 **Processability of layers**

Each individual layer of the image can be processed and is not "frozen" in a complete image.

Potential level of compliance Preserved / not preserved
Test method Automatic for digital images

Possible matrix for images

	E1: Integrity	E2: Metadata	E3: Form	E4: Size / resolution	E5: Orientation	E6: Colour mode	E7: Composition of colours	E8: Transparency	E9: Scalability	E10: Separation of layers	E11: Use of layers	E12: Processing of layers
N1.1 Visual perception	X		X	X	X	X	X					
N1.2 Decorative object	X		X	X	X	X	X					
N2.1 Search for fragmented items of information	X	X	X	X	X	X	X		X	X	X	X
N2.2 Realisations obtained by combining information	X	X	X	X	X	X	X	X	X	X	X	X
N2.3 Search for evidence	X	X	X	X	X	X	X		X	X	X	X
N3.1 Visual or illustrative further use	X		X	X	X	X	X					
N3.2 Further use for conveying information	X	X	X	X	X	X	X		X	X	X	X
N3.3 Processing using technical tools	X		X	X		X	X	X	X	X	X	X
N3.4 Issue of a physical representation	X		X	X		X	X		X	X	X	X

X = The property should be fulfilled in full or in part for this designated use.

Example of determining a preservation group: Digital work of art (image file)
The information objects of the preservation group are to be used by those interested in art, art historians and artists (designated communities Z1.1, Z2.3 and Z4.3). Probable designated uses are: N1.1 (perception as work of art), search for evidence, e.g. art history research (N2.3) and possible artistic processing of the information object, possibly with technical tools (N3.1 and N3.3).

	E1:Integrity	E2:Metadata	E3:Form	E4:Size / resolution	E5:Orientation	E6:Colour mode	E7:Composition of colours	E8:Transparency	E9:Scalability	E10:Separation of layers	E11:Use of layers	E12:Processing of layers
N1.1 Visual perception	X		X	X	X	X	X					
N2.3 Search for evidence	X	X	X	X	X	X	X		X	X	X	X
N3.1 Visual or illustrative further use	X		X	X	X	X	X					
N3.3 Processing using technical tools	X		X	X		X	X	X	X	X	X	X
Significant properties	X	X	X	X	X	X	X	X	X	X	X	X
Degree of implementation	largely preserved	Preserved	Preserved	Preserved	Preserved	Preserved	Preserved	Preserved	Preserved	Preserved	Preserved	Preserved

III. Audio

Definition

By "audio" we mean a fixed sequence of signals which are audible to the human ear. Audio events include language, music and other sounds.

Possible designated communities

Z1: **Listeners**
- Z1.1: Those interested in music as an art form
- Z1.2: Those interested in literature
- Z1.2: Those interested in language and speech

Z2: **Academics**
- Z2.1: Music and literary specialists
- Z2.2: Social scientists and ethnologists/cultural scientists
- Z2.3: Historians
- Z2.4: Biologists
- Z2.5: Technical scientists

Z3: **Those interested in history**
- Z3.1: Local historians
- Z3.2: Journalists
- Z3.3: Genealogists

Z4 **Media representatives**
- Z4.1: Journalists
- Z4.2: Publishers

Z5 **Music producers**
- Z5.1: Musicians

Possible concrete designated uses

N1: **To perceive the entire object**

The information object should be perceived by the senses as a work of art in the broadest sense, but one with no intention of providing information.
- N1.1: Direct perception of the acoustic impression of the audio object; audio object (or parts thereof) as a linguistic work of art
- N1.2: Enjoyment of music as a work of art
- N1.3: Experiencing literary recitation as a rhetorical work
- N1.4: Perception of sound (e.g. the flapping of a brimstone butterfly's wings)

N2: **To extract and analyse information**

Individual or multiple pieces of information are to be discerned from the information object.
- N2.1: Search for a certain discrete or elementary information unit which can be defined and described in advance (e.g. an interruption in a parliamentary debate: Joschka Fischer in the German Bundestag in 1984: "Mit Verlaub, Herr Präsident, Sie sind ein Arschloch")
- N2.2: Search for a new realisation which is obtained by analysing and combining a number of different information units

N2.3. Retrieval and analysis of specific mental and emotional manifestations (e.g. the delivery or rhetoric of a politician, certain interpretational techniques of a pianist)

N2.4. Search for special compositional elements

N3: **To process information**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

N3.1: Calculation based on the information of an information object (e.g. statistical analysis of the number of parliamentary interjections)

N3.2: Compilation of information from multiple information objects (e.g. writing a chapter on the biography of a politician or a composer on the basis of a speech or a musical performance)

N3.3. Further use of a part of or the entire information object for acoustic purposes outside the archive context (e.g. use of a speech in an exhibition, journalist writing e.g. a feature, use of a passage from a piece of music for advertising)

Possible properties of audio in narrower sense

E1: **Order and completeness of acoustic signals**

Potential level of compliance Full / not full

Test method Conceivably automatic (duration, number of acoustic signals etc.)

E2: **Automatic processability**

It should be possible to extract and modify individual parts of the audio recording (extracts, words, sounds) using a computer

Potential level of compliance Preserved / not preserved

Test method Human appraisal

E3: **Markup of specific speech or vocal passages or sequences**

The specific function of certain passages or sequences in the audio is highlighted by a different markup: Markup of different structural levels in speeches or pieces of music (2-3 title levels), markup of different semantic levels (volume, speed, comments by third parties).

Potential level of compliance Structural / semantic levels can be logically distinguished between (and therefore automatically processed); the layers can only be distinguished between acoustically; markup is not preserved.

Test method Conceivably automatic

E4: **Sufficient referenceability or addressability of certain speeches, pieces of music or sounds**

A reference to a certain point in the previous version of an audio must identify the same point in the new version. The same applies for references within the speech or the piece of music (especially the tracks)

Potential level of compliance Preserved / not preserved

Test method Manual check (tracks)

E5: **Distinguishability of different channels**

Background noise, recording quality, commentaries, processing tracks are distinguished between. It is assumed that the principal layer is accessible.

Potential level of compliance	Preserved in each case as separate layer / preserved as a separate layer to the principal layer / integrated in the principal layer / not preserved
Test method	Manual check/conceivably automatic

Possible matrix for audio in narrower sense

	E1 Sequence	E2 Process-ability	E3 Markup	E4 Address-ability	E5 Channels
N1.1 Acoustic perception	X		X		
N1.2 Musical enjoyment	X		X		
N1.3 Literary enjoyment	X	X		X	X
N1.4 Perception of sounds	X	X	X	X	X
N2.1 Search for fragmented information	X		X		X
N2.2 Realisations obtained by combining information	X	X			
N2.3 Search for specific emotional/mental manifestations	X	X	X	X	
N2.4 Rhetorical interest		X	X	X	
N2.5 Search for specific compositional elements		X	X	X	
N2.6 Specific interpretational techniques		X	X	X	
N3.1 Calculations from information		X	X	X	
N3.2 Compilation of information		X	X	X	
N3.3 Further use for acoustic purposes		X	X	X	

X = The property should be fulfilled in full or in part for this designated use.

Two examples of determining a preservation group

Example 1

The information objects of the preservation group are to be used by journalists and those interested in speech (designated communities Z1.1 and Z3.2). Probable designated uses are: N1.3 (experiencing the literary narration as rhetorical work) and N2.1-2.3 (gleaning and evaluating information).

	E1 Sequence	E2 Processability	E3 Markup	E4 Addressability	E5 Channels
N1.3 Experiencing literary recitation as a rhetorical work	X		X		
N2.1 Search for fragmented items of information	X	X		X	X
N2.2 Realisations obtained by combining information	X	X	X	X	X
N2.3 Search for evidence	X		X		X
Significant properties	X	X	X	X	X
Degree of implementation	Full	Preserved	Automatic processability	Preserved	Preserved as separate layer

Example 2

The information objects are to be preserved for designated community Z1.2, those interested in musical art

	E1 Sequence	E2 Processability	E3 Markup	E4 Addressability	E5 Channels
N1.1 Acoustic perception	X		X		
N1.2 Enjoyment of music as a work of art	X		X		
Significant properties	X		X		
Degree of implementation	Full	Not preserved	Visual presentation	Not preserved	Not preserved

IV. Moving images

Definition

"Moving images" refers to a fixed (order and speed of sequence) time-based sequence of "images". Such a sequence of images can, but does not have to be, accompanied by sound. In such cases, the sequence of sounds is synchronised with the sequence of images.

Possible designated communities (sample list)

Z1: **Viewers**
- Z1.1: Television viewers
- Z1.2: Cinema audience members
- (Z1.3: Gamers)

Z2: **Any individual seeking information**
- Z2.1: Consumers of news
- Z2.2: Viewers of documentary films

Z3: **Education / Instruction / Demonstration**
- Z3.1: Pupils (educational film)
- Z3.2: Adult education

Z4: **Documentation / research**
- Z4.1: Private individuals as producers (interested in documenting developments/situations/changes)
- Z4.2: Scientists (e.g.: long-term recordings, ethnographic films, experiments)
- Z4.3: Artists (documentation of performances etc.)
- Z4.4: Monitors (and their clients, secret services)
- Z4.5: Employees Instructions / visualisations

Z5: **Creative film artists**
- Z5.1: Film and video artists
- Z5.2: Trick film makers

Z6: **Producers / users**
- Z6.1: Companies as producers
- Z6.2: Companies as users

Possible concrete designated uses

N1: **To perceive the entire object**

The information object should be perceived by the senses as a work of art in the broadest sense, but one with no intention of providing information.
- N1.1: Direct perception of the visual (and acoustic) impression of the image (and sound) sequence as entertainment
- N1.2: Direct perception of the visual (and acoustic) impression of the image (and sound) sequence as appreciation (of art)
- N1.3: Use of the object (film, video) as emotional "memory support"
- N1.4: Use of the object (film, video) to generate profit

N2: **To extract and analyse information**

Individual or multiple pieces of information are to be discerned from the information object.

N2.1: Information on current affairs (e.g. news broadcasts/clips)
N2.2: Background information (e.g. documentary film)
N2.3. (Further) training (e.g. instructional films, film demonstrations)
N2.4. Use as evidence (e.g. ethnographic film about a dance of the Maasai or as storage of information for research, e.g. trajectory of projectiles)

N3: **To process and forward information**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

N3.1: Background for current affairs (clips of historical events in current news reports resulting in a new information object)
N3.2: Comparative sequential contrastive compilation for obtaining new realisations (e.g. collage of different clips of shaman rituals from different regions of the world to facilitate a direct comparison)
N3.3: Use of parts of a film as background for performance art
N3.4: Use of parts of a film (as "original quotation") for manipulation (advertising, propaganda, politics)

Possible properties of moving images/audio in narrower sense

E1: **Completeness of sequence of moving images and soundtrack**

The moving images can still be retrieved in full from the carrier (e.g. historical film)

Potential level of compliance Full / not full
Test method Visual/auditory check against existing object documentation

E2: **Integrity of complete moving image object**

The order of the images and sounds is exactly the same as originally intended

Potential level of compliance Identical / not identical
Test method Visual/auditory check against existing object documentation

E3: **Multiple soundtracks**

All soundtracks of the moving image object can be read and interpreted (e.g. for music and language or for multiple languages)

Potential level of compliance Full / partial / not interpretable
Test method Auditory check against existing object documentation

E4: **Synchronicity**

The reproduction of the sound and images is as intended at the time when the information object was created

Potential level of compliance full / partial / none
Test method Visual/auditory check against existing object documentation

E5: **Colour mode**

The colour mode corresponds to that at the time of production (either black/white or colour or discoloured)

Potential level of compliance: Fully preserved/ partially preserved
Test method: Visual check against existing object documentation

E6: **Play speed**

The original play speed intended at the time of production is known and is technically reproducible

Potential level of compliance: Reconstructible yes/no
Test method: Recourse to / comparison with documentation

E7: **Brightness / focus_resolution / grain**

Relevant properties for still images such as those listed here are known and can be measured / checked

Potential level of compliance: Full / partial
Test method: Visual check against existing object documentation

Possible matrix for moving images

	E1 Completeness	E2 Integrity	E3 Multiple tracks	E4 Synchronicity	E5 Colour mode	E6 Play speed
N1.1 Entertainment	X	X		X	X	X
N1.2 Enjoyment	X	X	X	X	X	X
N1.3 Memory aid				X	X	X
N1.4 Marketing	X	X	X	X	X	X
N2.1 Current affairs	X	X		X	X	X
N2.2 Documentation	X	X		X	X	X
N2.3 Education	X	X	X	X	X	X
N2.4 Evidence	X	X	X	X	X	X
N3.1 Old/new edit				X	X	X
N3.2 Comparative compilation of evidence	X	X	X	X	X	X
N3.3 Part of a work of art					X	X
N3.4 Manipulation					X	X

X = The property should be fulfilled in full or in part for this designated use.

Example of determining a preservation group

Example 1

The information objects of the preservation group are to be used by ethnologists (designated communities Z4.2). Probable designated uses are: N2.2 (documentation), N2.4 (use as evidence, scientific raw material) and N3.2 (comparison).

	E1 Completeness	E2 Integrity	E3 Multiple tracks	E4 Synchronicity	E5 Colour - mode	E6 Playspeed
N2.2 (Documentation)	X	X		X	X	X
N2.4 (Evidence)	X	X	X	X	X	X
N3.2 (Comparative compilation of evidence)	X	X	X	X	X	X
Significant properties	X	X	X	X	X	X
Degree of implementation	Full	Full	Full or partial	Full	Full	Full

Example 2

Parts of the information object are to be preserved to produce background reports for current affairs (designated community Z6 for Z2). (Creation of a new information object on the basis of parts of old information objects and combining them with new information objects)

	E1 Completeness	E2 Integrity	E3 Multiple tracks	E4 Synchronicity	E5 Colour - mode	E6 Playspeed
N3.1 (Old/new edit)				X	X	X
N3.4 (Manipulation)					X	X
Significant properties		X	X	X	X	X
Degree of implementation	Only partial required	Only partial required	Only partial required	Only partial required	Only partial required	Full (only for parts used)

V. Structured information

Definition

"Structured information" is the name given to a range of values or information types, the units, types and logical relationships of which are predefined in an explicit schema. Typical examples include databases, tables or XML data. In contrast to the other information types, structured information is normally not intended for direct sensory perception or for interaction, even if it contains other information types. Instead it should open up various enhanced processing and access possibilities for intermediary applications (e.g. requests for data records which simultaneously fulfil multiple criteria).

Possible designated communities (sample list)

Z1: **All designated communities wishing to use information or values which are not representation-dependent**

- Z1.1: Scientists (e.g. measured temperatures)
- Z1.2: Statisticians (e.g. economic data)

Z2: **All designated communities of other information types which require extended processing, accessing or linking possibilities**

- Z2.1: Photojournalists (e.g. search in picture libraries)
- Z2.2: University examination boards (plagiarism checks using text databases)
- Z2.3: Industry (e.g. patent databases)

Possible concrete designated uses

In contrast to the other information types, the direct perception of the entire object is generally of little use because the sensory perception of structured information requires selection, processing or evaluation. For example, a textual representation of structured information in a text editor only reveals its technical implementation. Any information types it contains (e.g. pictures in a picture library) first need to be selected before they can be displayed correctly.

N1: **To extract and analyse information**

Individual or multiple pieces of information are to be extracted from the information object.

- N2.1: Search for certain information units which can be defined and described in advance (e.g. a specific date)
- N2.2: Search for a new realisation which is obtained by analysing and combining a number of different information units
- N2.3: Search for evidence (e.g. way in which a scientist or an official body works)

N2: **To process information**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

N2.1: Calculating from the information of an information object (e.g. statistical analysis of historical economic data)

N2.2: Compilation of information from multiple information objects (e.g. collection of sources from different archive databases regarding a particular event)

Possible properties of structured information

E1: **Schema/structure**

Potential level of compliance	Full
	Logical equivalent (different structure without loss of information or function)
	Partial (implies loss of information, e.g. rounding error for different numerical data types or loss of data records)
Test method	Automatic (comparison of given unit/data type and actual value, e.g. whether entry is really a positive integer
	Manual check (sampling, generally excluded in practice)

E2: **Processing and access possibilities**

The processing and access possibilities of structured information are determined by the schema and also the application software. See information type Software for the properties of application software.

E3: **Number of data records**

Potential level of compliance	Percentage or number
Test method	Automatic
	Manual check (sampling, generally excluded in practice)

E4: **Values of the individual data records**

Potential level of compliance	Identical / not identical
	Quantifiable discrepancy (e.g. rounding error)
	Fulfilment based on information type
Test method	Automatic
	Manual check (sampling, generally excluded in practice)

Possible matrix for structured information

Structured information is basically a preliminary stage for actual use. This means that no evaluation of the properties can be undertaken on the basis of the usage possibilities described here.

	E1 Schema/structure	E2 Processing and access possibilities	E3 Number of data records	E4 Values of the individual data records
N2.1 Search for fragmented items of information	X	X	X	X
N2.2 Realisations obtained by combining information	X	X	X	X
N2.3 Search for evidence	X	X	X	X
N3.1 Calculation from information	X	X	X	X
N3.2 Compilation of information	X	X	X	X

X = The property should be fulfilled in full or in part for this designated use.

Two examples of determining a preservation group

Example 1

A database containing temperature measurement data is to be preserved for scientists (Z 1.2) for future climate models. The source is regarded as trustworthy and the survey methods are documented separately, which is why searches for evidence (N2.3) can be excluded as a designated use. Because the required methods and questions for scientific use cannot be predicted, it is not possible to narrow down the designated uses.

	E1 Schema/structure	E2 Processing and access possibilities	E3 Number of data records	E4 Values of the individual data records
N2.1 Search for fragmented items of information	X	X	X	X
N2.2 Realisations obtained by combining information	X	X	X	X
N3.1 Calculation from information	X	X	X	X
N3.2 Compilation of information	X	X	X	X
Significant properties	X	X	X	X
Degree of implementation	Logical equivalent	Specified by analysis software of the designated community	Full	Exact, none Rounding error

Example 2

A picture library with comprehensive context metadata is to be preserved for journalists (Z2.1).

	E1 Schema/structure	E2 Processing and access possibilities	E3 Number of data records	E4 Values of the individual data records
N2.1 Search for fragmented items of information	X	X	X	X
N2.2 New realisations from combining information	X	X	X	X
N2.3 Search for evidence	X	X	X	X
N3.2 Compilation of information	X	X	X	X
Significant properties	X	X	X	X
Degree of implementation	Logical equivalent	Determined by search portal	Full	Pictures of print quality **Identical metadata**

VI. GIS

Definition

Geoinformation is data with a spatial reference, i.e. the information is assigned to a geographical location (georeference). The location definition can be provided by a direct geographical reference in the form of coordinates or by means of an indirect reference to an administrative locality (country, city, road). The geodata information itself can be subdivided into geometrical data (position and form of the objects), topological data (stored spatial relationships), graphic features (signatures, labels) and attribute data (semantic descriptions).

The potential of geoinformation only becomes apparent in the context of a *geodata infrastructure*. This contains a geodatabase including geometrical and topological data and graphic features, thematic geodata for semantic descriptions, standardised web services for access, user interfaces and data exchange interfaces from and to external (specialist) applications. This provides the preconditions for capturing, modelling, analysing and visualising location-related data. Geoinformation systems are normally part of a more comprehensive specialised information system.

Standards such as the ISO-191xx series and the specifications of the *Open Geospatial Consortium (OGC)* are of pivotal importance for the compatibility and interoperability of the systems and data via standard interfaces and for the definition of the significant properties of individual components of geoinformation systems.

From an archive point of view, separate analysis of basic geodata (database view) and geoinformation systems (database, processing and visualisation view) lends itself as a means of accommodating the different evaluation approaches while taking the dynamics of the systems including the user view into account. *Basic geodata* are official geodata which describe the landscape (topography), the properties and buildings in a uniform geodetic spatial reference and in a manner which is application-neutral. They fulfil the function of basic data for geoinformation products and *geoinformation systems* with a high degree of utility in value chains. Geoinformation systems are information systems for capturing, processing, managing, analysing and visualising geodata.

Possible designated communities (sample list)

Z1: **Private individuals**
- Z1.1: Leisure activities such as geocaching, route planning

Z2: **Scientists: Education and Research**
- Z2.1: Historians
- Z2.2: Scientists such as biologists, geologists, physicists, meteorologists

Z3: **Providers and users of specialist geoinformation system from eGovernment**
- Z3.1: Land survey register
- Z3.2: Traffic information systems
- Z3.3: Environment information systems
- Z3.4: Transport telematics

Z4: **Specialist users and providers of additional services from private industry**
- Z4.1: Farmers

Z4.2: Extractive industry e.g. raw materials transport, water supply
Z4.3: Industrial services: Suppliers of guidance systems, transport monitoring and geolocation systems
Z4.4: Telecommunications

Possible concrete designated uses

N1: **To perceive the entire object**

The information object should be perceived by the senses as a work of art in the broadest sense, but one with no intention of conveying information.

N1.1: Visual perception of a geodata information object or parts thereof as a work of art

N2: **To extract and analyse information**

Individual or multiple pieces of information are to be extracted from the information object.

N2.1: Capture
Capture of attribute data to launch simulations; adding metadata or basic geodata to complete data collection

N2.1: Search
Search for a certain information unit which can be predefined and described (e.g. co-ordinates, subject headings, constructions, such as display of agricultural land, search for plots of land and agricultural units (felling areas), tour including information on Buga 2011 at the Ehrenbreitstein Fortress in Koblenz)

N2.2: Analysis
Search for a new realisation arising from the analysis and combination of multiple information units such as the visualisation of simulation results on the effects of pesticides; do I live in an area at risk of landslides?, route planning

N2.3: Retrieving attribute data, current information
Search for current information and retrieval of data such as traffic jam reports, road conditions in winter, earthquake in Nassau, flood warnings: river levels of the Mosel and Rhine

N3: **Map navigation**

N3.1: Navigation within a section of a map by using a zoom function to enlarge and reduce a section of it
N3.2: Representation of the maps in a window, the size of which can be enlarged or reduced dynamically by the user
N3.3: Panning function (moving a section)
N3.4: Displaying and hiding pre-configured layers
N3.5: Mixing of thematic layers (either two layers loaded in the map viewer, or one layer with a polygon digitised by the user)
N3.5: Use of north point and scales
N3.6: Retrieving data on one or more objects from one or more layers

N4: **Examples of the processing of basic geodata by geoinformation products**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

N4.1: Archiving-related example: Military leadership information systems (Michael Wunder, Jürgen Grosche: Verteilte Führungsinformationssysteme. – Springer, 2009, ISBN: 978-3-642-00508-4). A guidance information system is an IT instrument which can process information and present the strategic and tactical situation in a suitable information environment. The goal is the flexible generation and updating of a jointly used role-based deployment chart.
In many cases the geoinformation system is merged with parts of the guidance information system to create a monolithic block. There are methods for modularising this. This makes it difficult to archive situation maps (geoinformation system).

N4.2: FLOrlp (Flächeninformationen Online Rheinland-Pfalz) is a web GIS which supports farmers in applying for grants. FLOrlp is based on data from the land survey register and digital orthophotos (aerial pictures) of the surveying and cadastral authority of Rhineland Palatinate.

N4.3: The Wasser GeoPortal provides water engineering data in processed form for use by specialists in local communities, districts, associations, consulting engineers and universities, and also by interested citizens. Different instruments are used to grant simple and efficient access to the information.

N4.4: The State Office for Geology and Mining of Rhineland Palatinate provides online maps on selected geoscientific topics. The interactive maps for Rhineland Palatinate are available throughout Germany.

N4.5: The State Office for the Environment, Water Distribution and Trade Control allows users to obtain information on the implementation of the EU Environmental Noise Directive. This is a GIS-based map service which provides access to maps containing the noise levels in the vicinity of main roads in Rhineland Palatinate.

Possible properties of basic geodata

Basic geodata are official geodata which describe the landscape (topography), the plots of land and buildings in a uniform geodetic spatial reference and in a non application-based manner. Basic geodata are considerably simplified here and count as a spatially-related database in which the data records are geographical information with reference to a general data model (e.g. properties, raster displays, topologies etc.) This yields the following properties for basic geodata.

E1: **General properties of databases -> see above Structured information**

E2: **Description of the topographic area in a geotopographic database based on a model which is commonly used and widespread in Germany such as ATKIS AdV:**
- the Digital Agricultural Model (DLM)
This describes the topographic objects of the landscape and the relief of the earth's surface in vector format. The objects are assigned to a specific object type and are defined in terms of their geographical location, the geometrical type, descriptive

attributes and relationships to other objects (relationships). Each object has a unique identification number which is valid throughout Germany.
- Digital Topographic Maps (DTK)
These are raster data of the existing topographic maps. The raster data are structured by cartographic content elements into different layers and can be issued as individual monochrome layers and as a combined monochrome and colour edition.
- Digital Terrain Model (DTM)
DTMs are collections of geocoded points which are arranged in regular grids, or irregularly or linearly with regard to their location and altitude; they describe the terrain forms of the earth's surface (relief). Digital Terrain Models can also contain supplementary information (e.g. terrain edges, skeleton lines and individual peaks).

Potential level of compliance To be fully preserved
Test method Manual check

E3: **Support for a standard interface or for data exchange formats such as the EDBS for the exchange of ALK data or the OKSTRA interface (object catalogue for road and transport system) for transfer to a GIS database in the archive.**

Potential level of compliance To be fully preserved
Test method Automatic, can be automated, although manual checking should be considered during the prearchiving evaluation

Possible properties of geoinformation systems

Geoinformation systems contain the following components: basic geodata, thematic geodata, standardised web services for access, user interfaces and data exchange interfaces from and to external (specialist) applications. This permits capturing, modelling, analysis and visualisation of spatial data. It is therefore a complex system consisting of multiple components; the properties are assigned to these below. Underlying this is the question of which requirements the components must fulfil in order to have the potential to be preserved. The fact that the components can be replaced by alternative solutions in the preservation repository is taken into consideration.

E4: **Thematic geodata**
Thematic geodata are thematic data with a spatial reference. They consist of basic geodata and specialist data. In the modelling it is advisable to make a distinction between basic and thematic geodata. The modelling should be based on a proven model such as the AAA reference model (ALFIS-ALKIS-ATKIS model) in order to ensure ISO 191xx and OGC conformity and therefore interoperability.

Potential level of compliance To be fully preserved
Test method Manual checking during the prearchiving evaluation

E5: **Data transformation**
Data transformation functions which derive new geographical data from the existing data. These geo processing functions take information from the existing data records, apply analytical functions to them and enter the results into new data records. In order to preserve the software functions it is crucial to use a commonly used programming

language such as Java and to transfer the documented source code. Alternatively the functions can be replaced by open source solutions.

Potential level of compliance To be partially preserved
Test method Manual

E6: **Standardised web services for access, analysis, visualisation; front-end**
The large number of map views can be dynamically generated from underlying geographical information which can be regarded as a kind of "window into the database". These help support retrieval, analyses and processing. The requirements given in E5 apply for the software.

Potential level of compliance To be partially preserved
Test method Manual

E7: **Data exchange interfaces (DTA) from and to external (specialist) applications**

Potential level of compliance To be partially preserved
Test method Manual

Possible matrix for GIS data

	E1 Databases (general)	E2 Model of the geotopographic base data	E3 Standard interfaces	E4 Thematic geodata	E5 Middle-ware	E6 Web-services, visuali-sation	E7 Interfaces to external applications
N1.1 Visual perception as work of art						X	
N2.1 Capture	X	X					
N2.1 Search	X			X			
N2.2 Analysis	X	X	X	X	X	X	X
N2.3 Retrieval of attribute data				X	X	X	
N3.1 – 3.6 Map navigation	X			X	X	X	
N4.1 – N4.5 Specialist use of geoinformation systems	X	X	X	X	X	X	X

X = The property should be fulfilled in full or in part for this designated use.

Example of determining a preservation group

Example 1

The use is a scientist (Z2) who carries out analyses (N2,2) and has the results visualised (N3.1-3.6).

	E1 Databases (general)	E2 Model of the geotopographic base data	E3 Standard interfaces	E4 Thematic geodata	E5 Middle-ware	E6 Web-services, visualis-ation	E7 Interfaces to external applications
N1.1 Visual perception as work of art							
N2.1 Capture							
N2.1 Search							
N2.2 Analysis	X	X	X	X	X	X	X
N2.3 Retrieval of attribute data							
N3.1 – 3.6 Map navigation	X	X	X	X	X	X	
N4.1 – N4.5 Specialist use of geoinformation systems							
Significant properties	X	X	X	X	X	X	X
Degree of implementation	Full	Full	Preserved	Full	Preserved	Preserved	Preserved

Example 2

The user is a member of the military e.g.. of the Federal Ministry of Defence (Z3) who wishes to analyse situation maps (N2.2, N3.1-3.6, N4.1).

	E1 Databases (general)	E2 Model of the geotopo-graphic base data	E3 Standard interfaces	E4 Thematic geodata	E5 Middle-ware	E6 Web-services, visualis-ation	E7 Interfaces to external applications
N1.1 Visual perception as work of art							
N2.1 Capture							
N2.1 Search							
N2.2 Analysis	X	X	X	X	X	X	X
N2.3 Retrieval of attribute data							
N3.1 – 3.6 Map navigation	X	X	X	X	X	X	
N4.1 – N4.5 Specialist use of geoinformation systems	X	X	X	X	X	X	
Significant properties	X	X	X	X	X	X	X
Degree of implementation	Full	Full	Preserved	Full	Preserved	Preserved	Preserved

VII. Software

Definition

Software: The entirety or parts of programs, procedures, rules and all related documentation of an information processing system. (ISO 2382-1: 1993 from DIN-Taschenbuch 166, p. 311)

(*Computer) program*: A syntactical unit which obeys the rules of a given programming language and is composed of declarations and instructions; it is needed to solve a specific function, task or problem. (ISO 2382-1: 1993 from DIN-Taschenbuch 166, p. 311)

Possible designated communities (sample list)

Z1: **Computer gamers**
- Z1.1: Professional gamers
- Z1.2: Occasional players
- Z1.3: Games collectors

Z2: **Academics**
- Z2.1: Computer historians, computer scientists
- Z2.2: Social scientists, educationalists, psychologists
- Z2.3: Cultural scientists, media researchers, graphic designers, designers
- Z2.4: Games researchers, games theorists

Z3: **Those interested in history**
- Z3.1: Consumer nostalgics
- Z3.2: Technology nostalgics

Z4: **IT sector**
- Z4.1: Software producers
- Z4.2: Rights holders, jurists
- Z4.3: Youth protection workers

Z5: **Groups searching for information**
- Z5.1: Parents
- Z5.2: Teachers, educators
- Z5.3 Politicians

Z6: **Digital archivists (software as preservation planning tool)**

Possible concrete designated uses

N1: **To perceive, experience the entire object**

The information object should be experienced by the senses (visually, acoustically, haptically) with no intention of providing information.

- N1.1: Direct perception, experience of a program as a work

N2: **To obtain and evaluate information from the object**

Individual or multiple pieces of information are to be extracted from the information object.

N2.1: Discernible components of the software (e.g. graphic elements, interfaces, sound) are analysed and assessed.
N2.2: The source code or the sequence of the software procedures are analysed and assessed.
N2.3: The documentation of the software is analysed and assessed.

N3: **To process information**

A new information object is to be created from the information extracted from the information object through a process of selection, compilation and processing.

N3.1: Compilation of a source code for a new use platform
N3.2: Use of individual parts of the source code in other software
N3.3: Use of individual parts of the software (e.g. libraries) in other software
N3.4: Recording of a program sequence and representation as a series of screenshots or as a film
N3.5: Deducing the construction by observing a running program (reverse engineering)

N4: **To maintain the operability of the software**

The software is needed to carry out current tasks.

N4.1: The program is the target application
N4.2 The program supports the target application

Possible properties of executable software

The software behaves as its programmers intended.

E1: **Correct running of the processes in terms of sequence and completeness**

Potential level of compliance	Correct / partially correct / not executable
Test method	Conceivably automatic (compatibility check by the current operating system), otherwise manual check (samples)

E2: **Speed of running procedures**

Potential level of compliance	(in parts) faster / (in parts) identical / (in parts) slower
Test method	Automatic measurement and human assessment

E3: **Authenticity and completeness of the perceptible components (look and feel)**

Perceptible components are everything which the user can perceive with his or her senses when using the software. This can also refer to specially required hardware components, e.g. a special input device. In the case of software which is designed for a broad range of usage platforms (e.g. PC), it is difficult to define authenticity and completeness. A program can create very different impressions on a low- and on a high-performance PC. Furthermore, different user groups prefer different types of

presentation. For this reason, authenticity and completeness must always be viewed in conjunction with the designated use.

Potential level of compliance Full / reduced / minimum
Test method Human appraisal

Possible properties of source codes

The source code contains (as text) the formal description of the program in a (higher) programming language. The source code can be regarded as a representation of the program which cannot run itself. A translation program (compiler or interpreter) is generally required to translate the source code into a machine-executable code. Theoretically the source code could also be regarded as a preservation group of the information type text.

E4: **Automatic processability**

It should be possible to have the source code (sections, commands, characters) adequately translated (i.e. as executable program) by a compiler or interpreter into machine code.

Potential level of compliance Compilable / not compilable
Test method Automatic (using compiler or interpreter)

E5: **Assessability of the source through analysis by human expert**

The assessment depends not only on the quality of the source code and its visual representation, but also on the availability of the representation information (expert knowledge on the programming language used).

Potential level of compliance Assessable / not assessable
Test method Human appraisal

Possible properties of software documentation

The documentation of a piece of software corresponds to the information type text and therefore assumes all its properties.

Possible matrix for executable software

	E1 Correct running	E2 Speed	E3 Look and Feel	
N1.1 Experience of a program as a work	X	X	X	Z2; Z3; Z5; Z1.3
N2.1 Analysis of perceptible components	X & x	X & x	X	Z2; Z4
N3.3 Re-use of parts of the software	X			Z4.1; Z2.1
N3.4 Recording of the running program	X	X	x	Z1; Z2.4; Z4.3; Z5
N3.5 Reverse engineering	X	X	X	Z2.1; Z4.1
N4.1 Target application	X	X & x	X & x	Z1; Z4.1
N4.2 Target application supported	X	X		Z6; Z1; Z4.1

X = The property should be fulfilled in full for this designated use.
x = The property should be fulfilled in part for this designated use.

Possible matrix for source code

	E4 Automatic processability	E5 Assessability by experts	
N2.2 Analysis of source code		X	Z2.1; Z4.1; Z4.2
N3.1 Compilation of the source code	X		Z2.1; Z4.1; Z6
N3.2 Re-use of parts of the source code in other programs	X	X	Z2.1; Z4.1

X = The property should be fulfilled in full for this designated use.
x = The property should be fulfilled in part for this designated use.

Two examples

Example 1

A computer game is to be preserved for designated community Z1.1 (professional gamers). The designated use is therefore the use of the program (game) as a target application. Professional gamers strive to achieve regulated and comparable (sports) performance, meaning that properties E1 and E2 are of top priority. Players often deactivate perceptible components of the presentation which are not relevant for playing the game itself (e.g. high resolution of textures, real-time calculation of shadows etc.), meaning that property E3 is only of lower significance.

	E1 Correct running	E2 Speed	E3 Look and Feel	
N4.1 Target application	Correct	Same	Reduced	**Z1.1 Professional gamers**

Example 2

A computer game is to be preserved for designated community Z4.3 (youth protection officials). The designated use is therefore the analysis of perceptible components of the game to check for youth protection-relevant content. The entire program needs to be checked, meaning that E1 and E3 have top priority. Deviations in the speed (e.g. fast-forwarding) are acceptable, however, in order to accelerate the checking process.

	E1 Correct running	E2 Speed	E3 Look and Feel	
N2.1 Analysis of perceptible components	Correct	Faster	Full	**Z4.3 Youth protection workers**

Appendix B: Glossary

Preliminary remarks: The definitions are based where available on those contained in DIN 31644 and DIN 31645. They deal primarily with the use case of digital preservation.

Authenticity	In the context of long-term digital preservation and when related to representations, this means that the information to be archived has not been changed since the start of the transfer to the preservation repository, even if new representations have been created.
File format	Definition of the structure and meaning of the data in a file.
Data	Formalised digital representation of information which permits it to be interpreted, processed and exchanged.
Emulation	Strategy for preserving the long-term accessibility of representations. The strategy ensures that the system requirements for older representations can be recreated (emulated) through the use of special software on common systems then available on the market. The representations themselves remain unchanged where possible.
Level of compliance	The extent to which a significant property is achieved.
Preservation group	Group of information objects with the same significant properties which can be preserved using the same processes.
Information	All exchangeable types of knowledge; the term covers the content of works of intellectual creativity, the results of research and development, the documentation of political, social and economic events.
Information type	Group of archival documents with more or less the same properties.
Information object	A logically discrete unit of information.
Migration	The creation of a new representation in a new file format with the aid of a conversion program.
Designated use	Describes the functional possibility of how an archival document can be used.
Performance	The issue of information in a way which is perceptible to people on the basis of the interaction of data, hardware and software.
Representation	Logically discrete unit of digital data in an IT system which digitally represents an information object in full.
Significant properties	The properties of the information object being transferred which are to be preserved throughout the entire archiving process.
Designated community	Identifiable group of assumed users of an archive with specific and determinable interests and circumstances.

Appendix C: References

Addis, Matthew ; Wright, Richard: Audiovisual preservation strategies, data models and value Chains. – Southampton, 2010 https://prestoprimews.ina.fr/public/deliverables/PP_WP2_D2.1.1_preservationstrategies_R0_v1.00.pdf (accessed: 26.9.11)

AKEA [Arbeitskreis Elektronische Archivierung]: Was sollen wir erhalten? : Anmerkungen zu den Eigenschaften digitaler Objekte. – (no place of publication), 2008 http://www.wirtschaftsarchive.de/vdw/arbeitskreise/fachliche-arbeitskreise/elektronische-archivierung/EigenschaftendigitalerObjekte.pdf (accessed: 26.9.11)

Ball, Alexander ; Ding, Lian ; Patel, Manjula: Towards a curation and preservation architecture for CAD engineering models. – Bath, 2008 http://www.bl.uk/ipres2008/presentations_day1/17_Ball.pdf (accessed: 26.9.11)

Becker, Christoph: Trustworthy Preservation Planning (nestor edition 4). – (no place of publication), 2011 http://nbn-resolving.de/urn:nbn:de:0008-2011061603 (accessed: 26.9.11)

Becker, Christoph ; Heydegger, Volker ; Rauber, Andreas ; Schnasse, Jan ; Thaller, Manfred: Generic XML language for characterising objects to support digital preservation. – Vienna, (no date of publication) http://www.ifs.tuwien.ac.at/~becker/pubs/becker_sac08.pdf (accessed: 26.9.11)

Becker, Christoph ; Kulovits, Hannes ; Guttenbrunner, Mark ; Strodl, Stephan ; Rauber, Andreas ; Hofman, Hans: Systematic planning for digital preservation : evaluating potential strategies and building preservation plans. – In: International journal on digital libraries. – (no place of publication), 2009 http://publik.tuwien.ac.at/files/PubDat_180752.pdf (accessed: 26.9.11)

Beinert, Tobias ; Brantl, Markus ; Kugler, Anna ; Kulovits, Hannes ; Rauber, Andreas ; Schoger, Astrid: From TIFF to JPEG 2000? : preservation planning at the Bavarian state library

Using a collection of digitized 16th century printings. – In: D-Lib magazine. – 15.2009 http://www.dlib.org/dlib/november09/kulovits/11kulovits.html (accessed: 26.9.11)

Brown, Geoffrey ; Woods, Kam: Migration performance for legacy data acces. – In: International journal of digital curation. – 3.2008, 2, p. 74- 88 http://www.ijdc.net/index.php/ijdc/article/view/88/106 (accessed: 26.9.11)

CCSDS [Consultative Committee for Space Data Systems]: Open Archival Information System (OAIS). – (no place of publication), 2002 http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed: 26.09.2011), German version: nestor-Working Group OAIS-Translation/Terminology (Ed.): Referenzmodell für ein Offenes Archiv-Informations-System – German translation (nestor-Materialien 16). – Frankfurt am Main, 2012 http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2012051101 (accessed: 8.8.12)

CCSDS [Consultative Committee for Space Data Systems]: Producer-Archive Interface Methodology Abstract Standard, Magenta book. – (no place of publication), 2004 http://public.ccsds.org/publications/archive/651x0m1.pdf (accessed: 26.9.11)

Chanod, Jean-Pierre ; Chidlovskii, Boris ; Dejean, Herve ; Fambon, Olivier ; Fuselier, Jerome ; Jacquin, Thierry ; Meunier, Jean-Luc: From legacy documents to XML : a conversion framework. – (no place of publication), 2005 http://www.springerlink.com/content/5xnqptg4hrdqmy3g/?p=92854330c8c0411b8e5772deb4c9ef5eπ=8 (accessed: 26.9.11)

Dappert, Angela: Deal with conflict, capture the relationship: the case of digital object properties. – Boston Spa, 2009 http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/dappert-05.pdf (accessed: 26.9.11)

Dappert, Angela: Report on policy and strategy models for libraries, archives and data centres. – (no place of publication), 2009

http://www.planets-project.eu/docs/reports/Planets_PP2_D3_ReportOnPolicyAndStrategyModelsM36_Ext.pdf (accessed: 26.9.11)

Dappert, Angela ; Farquhar, Adam: Modelling organizational preservation goals to guide digital preservation. – In: The international journal of digital curation. – 2.2009, 4, p. 119-134 http://www.ijdc.net/index.php/ijdc/article/viewFile/123/126 (accessed: 26.9.11)

Dappert, Angela ; Farquhar, Adam: Significance is in the eye of the stakeholder. – Wetherby, 2009 http://www.planets-project.eu/docs/papers/Dappert_Significant_Characteristics_ECDL2009.pdf (accessed: 26.9.11)

Durand, David ; Liu, Alan ; Montfort, Nick: Born again bits : a framework for migrating electronic literature. – (no place of publication), 2005 http://eliterature.org/pad/bab.html (accessed: 26.9.11)

Ernst, Albrecht ; Keitel, Christian ; Koch, Elke ; Rehm, Clemens ; Treffeisen, Jürgen : Überlieferungsbildung bei personenbezogenen Unterlagen. – In: Archivar. – 61.2008, p. 275-278, also http://www.archive.nrw.de/archivar/

Giaretta, David ; Matthews, Brian ; Bicarregui, Juan ; Lambert, Simon ; Guercio, Mariella ; Michetti, Giovanni ; Sawyer, Donald: Significant properties, authenticity, provenance, representation information and OAIS (Proceedings of the 6th international conference on preservation of digital objects - iPres 2009). – San Francisco, 2009, p. 67-73 http://www.interpares.org/display_file.cfm?doc=ip3_italy_dissemination_cpr_guercio_ipres-2009.pdf (accessed: 26.9.11)

Gladney, Henry: Audio archiving for 100 years & longer: once we decide what to save, how should we do it?. – In: Journal of the audio engineering society. – 49.2001, 7/8 http://www.aes.org/journal/suppmat/gladney_2001_7.pdf (accessed: 26.9.11)

Granger, Stewart: Emulation as a digital preservation strategy. – In: *D-Lib magazine.* – 6.2000, 10
http://www.dlib.org/dlib/october00/granger/10granger.html (accessed: 26.9.11)

Hedstrom, Margaret ; Lee, Christopher: Significant properties of digital objects: definitions, applications, implications. – In: Proceedings of the DLMForum, Barcelona, 2002, p. 218-227
http://www.ils.unc.edu/callee/sigprops_dlm2002.pdf (accessed: 26.9.11)

Hedstrom, Margaret ; Lampe, Clifford: Emulation vs. migration: do users care? – In: RLG DigiNews. – 15.2001, 6
http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070513/viewer/file1628.html#feature1 (accessed: 26.9.11)

Helwig, Petra ; Roberts, Bill ; Nimmo, Emily: The Nationaal Archief of the Netherlands and the use of emulation. - (no date or place of publication)
http://www.planets-project.eu/docs/casestudies/PlanetsCasestudy_NationalArchiefandemulation.pdf (accessed: 26.9.11)

Heslop, Helen ; Davis, Simon ; Wilson, Andrew: An approach to the preservation of digital records. – Canberra, 2002
http://www.naa.gov.au/images/an-approach-green-paper_tcm2-888.pdf (accessed: 26.9.11)

Hitchcock, Steve ; Tarrant, David ; Carr, Les; Kulovits, Hannes ; Rauber, Andreas: Connecting preservation planning and PLATO with digital repository interfaces. – (no place of publication), 2010 http://eprints.ecs.soton.ac.uk/21289/1/ipres2010_submitted.pdf (accessed: 26.9.11)

Huth, Karsten: Probleme und Lösungsansätze zur Archivierung von Computerprogrammen - am Beispiel der Software des ATARI VCS 2600 und des C64. – Berlin, 2004 (Masters thesis)

Jenkins, Clare: Cedars Guide to: Digital preservation strategies. – (no place of publication), 2005
http://www.imaginar.org/dppd/DPPD/146%20pp%20Digital%20Preservation%20Strategies.pdf

Keitel, Christian: Benutzerinteressen annehmen und signifikante Eigenschaften festlegen: einige neue Aufgaben für Archivare. – In: Archive im digitalen Zeitalter. Überlieferung - Erschließung - Präsentation ; 79. Deutscher Archivtag 2009 in Regensburg (Tagungsdokumentation zum Deutschen Archivtag; 14). – Fulda, p. 29- 42 (VDA, 2010)

Keitel, Christian ; Lang, Rolf: Ingest von Fachverfahren im Landesarchiv Baden-Württemberg: Anmerkungen zu Authentizität, Prozessen und Softwareentwicklung. – In: Entwicklungen in den Bereichen Records Management / Vorarchiv - Übernahme – Langzeitarchivierung: Dreizehnte Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ vom 27./28. April 2009 ausgerichtet vom Staatsarchiv St. Gallen (Veröffentlichungen des Staatsarchivs St. Gallen). – St. Gallen, 2009, p. 35-45

Knight, Gareth: InSPECT Framework Report. – (no place of publication), 2009
http://www.significantproperties.org.uk/inspect-framework.pdf (accessed: 26.9.11)

Kol, Nada ; Oltmans, Erik: Comparison between migration and emulation in terms of costs. – In: RLG DigiNews. – 9.2005, 9
http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file959.html#article0 (accessed: 26.9.11)

Koninkliijke Bibliotheek & Rand Europe: Emulation testbed for digital preservation. – (no place of publication), 2008
http://www.kb.nl/dnp/e-depot/documenten/Rand_report_e-depot.pdf (accessed: 26.9.11)

KOST [Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen]: Katalog archivischer Dateiformate. - (no date or place of publication)
http://www.kost-ceco.ch/wiki/whelp/KaD/ (accessed: 26.9.11)

Kulovits, Hannes ; Becker, Christoph ; Kraxner, Michael; Motlik, Florian ; Stadler, Kevin ; Rauber, Andreas: Plato: a service oriented decision support system for preservation planning (Joint Conference on Digital Libraries from 16 to 20 June 2008). – Pittsburgh, 2008 http://publik.tuwien.ac.at/files/PubDat_172733.pdf (accessed: 26.9.11)

Long, Andrew Stawowczyk, Long-term preservation of web archives: experimenting with emulation and migration methodologies (IIPC project to evaluate emulation and migration as Long-term preservation solutions for web archives). – (no place of publication), 2009 http://www.netpreserve.org/publications/NLA_2009_IIPC_Report.pdf (accessed: 26.9.11)

McKinney, Peter: Preservation planning: a comparison between two implementations. – Wellington, 2010 http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/mckinney-74.pdf (accessed: 26.9.11)

nestor-Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff (Ed.): Wege ins Archiv : ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv (nestor-Materialien 10). – Frankfurt am Main, 2008 http://nbn-resolving.de/urn:nbn:de:0008-2008103009 (accessed: 26.9.11)

nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (Ed.): Kriterienkatalog vertrauenswürdige digitale Langzeitarchive (nestor-Materialien 8, Version 2). – Frankfurt am Main, 2008 http://edoc.hu-berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf (accessed: 26.9.11)

Neuroth, Heike ; Oßwald, Achim ; Scheffel, Renate ; Strathmann, Stefan ; Jehn, Mathias: nestor Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung. – Göttingen, 2009 http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2010071949 (accessed: 26.9.11)

Ockerbloom, John Mark: Archiving and preserving PDF files. – In: RLG DigiNews. – 15.2001, 5 http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070513/viewer/file558.html#feature2 (accessed: 26.9.11)

Oltmans, Erik ; Nanda Kol: A comparison between migration and emulation in terms of costs. – RLG DigiNews. – 9.2005, 2 http://www.rlg.org/en/page.php?Page_ID=20571 (accessed: 26.9.11)

Premis [Preservation Metadata: Implementation Strategies]: Version 2.1. – (no place of publication), 2011 http://www.loc.gov/standards/premis/version-2-1-announcement.html (accessed: 26.9.11)

Rathje, Ulf: Technisches Konzept für die Datenarchivierung im Bundesarchiv. – In: Der Archivar. – 55.2002, 2, p. 117-120 http://www.bundesarchiv.de/imperia/md/content/abteilungen/abtb/1.pdf (accessed: 26.9.11)

Reuben, Elizabeth: Migrating records from proprietary software to RTF, HTML and XM. – In: Computers in Libraries, 23.2003, 6, p. 30-33 http://www.infotoday.com/cilmag/jun03/reuben.shtml (accessed: 26.9.11)

Richter, Wolfgang: Standards für Archivformate : Archivische Anforderungen an Dateiformate vor dem Hintergrund der Migrationsstrategie. – In: Archiv und Wirtschaft 37.2004, 3, p. 124-130

Rothenberg, Jeff: Avoiding technological quicksand: finding a viable technical foundation for digital preservation. – Washington DC, 1999 http://www.clir.org/pubs/reports/rothenberg/contents.html (accessed: 26.9.11)

Rothenberg, Jeff: Ensuring the longevity of digital information. – (no place of publication), 1995 http://www.clir.org/pubs/archives/ensuring.pdf (accessed: 26.9.11)

Rothenberg, Jeff: An experiment in using emulation to preserve digital publication. – Den Haag,
2000http://www.studioautomata.com/itp/indestudy/emulationpreservationreport.pdf (accessed: 26.9.11)

Russell, Kelly: Digital preservation and the Cedars project experience. – (no place of publication), 2009 http://www.tandfonline.com/doi/pdf/10.1080/13614530009516805 (accessed: 26.9.11)

Siermann, Barbara ; Wheatley, Paul: Evaluation of Preservation Planning within OAIS, based on the Planets Functional Model; (no place of publication), 2010 http://www.planets-project.eu/docs/reports/Planets_PP7-D6_EvaluationOfPPWithinOAIS.pdf (accessed: 14.10.11)

Strodl, Stephan ; Becker, Christoph ; Neumayer, Robert ; Rauber, Andreas: How to choose a digital preservation strategy: evaluating a preservation planning procedure. – Vienna, 2007 http://www.ifs.tuwien.ac.at/~strodl/paper/FP060-strodl.pdf (accessed: 26.9.11)

Todd, Malcolm: Technology watch report: file formats for preservation (DPC Technology watch series report 09-02). – (no place of publication), 2009 http://www.dpconline.org/component/docman/doc_download/375-file-formats-for-preservation (accessed: 26.9.11)

Van Diessen, Raymond J. ; Rijnsoever, Ben J. van: Managing media migration in a deposit system (IBM/KB Long-term preservation study report series No. 5. Amsterdam: IBM Netherlands). – (no place of publication), 2002 http://www-5.ibm.com/nl/dias/resource/migration.pdf (accessed: 26.9.11)

Wijk, Caroline van: KB and migration : working document. – (no place of publication), 2006 http://www.kb.nl/hrd/dd/dd_projecten/KB%20and%20Migration.pdf (accessed: 26.9.11)

Wijk, Caroline van: KB and migration: test plan. – (no place of publication), 2006 http://www.kb.nl/hrd/dd/dd_projecten/KB%20and%20Migration%20Test%20Plan.pdf (accessed: 26.9.11)

Wilson, Andrew: InSPECT : Significant properties report. – (no place of publication), 2007 http://www.significantproperties.org.uk/wp22_significant_properties.pdf (accessed: 26.9.11)