

Project no. 269977

APARSEN Alliance for Permanent Access to the Records of Science Network

Instrument: Network of Excellence

Thematic Priority: ICT 6-4.1 – Digital Libraries and Digital Preservation

D31.1 REPORT ON DRM PRESERVATION

Document identifier:	APARSEN-REP-D31_1-01-1_4
Due Date:	2013-12-31
Submission Date:	2014-01-17
Work package:	WP31
Partners:	BL, DNB, KB, ONB
WP Lead Partner:	DNB
Document status	Released
URN	urn:nbn:de:101-20140516217

Abstract: This report places the subject of Digital Rights and Access Management (DRM) within the context of long-term digital preservation and examines the related risks and challenges which arise in connection with the long-term archiving and ongoing accessibility of DRM-protected objects, and also the safeguarding of associated rights. It reviews the results of initiatives and projects already undertaken in this field and provides the results of a recent survey. It also analyses current user scenarios both within and outside the APARSEN consortium before offering a summary of recommendations and best practices for dealing with digital rights and DRM protected objects.

Delivery Type	Report
Author(s)	Kirnn Kaur (BL), Stefan Hein (DNB), Sabine Schrimpf (DNB), Marcel Ras (KB), Manuela Holzmayr (ONB)
Approval (internal)	René van Horik (DANS), Eefke Smit (STM) (internal reviewers)
Approval (external)	Stefan Funk (Göttingen State and University Library)
Summary	
Keyword List	Digital rights, digital rights management (DRM)
Availability	<input checked="" type="checkbox"/> Public

Document Status Sheet

Issue	Date	Comment	Author
0.1	2013-10-11	Initial structure and methodology	DNB
0.2	2013-11-25	first DRAFT-Version	KB, BL, ONB, DNB
0.3	2013-11-28	some minor changes and additions (document structure, ch. 6 recommendations)	DNB
0.4	2013-12-02	Internal Review (remarks of ONB, BL) add Conclusion, Annex	ONB, BL, DNB
1.0	2013-12-05	DRAFT for internal and external review	DNB
1.1	2013-12-18	review remarks, updated ch. 6.5 research questions	BL, DNB
1.2	2013-12-20	textual amendments	BL, DNB, ONB
1.3	2014-01-16	Integration of the second review	DNB, BL
1.4	2014-01-16	Update chapter 2, finalizing	KB, DNB

Project information

Project acronym:	APARSEN
Project full title:	Alliance for Permanent Access to the Records of Science Network
Proposal/Contract no.:	269977

Project coordinator: Simon Lambert/David Giaretta

Address:	STFC, Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK
Phone:	+44 1235 446235
Fax:	+44 1235 446362
Mobile:	+44 (0) 7770326304
E-mail:	simon.lambert@stfc.ac.uk / david.giaretta@stfc.ac.uk

EXECUTIVE SUMMARY

As the result of the APARSEN work package 31 (WP31) about Digital Rights and Access Management the present report explores the subject of DRM by focusing on the various aspects, and above all the requirements, of long-term digital preservation. The terminology of digital rights and DRM are explained and defined in the beginning.

The report first examines how digital rights are represented by digital tools, and thus how they can ultimately be safeguarded on a permanent basis.

A further objective of this deliverable is to provide an overview of common DRM mechanisms and tools. The fact that DRM can be included as a part of the digital object itself and can also involve use and access restrictions necessitates an evaluation of the possible risks involved in the long-term preservation of DRM-protected objects. Access-restricted objects must be viewed as being potentially at risk, as the implementation of future preservation measures can be impeded or even prevented entirely by such restrictions. Nevertheless, DRM does not necessarily have to stand in the way of long-term archiving.

An overview showing how DRM and digital rights are handled within the long-term preservation community is also given. A study recently carried out in the Netherlands and also a study initiated by the work package are presented and assessed. The study results are supplemented by an appraisal of other DRM initiatives and by descriptions of actual user scenarios from the institutions represented in the work package.

An analysis of the results of a recent survey, conducted by the consortium is provided which reviewed how the participants handle DRM-protected materials and the associated rights. One of the primary targets was to discover how the community deals with archiving such objects and what is being done to protect the associated digital rights - both at present and, of course, in the future.

The report concludes with a summary of the different solutions which have been identified based on the results, and compiles a catalogue of best practices and recommendations supplemented with its own reflections. This deliverable therefore constitutes an important contribution to the establishment of the APARSEN VCoE (Virtual Centre of Excellence).

CONTENT

EXECUTIVE SUMMARY	4
LIST OF FIGURES	7
1 INTRODUCTION	8
1.1 DESCRIPTION OF THE DELIVERABLE (FROM DOW)	8
1.2 SCOPE.....	9
1.3 METHODOLOGY	9
1.4 DOCUMENT OUTLINE	10
1.5 LINKS TO THE OTHER APARSEN WPs	11
1.6 DEFINITION OF DIGITAL RIGHTS.....	11
1.7 DEFINITION OF DRM	12
2 REPRESENTATION AND PRESERVATION OF DIGITAL RIGHTS.....	16
2.1 WHAT ARE PRESERVATION METADATA?.....	16
2.2 PRESERVATION METADATA AND THE OAIS REFERENCE MODEL	16
2.3 PRESERVATION METADATA FRAMEWORK.....	17
2.4 PRESERVATION METADATA IMPLEMENTATION	17
2.5 PRESERVATION METADATA AND PREMIS	18
2.6 RIGHTS EXPRESSION LANGUAGES	18
2.6.1 OPEN DIGITAL RIGHTS LANGUAGE	19
2.6.2 METSRIGHTS	19
2.6.3 XRML	19
2.6.4 COPYRIGHTMD	19
2.6.5 MPEG21-REL	20
2.7 AN EXAMPLE	20
2.8 CONCLUDING.....	21
3 CLASSIFICATION AND EVALUATION OF DRM AND DRM SYSTEMS.....	23
3.1 CLASSIFICATION OF DRM AND DRM SYSTEMS.....	23
3.1.1 DATA CARRIER COPY PROTECTION	23
3.1.2 LIGHTWEIGHT DRM.....	23
3.1.3 ENCRYPTION-BASED PASSWORD PROTECTION	24
3.1.4 DRM SYSTEMS	25
3.1.5 SUMMARY OF DRM VARIANTS.....	30
3.2 EVALUATION	30
4 REVIEW OF DRM USE AND DEALING WITH DIGITAL RIGHTS.....	35
4.1 INITIATIVES DEALING WITH DIGITAL RIGHTS AND DRM	35
4.2 USER SCENARIOS IN THE WORLD OF LIBRARIES.....	37
4.2.1 AUSTRIAN NATIONAL LIBRARY (ONB).....	37
4.2.2 BRITISH LIBRARY (BL)	39
4.2.3 GERMAN NATIONAL LIBRARY (DNB).....	46
4.2.4 NATIONAL LIBRARY OF THE NETHERLANDS (KB)	50
4.3 THE DUTCH SURVEY	54
4.4 APARSEN DRM SURVEY	56
5 RESULTS OF THE APARSEN DRM SURVEY	57
5.1 SURVEY RESPONDENTS.....	57
5.2 THE LEGISLATIVE FRAMEWORK.....	57
5.3 THE DATA STOCK.....	57
5.4 THE APPROACH OF DEALING WITH DRM PROTECTED MATERIAL.....	58
5.5 THE APPROACH OF DEALING WITH DIGITAL RIGHTS	61
5.6 DIGITAL RIGHTS AND ACCESS	62
5.7 THE APPROACH OF DEALING WITH DRM PROTECTED MATERIAL AS A DATA HOLDER	65
5.8 THE APPROACH OF DEALING WITH DRM PROTECTED MATERIAL AS A DATA CREATOR	68
5.9 SUMMARY	69

6	RECOMMENDATIONS	70
6.1	GENERAL RECOMMENDATIONS	70
6.2	RECOMMENDATIONS FOR THE HANDLING OF DRM PROTECTED OBJECTS	71
6.3	RECOMMENDATIONS FOR THE PROTECTION OF DIGITAL RIGHTS	73
6.4	SUMMARY	75
6.5	POTENTIAL RESEARCH QUESTIONS	76
7	CONCLUSION	77
	REFERENCES	78
	ACRONYMS	80
	ANNEX 1: QUESTIONNAIRE OF THE APARSEN WP 31 SURVEY	81
	ANNEX 2: INVITATION FLYER	84
	ANNEX 3: MAILING-LISTS FOR SURVEY DISSEMINATION	85

LIST OF FIGURES

Figure 1: Document outline.....	10
Figure 2: Adapted SHAMAN LifeCycle.....	14
Figure 3: Architecture of a DRM System (adapted from (Rosenblatt, 2002))	26
Figure 4: Architecture of Microsoft's Windows Media Rights Manager (from (Microsoft Corporation, 2004))	29
Figure 5: Public Domain Calculation - example of a question.....	35
Figure 6: Example for a Pop-Up with information on the Q/A.....	36
Figure 7: Volume of content at the BL.....	41
Figure 8: Storage of digital collections at the KB	52
Figure 9: Since when do you deal with material that is protected by DRM? (n=18)	58
Figure 10: Differences between online and offline material (n=9)	60
Figure 11: Plans for analyzing the existing data stock (n=10)	61
Figure 12: Rights information on the object level (n=13)	61
Figure 13: Use of Metadata standards (n=13)	62
Figure 14: Kinds of software tools used (n=11)	63
Figure 15: Access categories (n=13)	64
Figure 16: Access allowed (n=13).....	65
Figure 17: Configurable access (n=11)	65
Figure 18: Data holder or data creator (n=18).....	66
Figure 19: Detection of DRM and rights information (n=10)	67
Figure 20: Do you accept DRM protected material? (n=10).....	68

1 INTRODUCTION

DRM stands for digital rights management and is the subject of intensive discussion in many quarters such as content providers, consumer and privacy groups. A closer examination of the term reveals two main aspects: firstly the rights to digital goods, and secondly to the management thereof, especially the checking, protection, description and above all the enforcement of these rights. These very concepts quickly reveal that this is a highly contentious field in which a number of stakeholders are all pursuing their own different interests. The subject of DRM and digital rights starts with authors and continues via producers and publishers right through to consumers, meaning that they remain relevant throughout the entire life cycle of a digital object.

The conflicts regularly portrayed in the media and literature are often between the interest groups of producers and consumers, i.e. the rights holders (copyright) and those wishing to make unrestricted use of a work. Within this discussion, the subject of long-term preservation and maintenance of accessibility, which clearly affect the last phases of the life cycle, are rarely accorded the attention they deserve. This deliverable represents an attempt to fill this gap.

1.1 DESCRIPTION OF THE DELIVERABLE (FROM DOW)

The following description of the deliverable is taken from the WP31 description in the Description of Work (DoW) document.

D31.1: Report on DRM preservation:

This report will summarize the research and requirements on DRM issues, including identification of appropriate tools. It will also contain an overview of appropriate best practices for dealing with DRM materials and its preservation, and recommendations for how best to undertake preservation within the constraints of DRM protection.

This deliverable is related to the following two tasks:

Task 3110: Review of DRM use and preservation:

Review the work that has been done in consortium members and beyond into DRM requirements and preservation. For example, the CASPAR project developed and tested tools which allowed one to deal with a changing legislative base on which to determine the rights associated with the aid digital object, and also to accommodate a multiplicity of legal systems. These capabilities are needed both in the current European environment and also to deal with digital rights created right now but which need to be honoured in the future in a different legal environment and with different methods of encoding the digital rights.

Furthermore this task will work out why long term preservation of DRM protected material is a current and important challenge, and how it has been overcome by partners. This will lead to recommendations for how these techniques may be adapted in future for other data where DRM considerations become important e.g. as data becomes more commercially valuable.

Task 3120: Digital Rights research:

In this task we compare the implementations of different digital rights tools and evaluate them by building up a classification that shows, which tools are well suited for digital preservation and which are not. A number of important questions will be identified and a research map will be produced.

1.2 SCOPE

The scope of this deliverable covers an examination of the technologies and approaches for handling DRM-protected materials in terms of long term archiving and ongoing accessibility. Other DRM-related topics which are also the subject of frequent discussion, such as data protection (privacy) and the public's right to information, are largely excluded or at best touched upon tangentially.

Because this work package consists mainly of participants from memory institutions (libraries), the majority of the findings obtained, especially with regard to the actual use cases presented later, are based on organisations concerned with the archiving and preservation of digital publications. Even though the experience and challenges with DRM protected publications gained by the libraries represented in the work package constitutes the main focus, other designated communities such as research data centres and industrial companies were not excluded in the studies. Anyway the presented concepts and findings are applicable to other sectors and communities, too.

Despite this, the report emphasises the perspective of the user of DRM and not necessarily the creator's or rights holder's views. The user, in this instance, represents the institution that is faced with the long-term preservation of digital material (that could be DRM-protected) including their associated rights.

1.3 METHODOLOGY

This deliverable is based mainly on the results of the two tasks (3110 and 3120) presented above. The following overview summarises the approaches which were selected and the steps which were necessary for the general organisation of the work, especially for the processing of these tasks. The order of the activities is not linear, i.e. some activities were conducted simultaneously and not necessarily in the order of the tasks set out in the DoW.

No.	Activities	related to
1.	Description of the internal WP use cases regarding the handling of DRM and digital rights	Task 3110
2.	Analysis of the use cases and of a DRM study already conducted in the Netherlands, yielding: input for step 4	Task 3110
3.	Finding a common definition of DRM	D31.1
4.	Conducting an online survey inside and outside APARSEN into the handling of DRM and digital rights	Task 3110 Task 3120
5.	Research into, and evaluation of, DRM systems and digital rights	Task 3110 Task 3120 D31.1
6.	Collecting individual contributions to create a DRAFT version, review and generation of final version	D31.1

1.4 DOCUMENT OUTLINE

The following chart (figure 1) shows the structure of the deliverable.

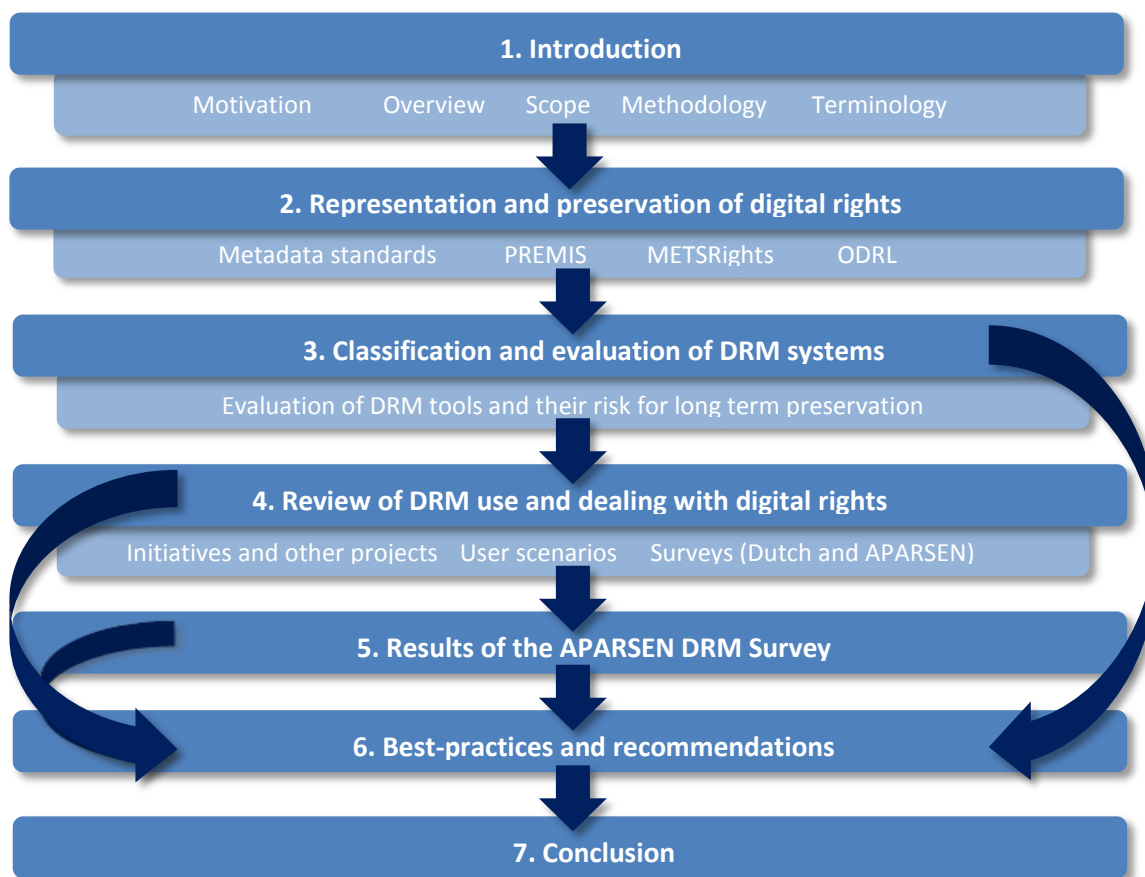


Figure 1: Document outline

Section 1 provides an overview of the structure of the deliverable and defines its scope and its designated community. There are also two separate subsections, which are dedicated to defining the main concepts of DRM and "digital rights". The latter, in particular, is presented within the context of current EU legislation.

Section 2 investigates the subject of representation and the long-term preservation of digital rights information. The emphasis here is on metadata standards, especially *Rights Expression Languages* (REL) for the marking of rights information.

Section 3 outlines common DRM systems and tools for the protection and management of digital rights. The DRM systems are classified and then evaluated in the light of the different long-term preservation aspects. This section therefore provides the main input for the recommendations given in Section 6.

Section 4 and **Section 5** examine how the long-term preservation community is currently coping with the challenges arising from DRM and digital rights. Both therefore provide crucial input for the target catalogue of best practices given in Section 6.

Section 6 contains a list of the previously identified current best practices in handling DRM and digital rights and presents the recommendations from the work package in the form of a catalogue. In addition it identifies some open research questions for the promised research map.

Section 7 draws conclusions and provides a tentative outlook for the future.

1.5 LINKS TO THE OTHER APARSEN WPS

This section describes the relationships between this and other work packages within the APARSEN project.

WP11 – Common Vision

WP31 examines the task of WP11 in order to arrive at a common understanding of the problems with regard to DRM and long-term digital preservation. Section 6 in particular, with its list of best practices and recommendations, makes an important contribution to the common vision.

WP13 – Coordination of Common Standards

Identification of common standards and gaps in the standards in the area of long-term digital preservation is one of the core tasks of WP13. The results of section 2 of this work package, especially the metadata standards for marking digital rights information, therefore provide input for WP13.

WP14 – Common Testing Environments

The preservation scenarios collected by WP14 include the "Detection of protection measures in online publications" scenario, a concrete test case for the testing of automatic recognition processes for DRM-protected materials. The main focus here is on recognising the type of DRM protection used. Given that certain types of DRM measures are known to be more or less compatible with long-term preservation, it can be determined whether a DRM-protected object should be imported into an archive system, or rejected or "repaired".

WP21 – Preservation Services

Similar to WP14, the preservation service "Characterization" covered in WP21 contains the possibility to determine the existence and the type of DRM measures, based on the technical metadata generated. Additionally, the "DRM Clearinghouse" service is closely related to section 2 of this deliverable. The principal task of this service is to: "Allow the digital rights associated with an object to be presented in a consistent way, taking into account the changes in legislation." (APARSEN 2013)

WP22 - Identifiers and citability

WP22 concerns issues related to persistent identifiers (PI). With Persistent Identifier Systems it is possible to determine the ownership/authorship of a file or object, also if it is embedded somewhere. Thus PIs can form a component of solution to do DRM.

WP35 - Data policies and governance

In the context of data policies WP35 has examined the role of digital licences. One of the findings is that the policy should give recommendations on the licenses to use, as well as possible reuse restrictions (e.g., CC licenses). The conducted survey of WP35 contains furthermore an overview of which licence types are part of the analysed policies. Thus this work package is related to section 2 which considers digital licences as a variant to express digital rights in a written form (APARSEN 2013).

1.6 DEFINITION OF DIGITAL RIGHTS

Digital rights refers to the 'rights' associated with accessing, using, creating and publishing digital content. The rights can relate to usage permissions as well as access preferences or limitations imposed upon digital content. In this respect, the digital content can be regarded as 'protected material', where the protection is on behalf of the 'creator' or 'owner' of the digital content. These rights can relate to

copyright legislation, intellectual property rights or contractual agreements imposed on the content. As stated on Wikipedia, digital rights describe the human rights that allow individuals access, use, creation and publishing of digital content as well as access and re-use of such content on electronic devices.

A recent communication from the European Commission (EC)¹ urges the delivery of solutions in order to provide greater access to online content. In relation to digital content one of the EC's objectives relates to the adherence to copyright and licensing relevant in the digital age. Within the World Intellectual Property Organisation (WIPO)², the EC negotiates on intellectual property rights (as well as industrial) in order to ensure adequate protection of intellectual property rights (IPR) at an international level. Copyright actions as identified in the EC Digital Agenda and EC Intellectual Property Strategy have delivered good progress; however, work is still to be done to ensure there is a single market for copyright. Two actions are to be undertaken by the Commission, which are given below:

1. Review and modernisation of EU copyright legislative framework (decision due in 2014); following elements will be addressed:
 - a. territoriality in the Internal Market;
 - b. harmonisation, limitations and exceptions to copyright in the digital age;
 - c. fragmentation of the EU copyright market; and,
 - d. how to improve the effectiveness and efficiency of enforcement while underpinning its legitimacy in the wider context of copyright reform.
2. Stakeholder (rights holders, licensing bodies, commercial and non-commercial users of protected content, as well as internet end-users) dialogue to deliver industry led solutions (by end of 2013), 'Licensing Europe', to explore licensing and technological solutions for EU copyright law. Topics to be covered:
 - a. Cross-border access and the portability of services
 - b. User-generated content and licensing for small-scale users of protected material
 - c. Audio-visual sector and cultural heritage institutions
 - d. Text and data mining

1.7 DEFINITION OF DRM

The following section defines how the term "Digital Rights Management" (DRM) is used in this paper. A selection of current types of DRM system is given in the following sections, especially in section 3. This is based on the definition given by Iannella (Iannella, 2002):

"Digital Rights Management (DRM) involves the description, layering, analysis, valuation, trading and monitoring of the rights over an enterprise's tangible and intangible assets. DRM covers the digital management of rights - be they rights in a physical manifestation of a work (eg a book), or be they rights in a digital manifestation of a work (eg an ebook)."

The definition is taken from the official specification of the *Open Digital Rights Language* (ODRL), a language created to describe rights information. The expansion of the term to include physical manifestations of a work is not covered in this paper which means that this report is focussed on managing the rights of digital content. Furthermore, this report looks at the rights related to certain kinds of digital use, for example; digital viewing/playing, copying, downloading, dissemination and resell of digital copies. Finally, as the rights protection and management is viewed as taking place

¹ Communication from the Commission on content in the Digital Single Market issued on 18.12.2012, accessed 25.11.13 http://ec.europa.eu/internal_market/copyright/index_en.htm

² World Intellectual Property Organisation (WIPO), website <http://www.wipo.int/portal/en/index.html>

digitally, there is the need of the digital representation of digital rights to also be reviewed in this report.

DRM is therefore mainly concerned with rights to digital objects which are protected and honoured by all stakeholders (e.g. producers and users). Here, the safeguarding of rights, are covered from the user's perspective. Set against this are the safeguarding and protection of the copyright and intellectual property rights of producers and authors. Iannella mentions the aspect of security and protection from improper duplication and use of works in an earlier definition (Iannella, 2001). These aspects are also often found in the DRM definitions given by content producers and companies such as Microsoft and Adobe (England, et al., 2002). One cornerstone of DRM linked to this aspect is **access management** which only grants access by authorised persons to the digital work and protects it from illegal use by means of encryption technologies (Picot & Thielmann, 2005, p. 16). In the case of Open Access material, access management could also mean granting access to every person without any limitation of use. Nevertheless organisations like research centers, which have to deal with this kind of object, have to take care about the preservation of the actual content and of course also of the type of digital rights applicable. Therefore it is necessary to preserve the appropriate rights information as mentioned in the following chapter.

This description and representation of digital rights in the form of licences represents a further cornerstone of DRM systems. It forms the basis of **usage control** - i.e. the manner in which the actual content can be used by users (Picot & Thielmann, 2005, p. 16). The marking of digital rights should be interpretable by both machines and people. The former is essential for the software-based implementation of DRM systems. Readability by humans also creates transparency and raises awareness of the fact that the use of a digital work may be subject to certain legal restrictions. In the digital world in particular, where it is easily possible to make perfect copies of digital objects, lack of transparency is one of the reasons for the "success" of music and software piracy.

Iannella's definition also includes the aspect of monitoring which is implied in (Picot & Thielmann, 2005, p.17) under "**Tracking of legal infringements**". According to this definition, DRM should also provide functions which facilitate monitoring the use of materials protected in this way. This also includes the tracking of legal infringements e.g. in the form of illegal duplication and provision on file-sharing sites.

Iannella's use of the term "trading" covers passing on rights to third parties but also touches upon a further aspect, that of **usage invoicing**. This permits revenue due to the publisher or producer to be collected based on use (Picot & Thielmann, 2005, p.18).

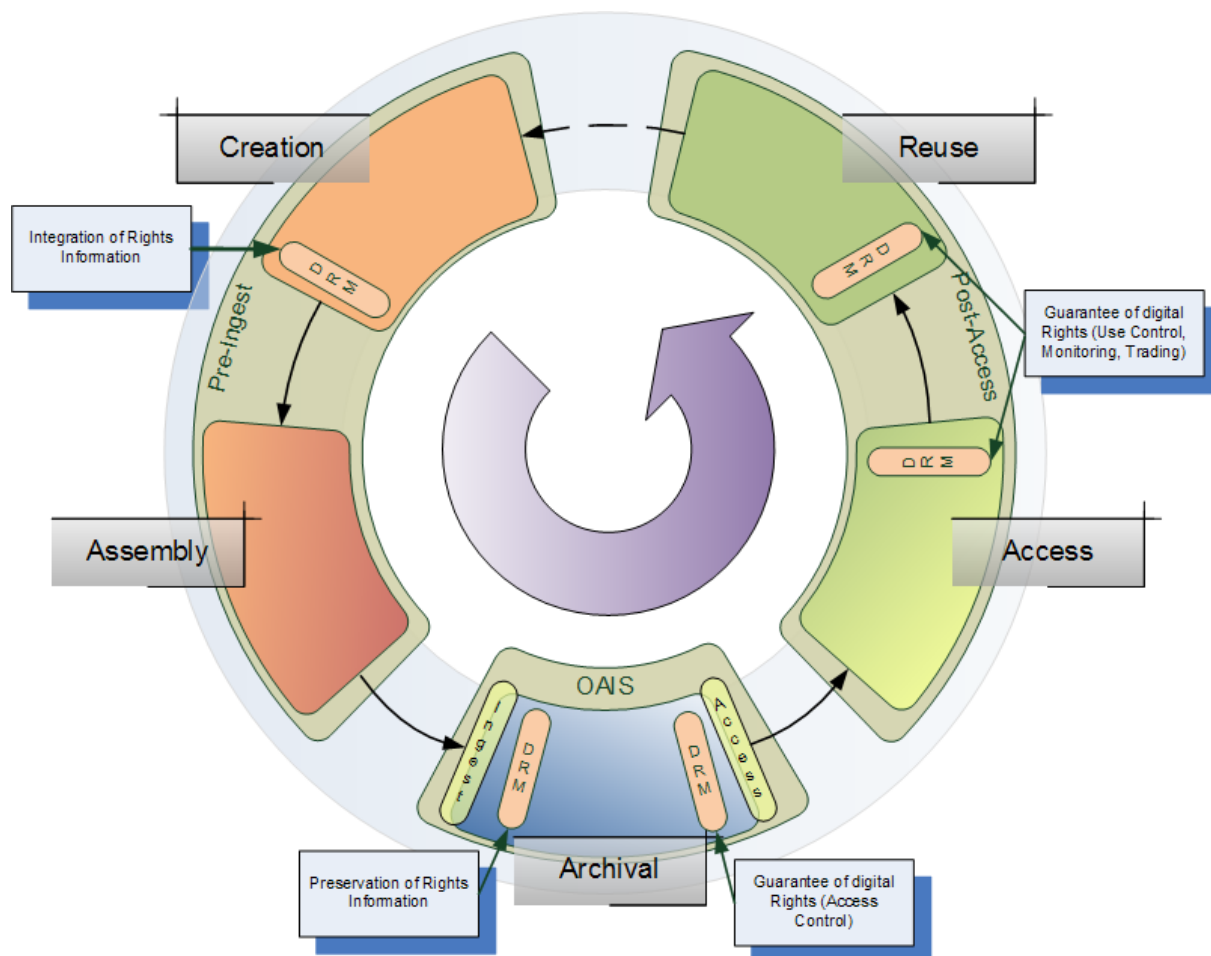


Figure 2: Adapted SHAMAN LifeCycle

Based on "The Life Cycle Phases of Digital Objects" developed in the SHAMAN³ project and the incorporation of DRM, it can be seen from the chart in figure 2 that long-term archiving also plays a key role in use besides the areas of content creation and the associated incorporation of digital rights (e.g. with the aid of rights expression languages) and the cornerstones mentioned above. The OAIS-compatible archiving system examined here is also a repository for use - i.e. not a dark archive. In this case, attention must be focused on digital rights during the ingest process e.g. through transfer to data management, such as administrative metadata, but this information must also be deployed during the access phases in order to safeguard these rights.

Finally, it should be mentioned that the study undertaken in this work package does not distinguish per se between DRM in the broader sense and DRM in the narrower sense, as the borders between them are relatively fluid in nature. The narrower sense here refers to concrete implementation within a file object e.g. password protection and encryption techniques. The broader sense includes all other DRM system mechanisms; these also include links outside the file object, e.g. to authentication servers and use-based invoicing technologies. The latter allows for example lending systems to manage the accounting depending on the duration of use. Furthermore any in-house implementation of a rights management system (i.e. see **Fehler! Verweisquelle konnte nicht gefunden werden.**), is considered as DRM in a broader sense. In most cases such solutions do not depend on the file format. Thus the preservation of the content and the preservation of the associated rights are two separate aspects. The

³ <http://shaman-ip.eu/>

first aspect is considered in the following chapter, the second falls outside the scope of the research undertaken within this WP.

2 REPRESENTATION AND PRESERVATION OF DIGITAL RIGHTS

This chapter gives answers to the question how digital rights of digital content can be preserved over the long-term. The first aspect of this rather challenging task is to find a way for representing digital rights in a digital form. Therefore, this chapter begins with an introduction into the concept of preservation metadata which is widely discussed within the long-term preservation community and which is an important concept of the OAIS Reference Model.

Subsequently, the chapter presents a couple of well-known metadata standards for digital rights and permissions and gives two examples for their use.

The chapter concludes with considerations about preserving such metadata, which are, according to OAIS, managed within a data management, a catalogue or another database component. That implies that metadata and the related content are managed separately from each other and could also have different storage locations.

2.1 WHAT ARE PRESERVATION METADATA?

All memory institutes have to deal with metadata and resource descriptions. Metadata is categorized according to what it is intended to accomplish: descriptive metadata helps in discovery and identification of resources, administrative metadata helps in managing and tracking them, and structural metadata indicates how complex digital objects are put together so that they can be displayed or otherwise used.

Preservation metadata is information that supports and documents the long-term preservation of digital materials. It addresses an archived digital object's *provenance*, documenting the custodial history of the object; *authenticity*, validating that the digital object is in fact what it purports to be, and has not been altered in an undocumented way; *preservation activity*, documenting the actions taken to preserve the digital object, and any consequences of these actions that impact its look, feel, or functionality; *technical environment*, describing the technical requirements, such as hardware and software, needed to render and use the digital object; and *rights management*, recording any binding intellectual property rights that may limit the repository's ability to preserve and disseminate the digital object over time. Preservation metadata addresses all of these issues and more. In short, preservation metadata helps make an archived digital object self-documenting over time, even as the intellectual, economic, legal, and technical environments surrounding the object are in a constant state of change. The principal challenge in developing a preservation metadata schema is to anticipate what information will actually be needed to support a particular digital preservation activity, and by extension, to meet a particular set of preservation goals (Lavoie & Gartner, 2005).

The scope and depth of the preservation metadata required for a given digital preservation activity will vary according to numerous factors, such as the “intensity” of preservation, the length of archival retention, or even the knowledge base of the intended user community.

2.2 PRESERVATION METADATA AND THE OAIS REFERENCE MODEL

The OAIS reference model provides a high-level overview of the types of information needed to support digital preservation, including representation information, preservation description information (which can be broken down into reference, context, provenance, and fixity information), packaging information, and descriptive information. These information types can be interpreted as the general categories of metadata needed to support the long-term preservation and use of digital materials, and have served as the starting point for a number of preservation metadata initiatives.

Libraries and archives have taken different approaches to preservation metadata. For libraries, the central goal must be to preserve information integrity; that is, to define and preserve those features of

an information object that distinguish it as a whole and singular work. In the digital environment, the features that determine information integrity and deserve special attention for archival purposes include the following: content, fixity, reference, provenance, and context (Waters & Garret, 1996).

Content refers to the object to be preserved. Some preservation strategies, for example, format migration, require that an object be changed in order to be preserved. Content therefore might not always be defined as a particular set of bits but may have to be abstracted to qualities of structure and format, or even to abstract intellectual meaning. Fixity refers to the way that content was fixed as a discrete object, and mechanisms for preventing or detecting change. Reference refers to means of identifying, citing and locating digital works. Provenance means the record of the origin and chain of custody of the digital object. Context can be defined rather broadly as the ways in which digital objects "interact with elements in the wider digital environment." It included hardware, software and media dependencies, as well as linkages among digital objects and even "social context."

The drafters of the OAIS model moved these attributes into a metadata context when they used the same categories in the OAIS information model (Consultative Committee for Space Data Systems, 2002).

2.3 PRESERVATION METADATA FRAMEWORK

A number of institutions and projects have released preservation metadata element sets, reflecting a wide range of assumptions, purposes, and approaches⁴.

In 2002, the OCLC/RLG Preservation Metadata Framework Working Group consolidated existing expertise in the form of a preservation metadata framework. Using the broad categories of information specified in OAIS as a starting point, the Framework enumerates the types of information falling within the scope of preservation metadata. The working group then expanded each category of information, providing additional structure to articulate the OAIS information requirements in progressively greater detail and ending with a set of "prototype" preservation metadata elements (OCLC/RLG Working Group on Preservation Metadata, 2002).

The Framework effectively superseded the specifications it was based upon and represented "a good starting point for future practical implementations of preservation metadata. The Framework did not define a metadata scheme that could be used in practice by a preservation repository. It had no underlying data model.

2.4 PRESERVATION METADATA IMPLEMENTATION

In 2003 OCLC and RLG established another international working group to take the analysis of the Preservation Metadata Framework group to the next step, and develop an implementable core set of preservation metadata elements generically applicable to preservation repositories. Called PREMIS (PREservation Metadata: Implementation Strategies), the group was composed mostly of representatives of institutions developing or operating preservation repositories.

This group defined core metadata as "the things that most working preservation repositories are likely to need to know in order to support digital preservation" (PREMIS Working Group, 2005). The PREMIS Data Dictionary was issued in 2005.

⁴ Some examples are: National Library of Australia, 15 October 1999, *Preservation Metadata for Digital Collections*, <http://www.nla.gov.au/preserve/pmeta.html>; *Metadata for Digital Preservation: The CEDARS Project Outline Specification Draft for Public Consultation*, March 2000, <http://www.leeds.ac.uk/cedars/cedars.pdf>; the NEDLIB project: Lupovici, Catherine and Julien Masanès, July 2000, *Metadata for Long-term Preservation*, <http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm>

2.5 PRESERVATION METADATA AND PREMIS

The PREMIS data model defines five types of entity: Intellectual Entities (that is, conceptual objects that might be composed of one or more digital files), Objects, Rights, Agents and Events⁵.

The PREMIS Data Dictionary is organized around these types of entities rather than the categories of the OAIS information model, but they can be mapped to each other. Metadata pertaining to Objects includes what the OAIS information model would call Reference Information (identifiers), Fixity Information (message digests and digital signatures), some Context Information (relationships and environment), and Representation Information (object characteristics). OAIS Provenance Information and some Context Information is expressed through metadata pertaining to Agents and Events.

The PREMIS Data Dictionary defines preservation metadata as "the information a repository uses to support the digital preservation process." Some examples of preservation activities and how metadata can support them:

- A resource must be stored securely so that nobody can modify it inadvertently (or maliciously). Checksum information stored as metadata can be used to tell if a stored file has changed between two points in time.
- Files must be stored on media that can be read by current computers. If the media are damaged or obsolete (like the 8" floppy disks used in the 1970s) it can be difficult or impossible to recover the data. Metadata can support media management by recording the type and age of storage media and the dates that files were last refreshed.
- Over long periods of time even popular file formats can become obsolete, meaning no current applications can render them. Preservation managers must employ *preservation strategies* to ensure the resources remain usable. This might mean transforming old formats to newer equivalents (*migration*), or imitating the old rendering environment on newer hardware and software (*emulation*). Both migration and emulation strategies require metadata about the original file formats and the hardware and software environments supporting them.
- Preservation actions may entail changing original resources or changing how they are rendered. This can put the authenticity of the resource in doubt.

2.6 RIGHTS EXPRESSION LANGUAGES

The representation of rights information is necessary for future use of the digital objects and the rights itself. Digital archives and applications using the objects must be fed with rights information. Besides PREMIS there are other rights expression languages in use.

⁵ More information on PREMIS is to be found on the PREMIS standards page at the Library of Congress: <http://www.loc.gov/standards/premis/>

2.6.1 OPEN DIGITAL RIGHTS LANGUAGE⁶

The Open Digital Rights Language (ODRL) is a rights expression language developed to express rights, rules, and conditions - including permissions, prohibitions, obligations, and assertions - for interacting with online content.

ODRL was initially created in 2000 to address the needs of the DRM sector when media players were first introduced to the marketplace. The ODRL language, currently at Version 2.0, defines an information policy framework through publication of two specifications: the ODRL Version 2.0 Core Model, and ODRL Version 2.0 Common Vocabulary. Included within the ODRL documentation are a number of basic use cases demonstrating how to implement policy expressions using the Core Model with terms from the Common Vocabulary. ODRL is fully extensible and provides a mechanism for new communities to extend and/or deprecate the ODRL Common Vocabulary used in conjunction with the Core Model.

2.6.2 METSRIGHTS⁷

The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. METSRights allows the documentation of minimal administrative metadata about the intellectual rights associated with a digital object or its parts, and is to be used as an extension to the METS Standard. This Rights Declaration schema has 3 main elements:

- a simple declaration of type of rights (copyrighted, licensed, public domain, contractual, other) and the public statement of that Rights Declaration,
- the naming of the Rights Holder(s) with appropriate contact information,
- the Context(s) for the rights declaration based on type of users who have a set of permissions for a digital object or part of a digital object. If there are any constraints to the permissions, those are also expressed within the context by listing the constraints and explaining them in a constraint description element.

2.6.3 XRML⁸

XrML is the eXtensible Rights Markup Language which has also been standardized as the Rights Expression Language (REL) for MPEG-21. XrML is based on XML and describes rights, fees and conditions together with message integrity and entity authentication information. The eXtensible rights Markup Language (XrML) is becoming an increasingly popular language in which to write software licenses. XrML is also being used by international standard committees as the basis for application-specific languages that are designed for use across entire industries. For example, the Moving Picture Experts Group (MPEG) has selected XrML as the foundation for their MPEG-21 Rights Expression Language, henceforth referred to as MPEG-21.

2.6.4 COPYRIGHTMD⁹

⁶ <http://www.w3.org/community/odrl/>

⁷ <http://www.loc.gov/standards/mets/>

⁸ <http://www.xrml.org>

⁹ <http://www.cdlib.org/groups/rmg/>

CopyrightMD defines a set of metadata elements that can be used to express copyright information for digital resources. It is compatible with and can be used as an extension to the Metadata Encoding and Transmission Standard (METS). CopyrightMD is developed by the California Digital Library. There are three primary functional objectives of the copyrightMD schema:

1. To express factual information relating to copyright that allows users to make an informed copyright assessment of a given work.
2. To connect users to sources of further information about the copyright status, or to a person or institution that can be a resource for obtaining permissions relating to copyright-protected uses of the resource.
3. To explicitly associate item-level copyright information with discrete digital objects, which enriches digital objects with copyright metadata at the point of creation, thereby preventing the creation of "orphaned works" in the future.

2.6.5 MPEG21-REL¹⁰

MPEG-21 defines also a "Rights Expression Language" standard as means of managing restrictions for digital content usage. As an XML-based standard, MPEG-21 is designed to communicate machine-readable license information and do so in a "ubiquitous, unambiguous and secure" manner.

2.7 AN EXAMPLE

A good example of how rights metadata could be described and implemented is given by Stanford University: <http://www.loc.gov/standards/rights/METSRights.xsd>¹¹. This example shows the range of information that could be described in a Mets Rights schema and the way it could be described.

Another example is the use of PREMIS within a METS container at the National Library of The Netherlands.

```
<mets:rightsMD ADMID="RMDIP1384774082304-006" ID="RMDIP1384774082304-005">
<mets:mdWrap MDTYPE="PREMIS:RIGHTS" MDTYPEVERSION="2.1">
<mets:xmlData>
<premis:rights>
<premis:rightsExtension>
<kbmd:availability>free</kbmd:availability>
<kbmd:owner-id>00025</kbmd:owner-id>
<premis:linkingAgentIdentifier>
<premis:linkingAgentIdentifierType>KB-agent id</premis:linkingAgentIdentifierType>
<premis:linkingAgentIdentifierValue>14</premis:linkingAgentIdentifierValue>
<premis:linkingAgentRole>supplier</premis:linkingAgentRole>
</premis:linkingAgentIdentifier>
</premis:rightsExtension>
</premis:rights>
</mets:xmlData>
</mets:mdWrap>
</mets:rightsMD>
<mets:rightsMD ADMID="RMDIP1384774082304-005" ID="RMDIP1384774082304-006">
<mets:mdWrap MDTYPE="PREMIS:AGENT" MDTYPEVERSION="2.1">
<mets:xmlData>
<premis:agent>
```

¹⁰ <http://mpeg.chiariglione.org/>

¹¹ <http://www.loc.gov/standards/mets/news080503.html>

```

<premis:agentIdentifier>
<premis:agentIdentifierType>KB-owner-id</premis:agentIdentifierType>
    <premis:agentIdentifierValue>00025</premis:agentIdentifierValue>
</premis:agentIdentifier>
<premis:agentIdentifier>
<premis:agentIdentifierType>KB-agent-id</premis:agentIdentifierType>
    <premis:agentIdentifierValue>14</premis:agentIdentifierValue>
</premis:agentIdentifier>
<premis:agentName>Webloket monografieën</premis:agentName>
    <premis:agentType>organization</premis:agentType>
    <premis:agentNote>Vitha</premis:agentNote>
</premis:agent>
</mets:xmlData>
</mets:mdWrap>
</mets:rightsMD>

```

The example of the KB shows only a small range of the possibilities shown in the Stanford example. However it is much more realistic as content is usually not accompanied with a large set of rights information other than the information shown in the KB example above.

2.8 CONCLUDING

Since winning the 2005 Digital Preservation Award, the PREMIS Data Dictionary for Preservation Metadata has become the international standard for preservation metadata for digital materials. PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open source digital preservation tools and systems.

Any organization thinking about implementing a long-term preservation repository must think about preservation metadata. This includes not only information about the materials that will be stored in the repository, but also event or tracking information for pre-ingest and repository actions, and information about ownership, rights and permissions. Preservation metadata needs to be preserved in its own, that might seem obvious, but one has to think about it. How will it be preserved and what are the purposes of preservation? Rights information should be preserved along with the other representation information and the bibliographical information of records. The way that could be done depends on the goals you have. But it seems that the most common ground is to store preservation metadata (together with the other metadata) at least in the *Archival Information Package* (AIP) in the digital preservation system. In this case, however, the problem arises that rights and the legislative base will change in the future and that the respective metadata will need to be updated accordingly. In general the rights information in a metadata database can quickly be updated. It is recommended here that a system is used, which has revisions capabilities. This allows the tracking of changes on the managed metadata and the restoring of the original rights information if it is needed. If the rights information is stored in the AIP, the AIP needs to be updated with any change to the rights metadata – even though the related content files were not modified. An update of the AIP has to be documented in the provenance metadata and changes the bitstream of the AIP. That implies further the recalculation of the originally calculated checksums that are used to ensure the data integrity of the preserved bitstream of the AIP.

Scope and depth of the preservation metadata required for a given digital preservation activity will vary according to the purpose of use. This means there is no one-size-fits-all solution, and every repository must make and understand its own decisions.

Despite the impressive amount of effort that has been devoted to preservation metadata over the last decade, a great deal remains to be done. The standardization and refinement of preservation metadata

element specifications needs attention. As does the scope of use of preservation metadata, and the incorporation of more preservation metadata into more repository applications.

An interesting project in the latter field is the “preservation health check”. The Open Planets Foundation and OCLC Research are conducting a pilot that runs through 2013-2014. The activity involves the National Library of France as a pilot site that provides the preservation metadata from their operational repository and deposit systems. The project consists of a quality analysis of the real-life preservation metadata (METS/PREMIS) used by the pilot site, and intends to demonstrate the value of preservation metadata in mitigating risks by aligning the PREMIS Data Dictionary to risk factors¹².

¹² <http://oclc.org/research/activities/phc.html>

3 CLASSIFICATION AND EVALUATION OF DRM AND DRM SYSTEMS

Based on the definition of DRM given in the first section, this section provides an overview of the key DRM technologies, DRM systems and concrete examples of implementation. This is followed by an evaluation based on aspects related to the long-term preservation of digital materials featuring such DRM mechanisms. A full consideration of all available DRM tools and systems, as well as those which are under development, would go beyond the scope of this deliverable and so only a selection of the main representatives is examined; the approach of dividing the DRM systems into categories has been adopted. The evaluation in the second part of this section will therefore be based predominantly on the DRM categories thus identified.

3.1 CLASSIFICATION OF DRM AND DRM SYSTEMS

DRM has long been a subject of discussion e.g. in the entertainment industry, despite the fact that the term itself has only recently come into common use. The geo-location restrictions for playing DVDs based on regional codes is an example of access management. Other examples include so-called dongles, physical devices which are plugged e.g. into the USB port of a PC. These prevent e.g. the unauthorised use of software programs and are still used to protect transactions in home banking applications.

3.1.1 DATA CARRIER COPY PROTECTION

With regard to user management, the prevention of copying is a prime example within the context of the entertainment industry. One - albeit unreliable - method is e.g. the deliberate inclusion of errors in the data stream of an audio CD. These errors then prevent conventional CD-ROM drives equipped with error correction systems from reading the data stream, thereby foiling any attempt to copy the music from the CD to another data carrier. It should be noted, however, that the copying of data is an acknowledged long-term digital preservation measure. Users wishing to exercise their legal right to play the music on a CD-ROM drive are therefore denied such use. Not infrequently it also leads to problems in playing the CD in normal CD players or car audio systems.

3.1.2 LIGHTWEIGHT DRM

For the purposes of this report, lightweight DRM (LWDRM) refers to all mechanisms which do not of themselves restrict access to digital objects or their use, but which serve the **detection and tracking of legal infringements**. This is mostly achieved through the use of marking techniques such as digital watermarks. Digital watermarks may be applied to the digital object in a way which is invisible to the user but which allows the content providers to detect their works e.g. on illegal file-sharing sites. In music files, for instance, this additional information is embedded in the form of slight, audibly imperceptible frequency modifications. *Robust* watermarks, in contrast to *fragile* marks, can be destroyed by manipulating the data stream (e.g. through format conversion, compression). Fragile watermarks therefore also provide proof of the intactness and integrity of a digital object. This marking technique is deployed above all as a means of authentication in DRM tasks such as access management or usage accounting.

If the watermark is generated during the course of the purchase transaction and is linked to the object and the transaction, this allows any legal infringement to be traced back to the purchaser. The resulting deterrent effect is probably the strongest argument in favour of LWDRM.

Various methods are now available which can be used to implement this form of DRM. One example of a group of companies which has joined forces for this purpose is the Digital Watermarking Alliance¹³ (DWA), which also includes Philips. Philips offers NexGuard¹⁴, a marking technique for pay-tv broadcasts, in which digital watermarks are embedded in the video content. Digimarc¹⁵ is a

¹³ <http://www.digitalwatermarkingalliance.org>

¹⁴ <http://www.civolution.com/applications/media-protection/nexguard-paytv>

¹⁵ <http://www.digimarc.com>

further member of the DWA and its technology is used e.g. in Adobe Photoshop in the form of a fee-payable plug-in. This requires prior registration which allows Digimarc to build up an extensive database of artists, photographers and designers, including their contact details. Digimarc Guardian¹⁶ is also offered as the corresponding monitoring service for tracking the illegal distribution of objects e.g. on file-sharing sites.

A representative of the non-commercial sector is the Fraunhofer Institute for Secure Information Technology SIT (Fraunhofer SIT). The Institute offers solutions for marking images, eBooks, documents and audio and video material. It has developed its own tool, *ImageMark*¹⁷ which is available as an executable program and software library (Windows, Linux); it is suitable for use with most common image formats (e.g. JPEG, PNG, BMP, GIF, TIFF). The watermarks created using ImageMark are transparent and are reportedly very robust in preventing image processing actions such as

- JPEG compression
- Scaling
- Rotation
- Cutting
- Format conversion
- Colour/grey conversion

The Fraunhofer SIT offers *DocMark*¹⁸ for the marking of documents and eBooks. It is based on the technology behind ImageMark and deploys methods which embed information directly into the format or structure of the text. Watermarks generated using DocMark, it is claimed, are not lost even after activities such as printing and rescanning (analogue digital conversion).

Further parts of the Fraunhofer organisation include the institutes for Integrated Circuits IIS and Digital Media Technology (IDMT). These have developed an LWDRM method which offers customers two format variants (Grimm & Neubauer, 2004). In the first variant, *Local Media File (LMF)*, the digital object is tied to a single player (e.g. PC). This is achieved through cryptographic methods which take into account the hardware constellation of the PC. This controls access to the object, meaning that this variant does not actually constitute LWDRM as interpreted above. The second variant, *Signed Media File (SMF)* is signed by the user and can be generated from the original file or from an LMF. Users can then separate an LMF from the pre-selected player and e.g. copy it freely to their home network. A personal certificate and related registration is required for the signing, meaning that if such an SMF is found e.g. on illegal file-sharing sites, this could have legal consequences. The individual marking within the "creation and assembly" process (cf. SHAMAN LifeCycle) permits forensic tracking in the "reuse" phase.

In conclusion it should be clear that, despite the use of a range of different techniques, LWDRM is targeted primarily at the management and tracking of legal infringements and not at the management of access and use.

3.1.3 ENCRYPTION-BASED PASSWORD PROTECTION

This section focuses on DRM mechanisms which require no connections to external components (such as authentication servers) during use and which basically manage the **access and usage possibilities** of objects. The term "access" here signifies the opening of a file object using pre-defined player and display software - even though the act of opening could itself be interpreted as the most basic form of use. Use is therefore always conditional upon having access to the object.

¹⁶ <http://www.digimarc.com/guardian>

¹⁷ <https://www.sit.fraunhofer.de/en/offers/projekte/digital-watermarking/bild-wasserzeichen>

¹⁸ <https://www.sit.fraunhofer.de/en/offers/projekte/digital-watermarking/ebook-wasserzeichen>

If an organisation wishes to securely protect access or individual usage possibilities, the objects concerned must be encrypted. Without encryption, the content can always be viewed by interpreting the individual bytes of a bytestream (assuming the file format specification is known). The simplest way of restricting access is e.g. to package the file object in a container and to encrypt the container. The container format ZIP offers e.g. the ZipCrypto and AES-256 processes. It is not specified here, how users obtain the password required to open the container. Possible solutions include the individualised issue and distribution of passwords via a different distribution channel (e.g. by e-mail). Yet this method is clearly not secure, as once the unpacked file object has been saved, then it is free of DRM. Despite the rudimentary nature of this variant, it was used in certain cases until recently for the submission of eBooks and online dissertations to the German National Library (DNB). This works in a situation in which the “receiver” of the digital object is trustworthy and can guarantee that the unpacked digital object will not be used in an illegal way

The above example shows that reliable management of access and usage should be embedded in the file object itself, i.e. it should be an integral part of the file format specification. An example of this is Adobe's PDF format. It contains functions which render access and - as shown in the list below, above all usage is manageable in a variety of forms:

- Print
- Edit document
- Copy content
- Extract pages
- Add comment
- Complete form fields
- Add digital signature
- Generate new pages

Different levels of RC4¹⁹-based encryption can be selected. In addition, two different passwords can be used for access and for issuing usage rights. The format also permits specification of whether metadata and content or only content is to be encrypted. Microsoft Office 2007 offers similar functions for protecting Office documents such as Word files.

The fact that the standard protection system for PDF files presented here is not regarded as secure, primarily on account of the RC4 encryption routine, must be viewed critically. It is easy to find tools on the Internet which can be used to circumvent this protection mechanism. It is therefore a disadvantage that the password is integrated within the document itself and that the identity of the user is not checked. Anyone who knows the password can override the protection mechanisms with this form of symmetrical encryption. As a consequence, Adobe decided to refine its security functions - which led consequently to the kind of DRM systems presented in the following section.

3.1.4 DRM SYSTEMS

The DRM category presented in this section focuses not only on selected aspects already presented in section 1, but also attempts, by means of a system of diverse components and technologies such as the digital watermarks and encryption methods already examined, to cover all the core DRM areas. Moreover, this DRM category also refers to DRM in the narrower sense, although most of the presented concepts are applicable on any kind of an in-house Rights Management solution (DRM in the broader sense). However the exception is that in this variant the DRM mechanism is directly integrated into the content file. In cases of DRM in the broader sense, the entire rights and access management lies outside, for example as a part of the retrieval system or traditional Access & Entitlement systems that manage the access by IP-recognition. The latter means that the IP address of the user can be identified as being part of a group that has been assigned to a company or person or the

¹⁹ <http://en.wikipedia.org/wiki/RC4>

Internet Service Provider. This enables the access to be restricted to specific countries or institutions as an example.

The different DRM components of a DRM system can be geographically distributed and communicate e.g. via the Internet. This results in a range of dependencies which can affect everything from generation and content through to use. The client, e.g. the media player or the document reader, therefore no longer functions independently as a gateway to the actual content. It is apparent that precisely this interaction between the different components markedly increases the complexity of DRM systems in comparison to the DRM variants already presented.

The architecture of a DRM system is outlined in (Rosenblatt, 2005) and consists of the three linked components of content server, licence server and client. The first two components therefore form the back-end for the content provider in relation to the client on the user side. Figure 3 illustrates the interaction of these three components and their constituent parts. The content server has three further constituent parts: the content repository for the content itself, a database for the metadata which describe and mark content, and the DRM packager. The latter is responsible for combining content and metadata and provides both together as a content bundle using encryption mechanisms.

The diagram also shows how the DRM packager is linked to a key store which is part of the licence server. The licence server, for its part, generates and manages licences which are linked to the content and user. For this, both must be uniquely identifiable and the user and hardware-related usage rights must be specified.

On the client side, the DRM controller takes care of processing access requests from users. It verifies the user's identity, requests relevant licence information from the licence server and finally releases the encrypted content, including the licence, for a suitable media player (player, viewer, etc.) (Kuch, 2007, p. 26-27).

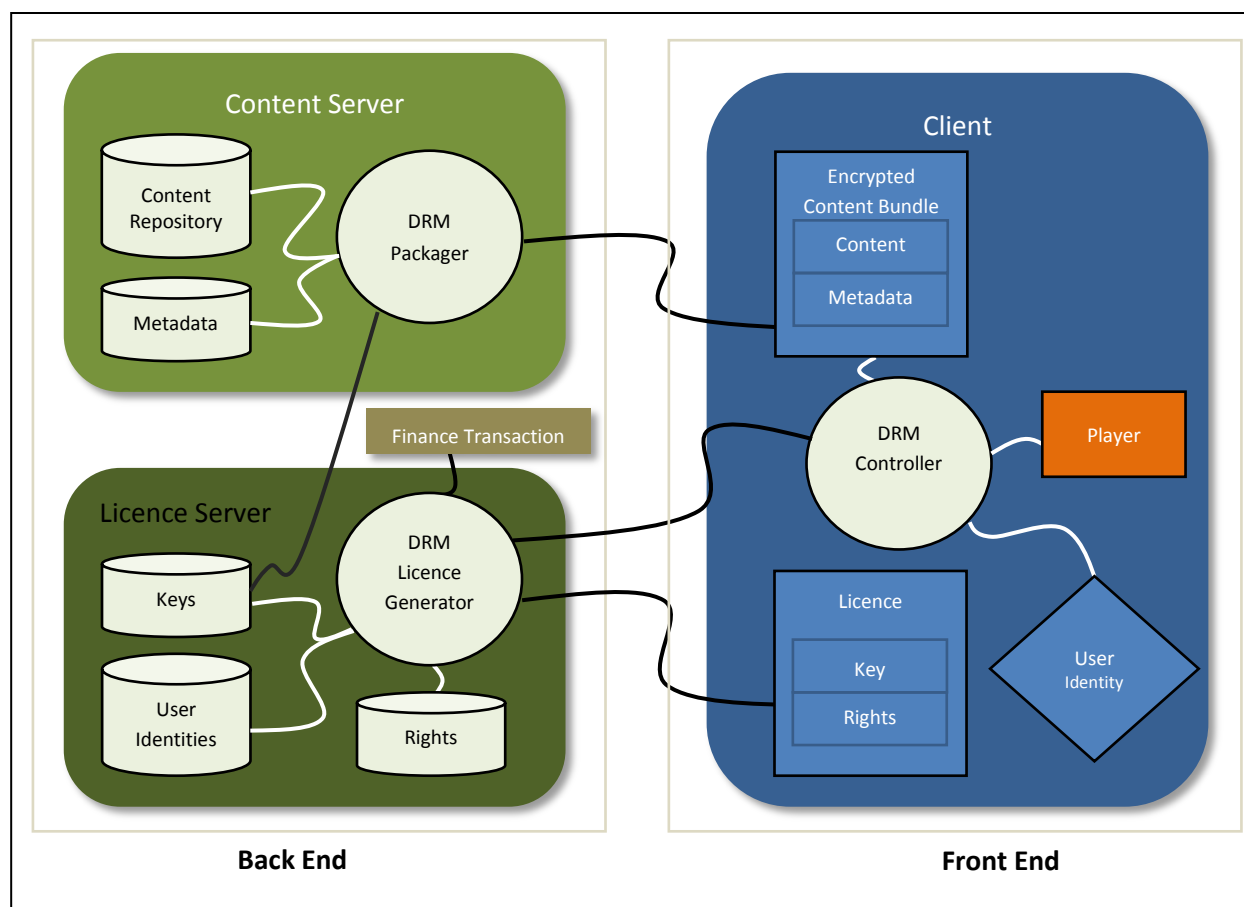


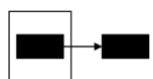


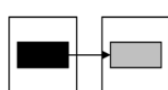


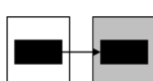


Figure 3: Architecture of a DRM System (adapted from (Rosenblatt, 2002))

In a DRM system, access is therefore only possible with a valid licence which is tied to the content by means of cryptographic methods, but the licence is not an integral part of the content.

The aspect of usage accounting is taken care of within this architecture in the connection between the DRM licence generator and a separate finance transaction component in the back-end. All access to the content has to be granted by the licence server, allowing it to record all usage activity.

In comparison to the usage restrictions on PDF files already described, a DRM system permits additional management possibilities for the use of digital content. Rosenblatt distinguishes here between the following possibilities:

Reproduction right	Transport right	Right to create derivative works
Print 	Copy 	Extract 
View 	Forward 	Edit 
Play 	Loan 	Insert 

Accordingly, DRM systems facilitate the known usage rights but they also permit the influence of content transport, especially through flexible licence models. This makes it possible to implement business models for the forwarding (reselling) and lending of digital objects.

STANDARDIZATION

Currently, there is no industry-wide standard for DRM. This would, of course, be highly desirable especially with regard to compatibility and interoperability. The restrictions on legally purchased content featuring DRM (such as music or eBooks) which confine use to the devices or the software of a particular producer is unacceptable for many users. Despite this, DRM is essential to protect the rights of creators and rights holders against infringements and illegal use or other forms of piracy. This also ensures the protection of the investment which has been made in the creation of these digital products.

In (Picot & Thielmann, 2005, p. 80) Microsoft demanded that it should be irrelevant how usage licences have to be transferred to the device or which hardware and system components the user makes use of. It should be of no consequence whether the operating system is from Microsoft, Apple, Linux or a relatively unknown provider, as long as it is capable of handling the DRM system and decrypting the transferred content. Microsoft has supported this approach by e.g. providing its DRM client technology as ANSI-C source, allowing these DRM components to be licensed by any manufacturer and integrated in their own platform, regardless of which operating system is actually used. Microsoft also uses the open rights description language *eXtensible Rights Markup Language* (XrML), thereby laying the foundations for interoperability between the DRM systems which the XrML can understand. Standardisation of DRM technologies brings benefits not only for users but also for long-term preservation. Standardisation leads, for example, to less dependence upon a particular product or producer. This makes it easier to exchange or update DRM components. Where open standards are used this also leads to the proliferation of tools (e.g. conversion tools) which are used e.g. to

implement long-term archiving measures. Open and well-documented standards also help ensure that components can be maintained and developed by a larger number of technicians (software developers). A good example of successful standardisation efforts is the Open Mobile Alliance²⁰ (OMA), a group of leading mobile telecommunications companies. The OMA developed a DRM which is now into its second version and which makes DRM-protected content available to mobile phones. It is now incorporated in many mobile telephones (Picot & Thielmann, 2005, p.112).

EXAMPLES OF DRM SYSTEMS

The best-known and most widespread implementations of DRM systems stem from large IT companies such as Apple, Adobe and Microsoft, although Apple has since developed into a major content provider like Amazon. Some open source implementations such as Open IPMP²¹ can also be found in this environment. It is to be expected that DRM will soon play a significant role within HTML5, too. Here it is planned to integrate the support of playing copy-protected movies without the need of third-party plugins like Flash and Silverlight²².

Audible

The Amazon subsidiary Audible²³ is one of the leading providers of audio books. The company not only offers content itself, it has also developed and successfully marketed its own DRM system. Audible provides its audio books in a personalised proprietary format (.aa) which requires the entry of a user name and password and a connection to an Audible server every time a book is accessed via the Audible Manager. Each audio book purchased from Audible is therefore marked and traceable. The Audible Manager functions as the DRM controller and permits e.g. the transfer of audio books to mobile MP3 players or the burning of CDs. If a compatible MP3 player is to be used, this requires prior registration and activation within the Audible Manager. This prevents copying of content to an indefinite number of MP3 players. There are now apps for smartphones and tablets which facilitate direct downloading and playing of the audio books.

Adobes LiveCycle Management

As mentioned above Adobe, too, has developed its document protection functions into a fully-fledged DRM system. Here, too, a native PDF object is marked. At the heart of the system is the Adobe LiveCycle Policy Server which plays the role of licence server. During the DRM enrichment process of PDF files (cf. DRM Packager), the selected usage rights are saved on this policy server. The usage rights are coded in a proprietary REL, preventing Adobe's DRM from being used with other document formats. The authorisations can also be assigned to individual user groups or individuals. After distribution, users on the client side must first log in to the Adobe LiveCycle Policy Server and provide authentication. The server safeguards the respective rights for the requested document. Furthermore, all the user's actions are logged and rights can be retroactively withdrawn by the rights holder (Kuch, 2007, p. 81). The usage rights can also be embedded in the document itself, meaning that the client software (in this case the Adobe Reader) can also grant access to the document when offline, assuming it has been appropriately configured. In the worst case, access to the content is denied if no contact can be made to the policy server²⁴.

FairPlay (Apple)

²⁰ www.openmobilealliance.org

²¹ <http://sourceforge.net/projects/openipmp/>

²² <http://www.pcworld.com/article/2052148/webs-gatekeepers-embrace-drm-for-next-html5-standard.html>

²³ <http://www.audible.com> and <http://en.wikipedia.org/wiki/Audible.com>

²⁴ <http://aecomag.com/software-mainmenu-32/57-adobe-livecycle-policy-server>

Originally devised for use with music, FairPlay is an integral part of Apple's iTunes music platform and is one of the most commonly used DRM systems. Since 2009, however, Apple has been offering music without DRM and so FairPlay is now only used for all other products offered in the iTunes store (films, books, apps). Apple's DRM system permits registered users to copy content to other devices to a limited extent. In order to use a device it must first be registered, like Audible. At present five devices can currently be activated for use at any one time.

Windows Media Rights Manager (Microsoft)

The architecture shown in Figure 4 of Microsoft's multimedia DRM system (primarily for Windows Media Audio (WMA) and Windows Media Video (WMV) files) is very similar to the reference architecture described by Rosenblatt in section 3.1.4.

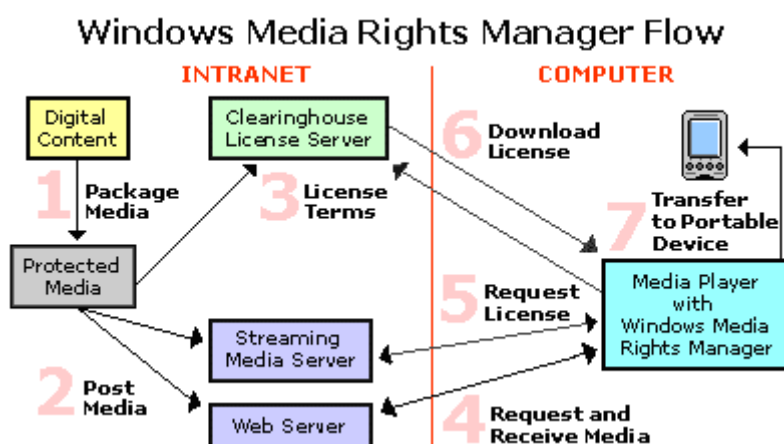


Figure 4: Architecture of Microsoft's Windows Media Rights Manager (from (Microsoft Corporation, 2004))

In the back-end, in this case the Intranet, there are the licence server (clearing house licence server) and content server (streaming media server and web server) components as described in Rosenblatt's draft. At the front-end, the DRM controller and player are combined in the Media Player with the Windows Media Rights Manager which also permits transfers to mobile devices. Rights information is represented in Microsoft's licence server by means of XrML. It permits e.g. the definition of the following usage possibilities:

- Frequency of use, number of plays
- Time-restricted use
- Restricted compatibility with certain devices
- Copy activation
- Permission to backup and recover the usage licence

The part of the life cycle described here is also based on the reference model. The process presented in step 1 includes the tasks of the DRM packager. It also covers provision on the content server (streaming media server and web server) and registration within the licence server (clearing house licence server) (Step 3). On the client side, use of the content starts with step 4, a request for the digital object made to the content server by the DRM controller (here the Media Player with Windows Media Rights Manager). Before the object can finally be activated for use, a valid licence must first be obtained which was requested in steps 5 and 6 (Microsoft Corporation, 2004).

In its Windows Rights Management Services (RMS)²⁵, Microsoft also has a DRM system with a stronger focus on documents for companies. The system can, however, also be used for other types of objects such as images.

3.1.5 SUMMARY OF DRM VARIANTS

The following table provides an overview of the DRM variants presented in this section. It shows the DRM aspects, the implementations and the technologies used.

No.	Type		DRM aspect	Examples	Technologies used
1	Data carrier copy protection		Use	Audio CDs	Illegal CD-TOC, deliberate reading errors
2	Lightweight DRM		Tracking of legal infringements	DigiMarc, Fraunhofer (ImageMark DocMark, LMF, SMF), Philips (NexGuard)	Digital watermarks, digital signatures
3	Encryption-based protection	password	Access and use	Adobe PDF, encryptable container formats such as ZIP, RAR, Microsoft Office	Encryption, REL
4	DRM Systems		all	Audible, Adobe LiveCycle Management, FairPlay (Apple), Microsoft Windows Media Rights Manager, Microsoft Windows Rights Management Service, Authentica Active Rights Management	Digital watermarks, signatures, encryption, REL

3.2 EVALUATION

The following section evaluates the DRM variants presented in the previous section in terms of the ability of digital objects which incorporate these technologies to undergo long-term preservation activities. It represents an appraisal on the part of the authors of this deliverable. This appraisal also

²⁵ [http://technet.microsoft.com/en-us/library/cc747763\(v=ws.10\).aspx](http://technet.microsoft.com/en-us/library/cc747763(v=ws.10).aspx)

contains a prediction component, meaning that 100% guarantees cannot be offered. If, for example, a DRM method is classified as low-risk for long-term preservation, this does not mean that the long-term archiving of materials with this protection will not give rise to problems in the future - even if the authors assume that it is highly improbable that such problems would be attributable to the DRM system used.

EVALUATION SCALE

A scale is introduced for the evaluation in order to make the DRM variants easier comparable, regardless of which file format. This is referred to below as the *Long-Term Preservation Risk* (LTPR) and is defined as follows:

LTPR	Characterization
no risk	No risk for future LTP measures
medium	Possible to use at present (at time of publication) in up-to-date hard and software environment, current LTP measures restricted, no external dependencies, medium risk for future LTP measures
high	Use and LTP measures already (currently) restricted, high risk for implementation of LTP measures in the future as result of external dependencies

The evaluation can basically be carried out for any DRM implementation. This should be conducted at the level of the DRM variant.

A DRM variant assessed with the first LTPR value poses no risk either for current use or for the execution of LTP measures (either now or in the future). An object featuring this type of DRM can therefore in all likelihood be ingested without any risk or need for any corrective measures into a long-term archive or repository. In the following it should become clear that “no risk” does not mean the absence of DRM.

With a medium LTPR, use is possible within the current framework. This includes correct installation of the hardware and software environment and also knowledge or provision of access details such as passwords. Offline use should be possible, i.e. with no external dependencies such as licences or authentication servers. It is assumed that LTP measures can be performed under these conditions, even though restrictions in characteristics such as sound quality may have to be accepted.

The use of DRM variants with a high LTPR carries risks at present. This hinders preventive LTP measures such as standardisation. This aspect combined with external dependency upon online components (e.g. licence servers) make permanent preservation and the execution of future preservation measures exceptionally difficult.

EVALUATION TEMPLATE

The evaluation is documented based on the following system:

DRM variant	Name of DRM variant (cf. Section 3.1)
-------------	---------------------------------------

LTPR	Assigned LTPR
Rationale	Why was the above LTPR given (examples)?
Restrictions	Which exceptions / restrictions should be taken into consideration in the evaluation.

EVALUATION OF THE DRM VARIANTS

1. Data carrier copy protection

LTPR	medium
Rationale	Data carrier migration is a key LTP measure, meaning that the prevention of all activities aimed at separating the data stream from the carrier should be regarded as risky. The data carrier copy protection currently prevents copying e.g. of audio CDs. If the data stream cannot be separated from the data carrier, this carries a high risk for future LTP measures because the necessary players and/or software may no longer be available. Use is, however, possible at present with common player devices (e.g. hi-fi CD players). Based on the principle of "what you can hear/see, you can copy", this permits LTP measures to be performed, albeit with restrictions e.g. in the form of loss of quality (digital-analogue conversion).
Restrictions	Some players already experience problems when confronted with data carrier copy protection. If it is possible to circumvent the copy protection, a lossless digital copy made at present would represent an LTP measure with negligible restrictions. This would therefore reduce the LTPR.

2. Lightweight DRM

LTPR	no risk
Rationale	Lightweight DRM involves no restrictions on access or use; the data stream is therefore accessible and the content usable at all times. The marking of digital objects therefore poses no risk for use or LTP measures.
Restrictions	When carrying out LTP measures such as format conversions, the use of digital watermarks can cause such marking information to be lost. The LMF defined by the Fraunhofer Institute (not strictly LWDRM) should be regarded more critically, as it can only be used on a single device. This precludes future use on a different device.

3. Encryption-based password protection

LTPR medium

Rationale	Access to the data stream and use of the content is predicated upon knowing the password. The password must be saved separately and linked to the actual content. The user must be given the password when access is granted. If only limited usage rights, such as text extraction, are granted yet the content can still be displayed, it can no longer be predicted with any certainty whether the conversion tool will require precisely this feature in the future. The execution of current and future LTP measures therefore carries risks.
Restrictions	Preventive conversion measures (normalisation) may be possible if the password is known, and this would justify an even lower LTPR. The same applies for the circumvention of protection features e.g. in the form of a brute-force attack. If the connection is lost between the password and the content, then the content is also lost.

4. DRM system

LTPR high

Rationale	Given that access to and use of the content is restricted similar to the "password protection with encryption" variant, objects protected by DRM systems also carry the same risks. A further problem factor is the existence of an external licence server, and connection to it is a precondition for encryption. Even today, use may be impaired or prevented entirely in the event of the content provider going out of business, network problems etc.
Restrictions	Fall-back solutions such as <i>Adobe's LiveCycle Management</i> can mitigate the external dependencies and, accordingly, justify the issue of a low LTPR value.

In summary it should be clear that the DRM techniques which should be regarded most critically are those which control access to and use of digital objects. If access to the content is already blocked, the problems involved in executing LTP measures are clearly apparent. Preservation measures without access to the actual content are not viable. Technical or other types of metadata (e.g. bibliographic) can only be extracted to a limited extent from protected files. According to OAIS, however, these data are incorporated in the data management and are essential for meaningful preservation planning and the execution of preservation actions. The encrypted content could also conceal malware (viruses, Trojans) which could enter the archive and remain undiscovered by virus scanners. User management should be analysed in more detail:

- What effect do restrictions have on the duration and frequency of use?
- What happens when usage rights later expire or are withdrawn?
- What happens when other devices are connected?

It should be apparent that time restrictions are basically impracticable for LTP measures. It is very difficult to define at present when an LTP measure, e.g. format conversion, should be conducted. Restrictions on the frequency of use would equate to restricting the number of uses of LTP measures. The question remains unanswered whether e.g. analysis tools for preparation or post-processing (e.g. quality assurance) constitute an incidence of use and therefore reduce the number of uses.

DRM gives rights holders the possibility to withdraw usage rights retroactively. Such withdrawal can affect all copies of a work currently in circulation. Naturally, institutions dedicated to safeguarding the cultural heritage find it difficult to reconcile this situation with their responsibilities.

The close ties to individual devices (cf. LMF) or categories of device should also be regarded critically. The future existence of such devices cannot be guaranteed - even if hardware and software museums are attempting to do just this. From the user's point of view, however, such attempts are seldom practicable and are of questionable longevity.

Considering the composition of usage rights as described by Rosenblatt, and given the uncertain nature of the future, the impression remains that all restrictions imposed upon reproduction rights, transport rights and rights to create derivative works pose risks for long-term preservation.

If it is accepted that providing the full range of reproduction possibilities should be the goal of a memory institution, then restriction of **reproduction rights** goes against the concept of ensuring long-term usability of the archived objects.

Restrictions on the **transport rights** inevitably interfere with the running of a preservation repository (data carrier migration).

The **right to create derivative works** restricts the usage possibilities. Even if restrictions such as prohibitions on text extraction only result in minor restrictions to use, conversion tools may rely upon them. The automatic extraction of bibliographic and structural metadata could also be prevented.

Even if there is already a solid, standardised foundation in the form of RELs such as XrML or ODRL for describing such rights, the mere existence of DRM mechanisms of this kind represents a further factor which contributes to the obsolescence of hard and software. This is because player devices and software age due to the ongoing development of these technologies (e.g. developments in RELs) and also due to technology development at the level of the file format, the viewer/player software or the devices themselves (e.g. DVD players).

4 REVIEW OF DRM USE AND DEALING WITH DIGITAL RIGHTS

4.1 INITIATIVES DEALING WITH DIGITAL RIGHTS AND DRM

In recent years, more and more initiatives related to Digital Rights Management (DRM) arose in the digital market with varying priorities, such as, the objective of simplifying licensing through cooperating with different organisations in the field of digital rights, or the focus of increasing visibility of necessary information regarding media and rights. These rights for object reuse or restrictions should be more transparent, simplified and easy to use. Another approach is to defend civil rights in the information society.

Consumers face a lot of confusion, once a movie, a piece of music, a picture, a text file and so on was found. How can this digital content be used and re-used? A good example for the clarification of this issue is the Public Domain Calculator which resulted from the Europeana Connect project.

The **Europeana Foundation** aims to make cultural heritage openly accessible in a digital way and therefore deals with rights management of data provided. The Data Exchange Agreement (DEA) is part of the Europeana Licensing Framework. It states that all descriptive metadata, provided to Europeana, can be re-used by third parties without restrictions (the extent of this information is reliant on the provider) and that all of those digital objects have to be provided with a rights label. There are four types of rights statements, which can further be divided into 12 statements (Europeana Foundation)

The **Public Domain Helper Tool** was implemented as part of the Europeana Connect project. It deals with content where the rights have already expired and passed into public domain across Europe, bearing national laws in mind. This tool provides a simple interface so that users can rapidly find out if the requested content lies in the public domain. (Angelopoulos & Jasserand, 2011)

It is a user friendly tool with a simple interface and intuitive workflow. The first choice one has to make is the jurisdiction, a country has to be chosen out of a drop down list. After this, one is guided through questions (see Figure 5). If a question or answer is unclear, the information icon can be clicked and a pop-up box opens with further information (see Figure 6).

There are three calculations offered. The *Button-Based Public Domain Calculation* offers only selecting answers from a list, where the *Form-Based Public Domain Calculation* provides text boxes to put relevant data in. The third one is the *Automatic Public Domain Calculation*, it is a mixture of text boxes and drop downs to find out a works status.

Into which of the following categories does the examined item fall? ⓘ

- ☐ First fixation of a film ⓘ
- ☐ Unoriginal Database ⓘ
- ☐ Broadcast ⓘ
- ☐ Performance ⓘ
- ☐ Phonogram ⓘ
- ☐ Literary or artistic work ⓘ
- ☐ Non-original photograph ⓘ

Figure 5: Public Domain Calculation - example of a question

Additional Information: Into which of the following categories does the examined item fall?

Please note that a single item might be protected by multiple layers of rights. For example, a CD will often comprise four layers of rights: if it contains music, that may be covered by copyright; any lyrics accompanying the music may also qualify for copyright protection, while the performers (musicians and singers or any other performer) as well as the phonogram producer may be protected by related rights. Similarly, a book may consist of text and illustrations, both of which may be protected by copyright. If an illustration is a photograph of a painting, a third layer of protection may be added. The term of protection of all relevant rights should be examined in order to determine whether the item as a whole is in the Public Domain or not. Please make sure you correctly identify and apply the Public Domain Calculator to all subject matter that qualifies for protection.

To this end, please also keep in mind that in accordance with Austrian law and for the purposes of this Public Domain Calculator:

- A volume, part, instalment, issue or episode of a work shall be treated as if they are whole independent works. Individual parts or issues of magazines, newspapers and other periodical works should also be considered to be independent works. The Public Domain Calculator should accordingly be applied to each of these individually.

Close

Figure 6: Example for a Pop-Up with information on the Q/A

A series of questions appear, depending on the previous answers selected.

One outcome, depending on the series of answers the user gave, would for instance be: “The rights in the unoriginal photograph have expired. Please note that additional rights might also apply, as a single item might be protected by multiple layers of rights. You may need to apply the Public Domain Calculator again to examine whether other relevant copyright, related rights or sui generis rights protection has expired.”

Furthermore, it is possible to embed the Calculator into projects in order to provide more technical information. (public domain calculation)

The following section gives a rough overview on further existing Digital Rights Management Initiatives and their central statements.

Informed Dialogue about Consumer Acceptability of DRM Solutions in Europe (INDICARE) is an open platform, built as part of a research and dialogue project that ended 2006. The paper, “Consumer’s guide to Digital Rights Management”, was provided to inform consumers about the topic of DRM. (INDICARE project, 2006) The project focuses on the usage within the movie and music industry, points out the consumer side of the whole topic and proposes ways of reconciling heterogeneous interests. INDICARE is financed by the European Commission under the eContent Programme of the Directorate General Information Society. (indicare.org, 2008)

European Digital Rights (EDRI) was founded in June 2002 and is an international non-profit association. There are 35 privacy and civil rights organisations that have an EDRI membership, which is restricted to not-for-profit, non-governmental organisations. The goal is to defend civil rights in the information society. (About European Civil Rights in Europe, 2010) The main keywords within this initiative are *privacy* and *data protection*. The goal is to spread the word of greater respect for and awareness of the fundamental right to data protection and to privacy for European citizens, as well as a safer Internet for children. (European Digital Rights, 2012)

Creative Commons is a non-profit organization with headquarters in the United States, their released copyright licences are free of charge to the public. It is an easy to use and easy to understand toolkit to adapt the licencing statements to the user’s needs. Creative Commons was inspired by the open source software movement. They inform the user about copyright issues and grant them permissions of use.

The focus here is on using and sharing creative work. Offered are six different licences that can be applied to one's work, depending on the purpose of it. (Creative Commons)

Persistent Identifier Systems (PI) provide an infrastructure to identify objects of any kind giving it a unique number-string. With DOI or other Persistent Identifiers such as URN systems it is possible to determine the ownership/authorship of a file, especially if it is embedded somewhere. One of the many purposes for which DOIs can be used is to enable automated copyright management, for this specific metadata on the rights is provided and available to the public. This metadata is persistently associated with the object identifier and can therefore be traced. Each entity is distinct. The Software Information Industry Association compares DOI with bar codes. It identifies a content and points it to a directory on the internet. This identification is a permanent number which never changes and therefore is consistent. Interoperability among media types is possible and handled at metadata level. DOI has become an ISO standard (TC46/SC9) and through this process it will update some of its parts in order to be compliant with other standards such as the US National Standard ANSI/NISO Z39.84, Syntax for the Digital Object Identifier. APARSEN has also conducted research into different PI systems and its interoperability within WP22 (see chapter 1.5).

Linked Content Coalition (LCC) was established in 2012 and is closely related to DOI. A goal is to offer an easy discovery of the ownership of content, no matter what kind of media type it is. Its focus is on the description of rights, the standardisation of metadata and the interoperability of databases. (CEPS DIGITAL FORUM, 2013) The International DOI Foundation will use LCC specifications in developing further DOI applications completing techniques already available for using DOI names with Linked Data applications.

The **RRM** (Rights Reference Model) is produced as part of the LCC framework. This model enables first of all the conversion of languages to the RRM with keeping its meaning. And second of all, it claims a standardisation and clarification of digital rights. (CEPS DIGITAL FORUM, 2013) It is extensible, flexible and publishers can optimize it according to their needs. Copyright lies with LCC but can be freely used by the terms of free use licence. It is adaptable to all types of rights and content and to different use and control of this data. (Rust, 2013)

Accessible Registries of Rights Information and Orphan Works (ARROW) is a standard-based rights management infrastructure. It enables exchange of information and right information management of text-based works in any digitisation projects. It helps in identifying rights-holders of media files and which steps one has to take to get the licence for a piece of work. (CEPS DIGITAL FORUM, 2013)

4.2 USER SCENARIOS IN THE WORLD OF LIBRARIES

The following section presents examples of user scenarios which are provided by the work package participants.

4.2.1 AUSTRIAN NATIONAL LIBRARY (ONB)

As the central academic library of the Republic of Austria, the Austrian National Library (ONB) looks back over a rich history in tradition going as far as the 14th century.

As a service-oriented information centre, the Austrian National Library offers its users access and professional advice on its own holdings (almost eight million objects) and also on international data pools. In addition, it accepts commissions on scientific research, and operates documentation centers and service facilities. Since the beginning of the digital age a steadily increasing part of the services are processed online via the website of the Austrian National Library. Because of the Austrian Media Law, the Austrian National Library receives – as the only library in the country – copies (Legal deposit) of all publications published in Austria, including digital media and dissertations approved by Austrian universities. On top of that, the Library systematically selects and collects literature specifically concerning Austria but published in other countries, and literature on the humanities with special attention paid to works that are relevant to its own departments.

Web@rchive Austria: Through a combination of the strategies known as domain harvesting, selective harvesting, and event harvesting, the aim is to capture as vividly as possible snapshots of Austrian web

space and perpetuate them for future generations. Web@archive Austria can be accessed on special terminals at the premises of the Austrian National Library.

DIGITAL RIGHTS AND LEGAL BACKGROUND IN AUSTRIA

Following laws provide the basis for activities of the Austrian National Library:

- Bundes-Museengesetz 2002 (Federal Law of Museums)
- ÖNB Bibliotheksordnung (ab 2.12.2009) (Austrian National Library rules)
- Mediengesetz (Austrian Media Law)
- Pflichtablieferungsverordnung (PflaV, ab 27.3.2010) (Legal Deposit Regulation)
- Urheberrechtsgesetz (UrhG) (Copyright Law)
- Datenschutzgesetz 2000 (Data Protection Act)
- ÖNB-Archiv-Verordnung (ONB Archive Regulation)

The new Austrian Media Law was approved by the National Assembly in January 2009 and became operative in March 2009. This amendment to the law is the legal basis for web archiving and governs the collection of online publications.

DATA TYPES AND VOLUME

In general the Austrian National Library collects offline and online material, e.g. CD-ROMs, DVDs, online publications, e-books, e-journals, images and internet sites in the Web@rchive mentioned above.

Following table gives an overview on the holdings of electronic documents at the Austrian National Library.

Type	Holdings on 31.12.2012
Austrian National Library overall	10.229.611
Electronic documents	
Electronic documents offline (physical unit)	6.456
Electronic documents online (bibliographic unit; excluding Web Archiving)	9.655
Web Archiving, stored (diverse) domains	1.177.810

The number of files saved in Web Archiving is 1.120.657.055 with a data volume of 28,57 TB. Digitisation of holdings required a data volume of 30,15 TB by end of 2012.

The proportion of protected material will be decreasing as data carriers as CD-ROMs and DVDs will decrease in volume.

DRM PROTECTED DATA

The typical data types stored at the Austrian National Library, which are protected are offline databases, videos, applications on DVDs, books like e-learning material, e.g. from Manz Verlag; the technical protection of off-line material lies below 1 % at the Austrian National Library.

Only if a media owner signs an agreement for access to the resource on the internet, the Austrian National Library releases the resource worldwide.

With the ingest of publications into the repository the data set with access rights is included, e.g. also if there are restrictions concerning number of concurrent user per minute. Those metadata are assigned in xml and are examined according to access rights controlled by the IP range. Licenced databases are not included.

Generally access to copyright resources is restricted to “single concurrent user onsite”. This implies to offline resources as well as to online-resources including access to the Web@archive.

The Austrian National Library has not implemented an in-house DRM technique.

LONG TERM PRESERVATION OF DIGITAL RIGHTS AND DRM PROTECTED MATERIAL

The Austrian National Library deals with digital rights according to copyright and copy-protected material since 2004. Copy-protected material is not accepted, apart from off-line material as CD-ROMs and DVDs, to ensure digital preservation.

The Austrian National Library always demands to get a non-protected version of the object by negotiating with the media owners. This strategy to exclude protected material from efforts towards digital preservation, involves the risk of having gaps in the collections.

According to the media law the Austrian National Library is allowed to create copies of an object for preservation purposes. The argument of proper archiving of digital objects mostly convinces media owners to provide non-protected material.

Another access restriction is password protection: the password has to be submitted together with the object for archiving. The case of encrypted material the Austrian National Library hasn't had before.

The Austrian National Library does not preserve digital rights and/or DRM information as protected material is not accepted. Offline material is not long-term preserved if it is copy protected. The Austrian National Library complies with digital rights including this information in the metadata description; preservation metadata, access rights are included in the metadata fields as well.

Austrian Books Online: This project is an example for the efforts the Austrian National Library is taking in regards of digital preservation. In this joint project with Google, the complete historical and public-domain book holdings of the Library are digitised. This comprises some 600.000 volumes including titles from the early 16th century up to works from the second half of the 19th century. The full scope will be approximately 200 million digitised pages.

The volumes that are being digitized in *Austrian Books Online* are available via the Digital Library of the Austrian National Library, as well as via Google Books. It is planned to make these items also accessible via Europeana (<http://www.europeana.eu>), the European Digital Library.

The first phase of the project was launched back in June 2010. Until the end of 2010 the prerequisites for the fully operative project were established and the digitisation proper started early 2011. The first 100.000 books were released in April 2013.

Further information on Austrian Books Online is available at <http://www.onb.ac.at/bibliothek/austrianbooksonline.htm>.

4.2.2 BRITISH LIBRARY (BL)

The British Library is the national library of the United Kingdom and one of the world's greatest research libraries. The Library's collection has developed over 250 years and exceeds 150 million separate items representing every era of human history and includes books, journals, manuscripts,

maps, stamps, music, patents, photographs, newspapers and sound recordings in most known languages. It is a collection ranging from 3,000 year-old Chinese oracle bones to the latest e-journals.

The British Library's Corporate Strategy for the period 2011-15 identifies five key themes, of which three anticipate increased and improved access to digital content:

- Guarantee access for future generations
- Enable access to everyone who wants to do research
- Enrich the cultural life of the nation.

To deliver the strategies embracing these themes a corporate Digital Rights Management 'solution' is required that promotes the widest permissible access to digital content, and facilitates the widest possible reuse of that content, while providing safeguards to identify and protect the rights of rights holders. The word 'solution' is used in the widest possible sense, and includes technical systems, business processes, and management and governance systems.

DIGITAL RIGHTS AND LEGAL BACKGROUND

Legal deposit for printed books and papers has existed in English law since 1662. It helps to ensure that the nation's published output is collected systematically, and as comprehensively as possible, both in order to preserve the material for the use of future generations and to make it available for readers. Legislation for legal deposit in the UK has been governed by the Legal Deposit Libraries Act since 2003. Publications deposited with the British Library are made available to users in its various reading rooms, are preserved for the benefit of future generations, becoming part of the national heritage. The Act introduced a framework in which regulations for the deposit of non-print works could be made. Pending the implementation of formal regulations, the British Library has encouraged voluntary deposit of offline, online or electronic items, which are managed through its digital storage system. In 2007, a voluntary deposit scheme was initiated by the British Library for UK publishers of scholarly electronic journals and from publishing trade associations and the other legal deposit libraries.

Selective archiving of UK websites (with permission) was started in 2005 in collaboration with the Joint Information Systems Committee, the National Archives, the National Library of Scotland, the National Library of Wales and the Wellcome Library. Website owners and rights holders are encouraged to give copyright permission for the British Library to make snapshot copies of their website at regular intervals for the UK Web Archive.

Legislation came into force on 6 April 2013 extending the principle of legal deposit from printed publications to digital materials. The six legal deposit libraries, now have the right to receive a copy of every UK electronic publication, including blogs, eBooks and the entire UK web domain.

Throughout 2012/13 a strategy was developed to implement the new regulations. Henceforth we will be able to collect, preserve and provide long-term access to this increasing proportion of the nation's cultural and intellectual output.

As well as the deposited collection mentioned above, the British Library's digital collection includes donated items such as personal archives, purchased items, and large numbers of items created by digitising the Library's "analogue" collection such as manuscripts, books, newspapers and sound. Digitisation helps to preserve originals by providing a surrogate and thus reducing handling of the original, but also enables increased access to the collection.

Rights information for major publisher subscriptions and commercial licences for access to digital content are held in spread sheets and within paper licences in the Corporate Procurement Unit and the Copyright and Licensing Teams. Some information on rights is held within the Digital Asset Register, which is a spread sheet maintained by the Digital Scholarship team.

DATA TYPES AND VOLUMES

The Digital Library System (DLS) provides a shared technical infrastructure for non-print legal deposit in the UK. The volume in GB of digital content within DLS is given in the chart below. The

figures are given as percentages where the total volume is 346,388GB. The data provided covers the period March 2006 to November 2013.

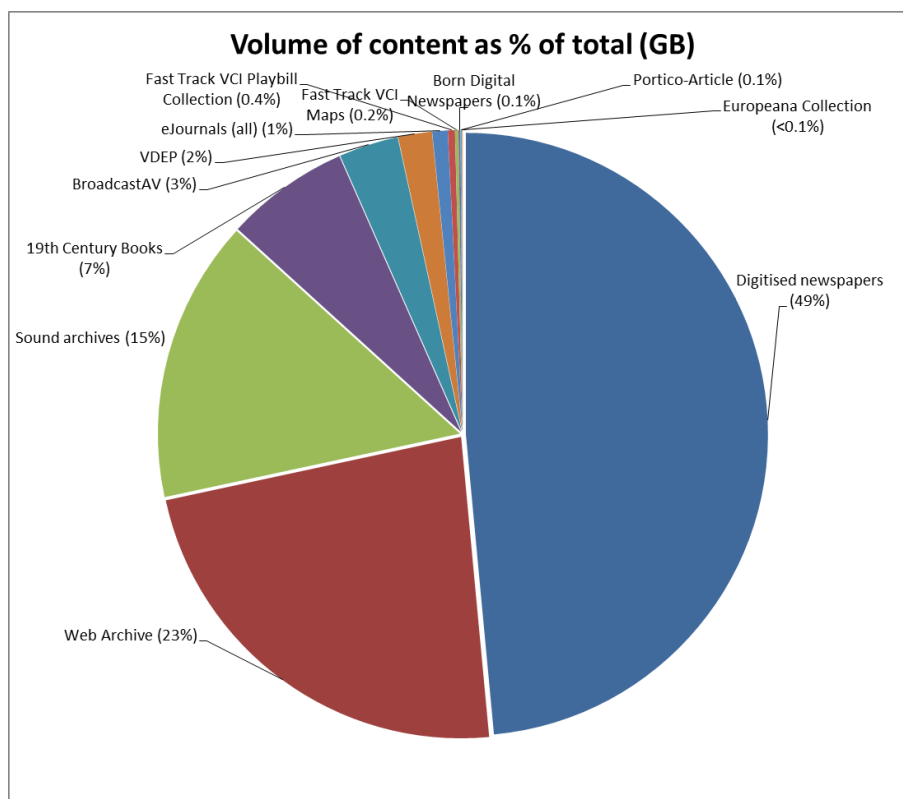


Figure 7: Volume of content at the BL

The digital objects in the DLS are in a wide variety of file formats. Some of these are container formats such as zip or gz, which contain other files of various formats. For some ingest streams the file formats are controlled, for example from internal digitisation projects. For others the Library has no control, such as with voluntary deposit or web archives. The following list gives examples of the file formats held: xlsx, xls, xml, jp2, gz, wav, jpg, pdf, tif, zip and doc. The zip files currently mainly contain jp2 and xml files and are generally from digitisation ingest streams. The gz files are generally web archive files, and although their contents are dominated by htm and gif file formats, there could be any type of file that can be part of a web page. The complexity of the objects and the range of formats held therefore has the potential to be huge.

The distribution of digital object sizes varies. Most of the very small objects are individual e-journal articles. The size distribution will change over time as the number of web archives in the collection increases with legal deposit, because these are typically around 1GB per object.

The projected increase in digital object storage volume per storage node over the coming years is that by 2020 it is expected that there will be between 4.5 and 5.0 Petabytes per node, totalling around 20 Petabytes for all four nodes.

DRM PROTECTED DATA

A project²⁶ on DRM Requirements was carried out from January 2013 to October 2013. The project looked at a DRM solution which would allow the BL's digital collections to be made available under the widest possible usage terms, whilst preventing digital content being used by unauthorised users, or

²⁶ This project, carried out internally by the BL, was led by Lucie Burgess, Head of Content Strategy Research and Operations, as the Project SRO, and David Sweeney, as the Technical Project Leader.

used by authorised users in unauthorised ways. The project delivered a set of high level principles required for designing the ideal solution. The focus then moved on to the implications of these principles for the BL and its users. The ideal DRM solution along with ideal policies and business processes was identified and specific high level outputs of this project will be taken forward by the BL in 2014.

The Library has a number of systems that contain DRM protected data. These include; the Digital Library System, BLDSS, Images Online, numerous public domain projects hosted by the Library, electronic theses online (eThOs), Endangered Archives, access to electronic resources in the reading rooms etc. Some details are provided below:

Digital Library System

The Digital Library System (DLS) has in place the following usage rights:

- Unrestricted use – can be made freely available in the reading rooms, and online
- Access without restriction to all users in reading rooms for content which is not designated as non-print legal deposit
- Non-print legal deposit (usage terms as per legal deposit regulations)
- No access to users, but access permitted to BL staff
- Other – the usage restrictions do not fall into one of the above categories, so access is not currently provided

Voluntary Deposit Content

Prior to Non-Print Legal Deposit regulations coming into force in April 2013, The Library has been collecting content through the Code for the Voluntary Deposit of Online Publications since 2000 ('VDEP'). VDEP content includes electronic newsletters, books, serials and other content generally sent to the BL by e-mail and is quite diverse. Access to VDEP content is currently not accessible in the reading rooms. Cataloguing the content remains a manual process. VDEP content has been ingested without digital rights metadata attached although this scheme is no longer in force. A large volume of complex usage rights information is held in spread sheets and not in machine-readable metadata, and therefore there is no automated way of dealing with these usage rights.

Non-Print Legal Deposit Collection

Since 6 April 2013, regulation for the Legal Deposit of non-print publications has been in force. This supersedes the VDEP scheme. As a result of the implementation of the regulations, business processes for ingest of digital content have changed. The digital usage rights for Non-Print Legal Deposit (NPLD) are well-defined and understood; they are enshrined in the Regulations, and are applied automatically to all digital content ingested. Content is accessible through a secure access system which is available on all PCs in the reading rooms, restricting access to a single concurrent user in each Legal Deposit Library for each single item of content, limited printing and no digital copying.

Subscriptions and e-resources

The British Library subscribes to, licences and purchases digital content for use in the reading rooms. Access to subscribed, purchased or licenced eBook and eJournal content not held in the DLS is also available within the reading rooms. Most of these collections are subscribed to by the Library and, under the terms of the subscription agreement, are only available either via British Library PCs within the reading rooms or to registered users on Library premises. A small but growing number of electronic resources have been licenced for online use by registered readers. Stakeholders expect there to be significant increases in the amount of content we licence for use by registered readers online. Restrictions on reading room use for all content are enforced by a block on download and copy capability on reading room PCs (additional restrictions apply to non-print legal deposit content). These terms and conditions of use are accepted by the reader during the registration and renewal process. In addition, separate usage rights exist for unsubscribed and subscribed electronic journals for Document Supply purposes

Audio-visual content

Rights issues surrounding audio-visual materials are variously complex: involving rights in recordings or performances; in the song, film, underlying work; with different lengths of term; and involving many different rights holders. Sound and moving image material is catalogued, although not all usage rights are currently captured. Access to audio visual material is primarily restricted to the reading rooms, providing access to over 100,000 recordings in the reading rooms with 50,000 sound records being rights cleared for global online access (listening only). Additionally, download is available for users in higher education and further education, through a separate platform provided by JISC. Broadcast news content is captured from 15 UK broadcast TV and radio channels, and is made available in the reading rooms only.

Digitised content

The British Library's technical infrastructure and rights management processes around digitised content are changing and developing. The British Library is digitising a significant amount of out-of-copyright content, and a growing amount of in-copyright content, funded through third-party partnerships. The IP and licensing team has simplified its third party digitisation agreements to a standard set of licence terms which govern usage rights. Most commercial digitisation projects result in paid-for use of a digital asset using a third-party platform with a licence for the British Library to provide free access to the content within the reading rooms during the term of the contract. The Google books project will provide public domain access to 400,000 15th-17th century European books through the Google Books site, and eventually through BL infrastructure, once the digitised content has been ingested into the DLS. The Europeana WW1 digitisation project will provide online, rights cleared access to 200,000 digitised images through the BL's own infrastructure. Any content that is held in the DLS, and which has "Unrestricted Access" usage rights associated with it, is available within reading rooms as well as externally over the Internet. Some are restricted to the reading rooms as they are still in copyright. Currently, work is being undertaken to develop generic metadata profiles and routines to ingest content, which includes basic metadata relating to usage rights.

British Library Document Supply Service

The British Library Document Supply Service (BLDSS) is a comprehensive online document supply service that draws on over a dozen different software components to provide an integrated end-to-end service. The system can deal with some relatively complex document supply rights requirements, but is currently unable to deal with more complex rights issues.

During the acquisition process, or at ingest, permissions (item status and acquisition method) and copyright information is recorded in the bibliographic record, and this information is populated. The BLDSS access component (which matches item right permissions against user rights permissions) queries this data to validate whether or not an access request should be permitted. The access component takes into account permissions information. Each registered user of the service is obliged to accept specific usage conditions in order to use the service. Once registered, the user is identified as having a specific type of user profile, with an associated and defined set of usage rights.

ACCESS CONTROL

Access control, another key aspect of DRM, is specific to each service under which content is delivered. For example, the Ericom system delivers its own access control for digital legal deposit content which enables content to be delivered in a manner consistent with the regulations. Document Supply has its own quite sophisticated rights management and delivery module which enables document supply content to be delivered to users under terms consistent with document supply licences. Third party services provide access under terms consistent with third party licences. Authentication of users according to their role is a pre-requisite for an automated DRM solution to work.

The DRM project has delivered a rights taxonomy to replace the Cosima Stanford rights metadata syntax that has been implemented in the Metadata Extension Repository (MER) which is currently under development and will set out 19 usage rights definitions.

FUTURE DRM REQUIREMENTS

When considering the design of a future DRM system, the following requirements should be met:

- The DRM approach will be unified and comprehensive where best practice is established and solutions are re-used;
- DRM systems should have a high-level of automation, balanced against business processes requiring human intervention that are standard across business areas and services;
- Usage rights definitions should be simplified and streamlined with agreed definitions that will be subject to controlled change. The usage rights matrix should be core to the DRM system;
- The system should be fully scalable and able to handle DRM for increasing volumes and new types of digital content and services;
- The system should be flexible, achieved through standardisation of process and extensibility of components to allow for the integration of new services;
- The system should be seamless, with all DRM components linking to common data held in centralised repositories and machine-readable databases. Data should be live, reliable and reusable.

What does this mean for our users?

Users of the British Library will be able to access and use permitted digital collection items through appropriate services, and use them in a way that complies with the conditions of the access service and the item's usage rights.

Registered readers using the BL Wi-Fi would be able to access and use content anywhere on Library premises, not just within reading rooms, although a DRM capability is not the only barrier to making this happen. Where permitted, registered and authenticated readers will be able to access and use content remotely, with DRM managed content delivered directly to their devices. A good DRM capability will enable richer user services, enhancing the researcher journey and experience with the BL.

Staff, and potentially users of the BL web site, will be able to search for certain types of content and then be able to reuse digitised material in other contexts. This is particularly the case for; public domain, creative commons, and openly licensed content, and orphan works; and potentially also for in-copyright rights reserved free access content, depending on the licence.

What does this mean for the BL?

New services and market opportunities can be exploited once the usage terms for BL digital items are made clearer. Certain categories of digital content will be made available for customers and partners to use in new and innovative ways, enabling certain levels of access to everyone who wants to do research. For some categories of content this will mean that physical barriers to access can be removed. Stakeholder interviews with the business areas have revealed a desire to challenge established boundaries regarding access to digital content.

In order to promote increased access to our collections, and develop new business ideas and services, a comprehensive DRM system would assign usage rights or user licences to collection items and permit subsequent access to digital content. The desire to make more collection items available has to be balanced against the need to respect intellectual property rights, and to prevent unauthorised access and usage of digital content.

What is the ideal DRM solution?

A comprehensive 'DRM solution' includes management of digital IP policies and application of usage rights, business processes and technical components (both software and hardware) both to control access, track and maintain usage rights. The ideal future 'DRM solution' could encompass systems and processes that deal comprehensively with everything connected with usage rights, access control, licences and permissions.

Any automated DRM solution will not function without a set of defined and agreed usage rights, as this is the cornerstone for any rights management system, so whatever the DRM system of the future is, this will need to be underpinned by a usage rights taxonomy.

The technical DRM component would consist of, at least the following main components (please note this is not an exhaustive list of all requirements):

- A master repository for all digital usage rights held as metadata, with the ability to track and audit changes
- A Graphical User Interface (GUI) to enable recording and editing of usage rights metadata;
- The ability to query the master repository, for example to find all content with particular usage rights attributes;
- The ability to authenticate users via a customer management infrastructure, because usage rights are a function of the user's role and location;
- Access control components to evaluate, for any particular request, whether or content should be delivered to the user, according to their role, location and the usage right being applied to the content, and the time at which the request is submitted
- Integration with applications and services so that usage rights can be appropriately displayed
- A management information component that gives business areas access to reporting information, facilitating good decision making;
- A loosely-coupled set of components that enable any one component to be replaced without needing to redesign the others.

Ideal Policy

Governance and policy-making for changing or adding usage rights will be streamlined and controlled. Decisions will be made on a scheduled basis by appropriate individuals or groups. Business processes will be designed to connect the policy-making and decision-making with the needs of the business areas to make content available, and to add or change rights and permissions.

Ideally, the decision making process on application to make new content available will be agile, and allow for quick decisions to be made when necessary. Policy decisions about removing access permissions or "takedown" of content have to be quick and should therefore be a designated responsibility for a specified post or department.

Ideal business processes

Where feasible, business processes will be designed to fit in with the principles of delivering a highly-automated system. The business processes are the tasks that will be performed that will create, delete or change rights metadata, for example: recording publisher licence information.

There will be defined specialist areas responsible for deciding the rights status of digital objects, for developing policies that balance the need managing a diverse set of rights against the need for simplification, for consulting widely regarding changes to policy, and so on.

The technical processes will, where possible, utilise publisher-supplied metadata to determine access rights at the point the content is ingested, but this will need to be supported by business processes for manually recording access rights information. The ideal business process for recording publisher licence information would be for a simple user interface to be available, where the licence terms can be recorded directly into a database.

Business processes are to be found everywhere in the current system and their efficiency generally reflects the maturity of the digital services to which they belong. A new DRM system would reduce manual processing and introduce new business processes to increase efficiency and maximise automation.

CHALLENGES

Our key findings of the DRM Requirements Project (January 2013 to October 2013), showed that rights issues are complex which include contractual agreements and the application of legislation to digital material. These complexities also relate to our content, business processes and systems. One of our key objectives was to simplify and unify this complex world as much as possible. The project identifies a number of key issues and challenges which are to be reviewed with specific tasks being taken forward in 2014. The challenges are listed below:

- **Vocabulary and terminology** must be defined, understood and agreed
- **DRM has a major impact on business processes.** In many areas our business processes are manual and could be automated
- **Maintain digital rights** as they change over time
- DRM capability to support **copyright and IP policy** as well as access and re-use policies
- **IT architecture** is complex. The main areas which need to support DRM are ingest processes including strategic ingest, our catalogues and the DLS; also our services and applications
- **Legacy digital content**, to which DRM needs to be applied; taking into account new content streams – more digitisation projects, more open access content
- Access control for our **services and applications** at the level of the individual application, needs relevant rights metadata so that content that can be used in different services
- **Licences and contracts** need to be automated with machine-readable rights metadata by focussing on the terms of use
- Finally, the **skills** to apply rights metadata need to exist across a number of operational processing areas

4.2.3 GERMAN NATIONAL LIBRARY (DNB)

As the central archival library for the Federal Republic of Germany, The German National Library (DNB) has a legal mandate to collect, catalogue and permanently archive German and German language publications.

Before 2006, DNB collected online publications on a voluntary basis. To accommodate for preservation needs many experimental approaches were tested. These activities prepared DNB for the upcoming legislative mandate. The approaches remained modular and were not integrated into the key operations of the organization. In 2006 DNB was legally mandated to engage in large scale digital preservation. Since then digital preservation activities were successively implemented in DNB workflows. The legal mandate helped to divert funds to the task.

The DNB hosts also the German Music Archive²⁷, which has the legal mandate to undertake the state-wide central collection of sheet music and sound recordings and serves as the music bibliography information centre for German.

DIGITAL RIGHTS AND LEGAL BACKGROUND

DNB takes care that all digital publications can be utilized in accordance with legal regulations. Depending on the rights that the content producer grants us during the submission process, some publications can be provided in-house only, while others are remotely accessible.

Format and way of delivery is determined in the Law regarding the German National Library (DNBG²⁸), the Legal Deposit Regulation and in accordance with the Law on copyright and related rights of legal protection.

All registered DNB users and deliverers accept a data protection statement.

Deliverers must confirm that they are entitled to deliver the publication as a deposit copy to DNB with the rights necessary for the legal mandate. Beyond the exceptions for use of the German Copyright Act the right holder may grant following rights:

- a) access for registered users via internet,
- b) unlimited access.

²⁷ http://www.dnb.de/EN/DMA/dma_node.html

²⁸ http://www.dnb.de/EN/Wir/Recht/recht_node.html

DNB receives DRM protected material but does not produce material that is DRM protected. In general publications which are published by the DNB are DRM free. Also DNB advises its deliverers to abstain from the use of DRM mechanism for the delivery to the DNB.

DATA TYPES AND VOLUMES

The following table gives a short current overview about the data volume and its annual increase for the most common data types.

Type	Annual quantity ²⁹	Total stock	Total size in TB	Average size in MB
Theses	13.745 titles	139.057 titles	~ 3	22,8/object
eBooks	153.669 titles	513.097 titles	~ 24,4	50/object
ePapers	161.126 editions	372.160 editions	~ 7,6	21/object
eJournals	5.433 magazines/ articles	28.479 magazines / articles	~ 1,36	50/object
DNB digitized print media	300.000 pages	~ 3 million pages	~ 95	11,5/page
Digitized audio	~ 100.000 CDs	450.000 CDs	~ 150	355/CD

DATA VOLUME AND ANNUAL INCREASE

The table shows that roughly 280 TB of data is split in categories with high divergences in storage size and object quantity. Meanwhile the total amount of digital born publications has increased to over one million objects.

So far, DNB has focused mostly on static online publications that are collected mainly in PDF or, recently, in ePub format. We accept and archive any PDF version that is submitted to us. Also we advise our deliverers to use PDF/A if it is possible.

We are in the process of gradually digitizing all audio CDs of the German Music Archive since 1983 by using the uncompressed Broadcast Wave format.

In order to fulfill the collection mandate in the field of internet websites, DNB is currently working on a project to establish the basic organizational and technical principles for automatically harvesting websites, so-called web harvesting. We are also seeking to collaborate with external partners in this field.

In the past DRM mechanism of digital objects were only detected manually. However, no statistical recordings of DRM mechanisms detected were implemented. It can however be assumed that the proportion of DRM protected material has been increasing in parallel to the further development of DRM techniques and format capabilities.

Approximately one third of the online publications are Open Access content. That means that the access to that type of content is unlimited and also available via internet without user registration.

²⁹ The quantities of theses, eBooks, ePapers and eJournals are figures from 2013, the others from 2012

MEASURES FOR SAFEGUARDING DIGITAL RIGHTS

The following measures are related to publications which are not considered to the category of Open Access.

The access to the reading room, access to digital holdings via the DNB portal requires the registration and authentication of the user.

The access mechanisms provided via the DNB portal ensure compliance of digital rights of the collected publications and digitalized music, generally speaking to all digital DNB holdings. Therefore the user can only access the publication within the reading rooms of the DNB on specific types of computers. To restrict copying these “read only” computers have the following limitations:

- no connection to the internet,
- no drives for writeable medias (Floppy Disk, CD-ROM),
- no USB port,
- access only via a specific viewer/player software.

In equivalence to the Law regarding the German National Library (DNBG) and its regulation that publishers have to deliver two copies of every published physical book, it is only allowed that two users access a publication at the same time.

The installed viewer software allows only the printing of a publication. Thus the user can't save the file under another location, and in particular not on an external storage medium. On every reading computer and printing station, the user gets a warning that he/she is also obligated to comply with the German Copy right Act. For printing that means that only 15% of the pages of a publication are allowed to print.

Currently there are no mechanisms to maintain and preserve digital rights information on the publication level. The access and use is assignable on the level of the mentioned data types above and is oriented on the respective legislation that is valid at the time of access. The data type is recorded in a data field on publication level and that is part of the catalogue metadata.

DRM PROTECTED DATA

The following data types are occasionally submitted with integrated DRM measures to the DNB:

- Doctoral theses and teaching theses of German universities
- DNB digitized print media
- eBooks
- eJournals
- ePapers

The use of DRM techniques and tools depends on the file format and its capabilities, the data type and the publisher. The following techniques were detected so far:

- PDF document restrictions (password protection and print, copy restrictions)
- Adobe DRM (mostly publishers)
- encrypted ZIP container

THE APPROACH TO ENSURE THE LONG-TERM PRESERVATION OF DRM PROTECTED MATERIAL

Since the end of 2012, DNB uses tools to detect DRM measure of digital objects during the ingest process. Before that time the detection were manually done by random sampling. DNB considers DRM measures as a potential risk to fulfill its legal obligation.

Protected publications are likely to cause difficulties throughout the preservation process and access now and in the future. A password protected file is hard to convert in another format. Also the success of preservation actions by emulation is threatened.

In accordance with the decision to preserve unaltered originals and to abstain from normalization measures at the time of ingest, the DNB tries to collect the unprotected version of the digital object whenever it is possible.

The approach for online publication contains the decision to refuse “DRM suspicious” material after detection and give the publisher or the delivering institution the possibility to remove the protection for a second delivery. DNB does not modify the publication ourselves.

For the automatic detection DNB uses the support of open-source tools. In the case of encrypted ZIP containers the regular unpack routine would report the protection measure. For some time now the automatic generation of technical metadata using metadata tools has been a recognized and established component of the ingest process. The DNB has long been using the File Information Tool Set (FITS)³⁰ as a framework for using an entire tool set. This framework provides access to a whole range of tools including the JSTOR/Harvard Object Validation Environment (JHOVE and JHOVE2)³¹ tool, the Digital Record Object Identification (DROID)³² tool and the NLNZ Metadata Extractor. Use of a tool set widens file format support and reduces the risk of errors in the identification and validation of the file format. Some of the above tools (e.g. JHOVE) also permit the recognition of document restrictions such as password-protected PDF files.

According to its legal mandate the DNB takes preservation actions like migration on archived publications. Where DRM mechanisms inhibit preservation actions, the law about the German National Library (DNBG) principally permits DNB to remove the respective DRM. In particular this is important for post processing the stock of already archived objects, which have unrecognized DRM mechanisms.

LIMITS

One limit of the approach of refusing “DRM suspicious” material lays in the limited capabilities of the used metadata tools. So the tools have to be up to date to support new formats and format versions. Unfortunately FITS is not able to determine all variants of PDF restrictions. But if that would be possible another question would arise: Which restrictions are real risks for long-term preservation activities? If the user is not allowed to print the document, it need not necessary be a risk for a conversion in the context of format migration actions. In cases of format transformations a further question still arises as to whether and how such usage restrictions should be preserved.

The alternative approach of removing DRM mechanisms implies many problems in itself. Removing technical mechanisms needs corresponding tools and might change the authenticity of the object. In general it is not easy to acquire a software tool that violates the current legislative. If there aren't any tools or they are not allowed to use the last approach for encrypted documents could be trying every combination of possible password characters. That approach is known as a brute-force attack and is very expensive, because it needs a lot of hardware resources like processor time. For long password lengths it takes a very long time to crack the password, in the worst case the cracking attempts are nearly infinite.

Finally it isn't always possible to get a DRM-free version of a publication, especially if the delivery dates back months or years. In that case, the DNB records this issue for further post processing measures by the use of future tools. As for instance that could mean the use of tools with more capabilities and high-performance hardware to bypass or defeat that kind of file protection.

³⁰ <http://code.google.com/p/fits/>

³¹ <http://jhove.sourceforge.net/>

³² <http://www.nationalarchives.gov.uk/information-management/projects-and-work/droid.htm>

CHALLENGES

In general the increase and change of file formats, their implementations and the DRM techniques that they contain are some of the biggest challenges. Therefore it is necessary to keep the used analyzer tools and reading platforms up to date. Furthermore new technologies like tablet PCs and portable eBook readers with new embedded techniques to protect digital rights have to be considered.

As mentioned above, it is important to detect DRM measure as early as possible – then there is a good chance to contact the author or publisher for a DRM-free version. The more time has passed, the smaller the chance to get in contact with the rights holder. That increases the risk to have, archive and use only a DRM protected restricted version of a publication.

4.2.4 NATIONAL LIBRARY OF THE NETHERLANDS (KB)

The KB, Koninklijke Bibliotheek, is the national library of The Netherlands. It is the central archival library of its country. Its mission is to offer everyone everywhere access to everything published in and on the Netherlands. So the task of the KB is to collect, catalogue and permanently archive and provide access. However, the KB has no legal mandate to act on.

The KB has as its mission to bring people and information together. With all its technological applications the Internet has in a short period of time become the resource of choice for people looking for information. This gives rise to great challenges for libraries whose traditional task has been to provide information on paper. Libraries have come to play an important role in the digital world. The KB is a pioneer in this respect, both nationally and internationally. The KB strives to exploit the potential of the Internet optimally in order to serve its customers as adequately as possible. The KB also hosts the national ISSN-center.

The KB is planning to implement a broader package of protection measurements: not only the technical limitations inside a file (like copy protection, print protection, Adobe DRM), but also in the delivery, for example:

- access only after user identification (e.g. only library pass holders or researchers)
- IP-address protection (only viewable in the reading room of our library, or only in the Netherlands)
- access only for digital objects with publication date before 1872 or 1900
- or perhaps use of a dedicated viewer which offers the necessary protection

The KB is running a project now that is aiming to implement an infrastructure and broad package of protection to use in providing our patrons access to protected material. This material is now mostly stored in 'dark archive', but using technical limitations we intend to do more with this material.

DIGITAL RIGHTS AND LEGAL BACKGROUND

In short, the Dutch Copyright Act is largely based upon the EU Copyright Directive. Works in copyright may not be made digitally available other than via a closed network within our building. For making works available for remote access or on the internet, permission is required from the right holders beforehand.

Deposit is based on voluntary deposit or based on agreements with publishers and other content suppliers. There is a copyright law. For in-copyright material deals have to be made with organizations representing the rights holders (PictoRight etc.) and publishers. Even for out-of-copyrights material there are sometimes restrictions e.g. when the digitization vendor claims certain rights (this happens with public-private partnerships).

The absence of a legal framework was not a real problem in the printed world, as the Netherlands is a small country with not so many publishers. These publishers are also well organized in the Dutch Publishers Association. When entering the digital world this becomes more and more a problem which is also felt by the KB and its collection policy.

In short and depending on the definition of DRM used:

- Employees create texts for our websites, and scientific articles (to which external publishers sometimes apply DRM), among other things
- We also digitise physical material, but do not claim rights in the scans other than database right. No search engines
- Scans made by Google from copyright free books in our collection must be protected against substantial downloading

More in detail:

a. Digital objects send by publishers

In 1996 and 1997 the KB made experimental agreements with a number of the larger international publishers based in The Netherlands (Elsevier, Kluwer Academic Publishers, SDU Publishers) for the depositing of digital publications with Dutch imprints. In 1998 a more general agreement on digital publications was made with the Dutch Publishers Association and the International Association of STM Publishers. This led to the implementation of the e-Depot in 2002, a system where digital publications (mainly e-journals articles) were stored and permanent access was guaranteed.

In the agreement with the publishers was agreed that KB e-Depot would take care for preservation but in return KB was allowed to provide on-site access to the publications (and ILL). Users may download a limited number of them and take them home. However, we must prevent mass download of whole issues and volumes. We do check on this (counter compliance).

Format and way of delivery is part of the agreement and determined in a set of guidelines on how to deliver. In the guidelines is stated that objects must not have any DRM protection, as that will make active preservation difficult or even impossible. E-Journal articles ingested in the e-Depot are as far as we know not DRM protected. But it is sure that KB receives DRM protected material. However active checks on this turns out to be difficult. In 2010 we discovered that there were files with DRM (password protection for some functions) in our e-Depot. We notified the publisher and it provided us with the necessary passwords. So, it could be that there are DRM protected files in our e-Depot, but we are mostly unaware of this. The problem is that we don't know if we receive this material. We do request the publishers to not send us DRM protected material for preservation reasons.

e-Journal articles are mostly in PDF format, problems with these are low. But journal articles are often accompanied with so called supplemental files, and this could be anything from tiff files to RAW data. DRM protection of these files, belonging to an article, could never be checked.

All provisions and agreements with publishers and other content holders are in line with copyright legislation.

All registered KB users and deliverers accept a data protection statement.

b. Digital objects created in own digitization projects

Since 2005 the KB has been doing mass digitization projects of books, newspapers and journals. In cases where this material is not out-of-copyright, deals have been made with publishers or rights holders organizations. We do not expect to use DRM on these objects.

The KB is momentarily undertaken a project called 'Access Rights'. We are currently in the stage of developing the architecture and basic functionality. In the near future we will look into producing DRM protected material (*water marking*, copy protection, Adobe DRM etc.). Publishers and other owners (or representative organizations, e.g. PictoRight) do require this from us, e.g. in the case of eBooks.

DATA TYPES AND VOLUMES

The KB collects a multitude of collections (both born digital as digitized material). These are mainly static publications, but also few multi-media. Digitized materials are Tiff and JPEG2000, text

materials are mostly PDF in different versions and, more recently, in ePub format. We accept and archive any PDF version that is submitted to us. However we have guidelines on how to use PDF.

KB is in the process of digitizing all text collections, books, journals and newspapers.

Data types

- E-journals (national and international)
- KB digitized print media (books, journals, newspapers)
- Doctoral theses and pre-prints of Dutch universities
- eBooks
- ePapers
- CD's, executables
- Selections of Dutch websites

The image below gives an overview about the data volume currently stored and the most common collections³³.

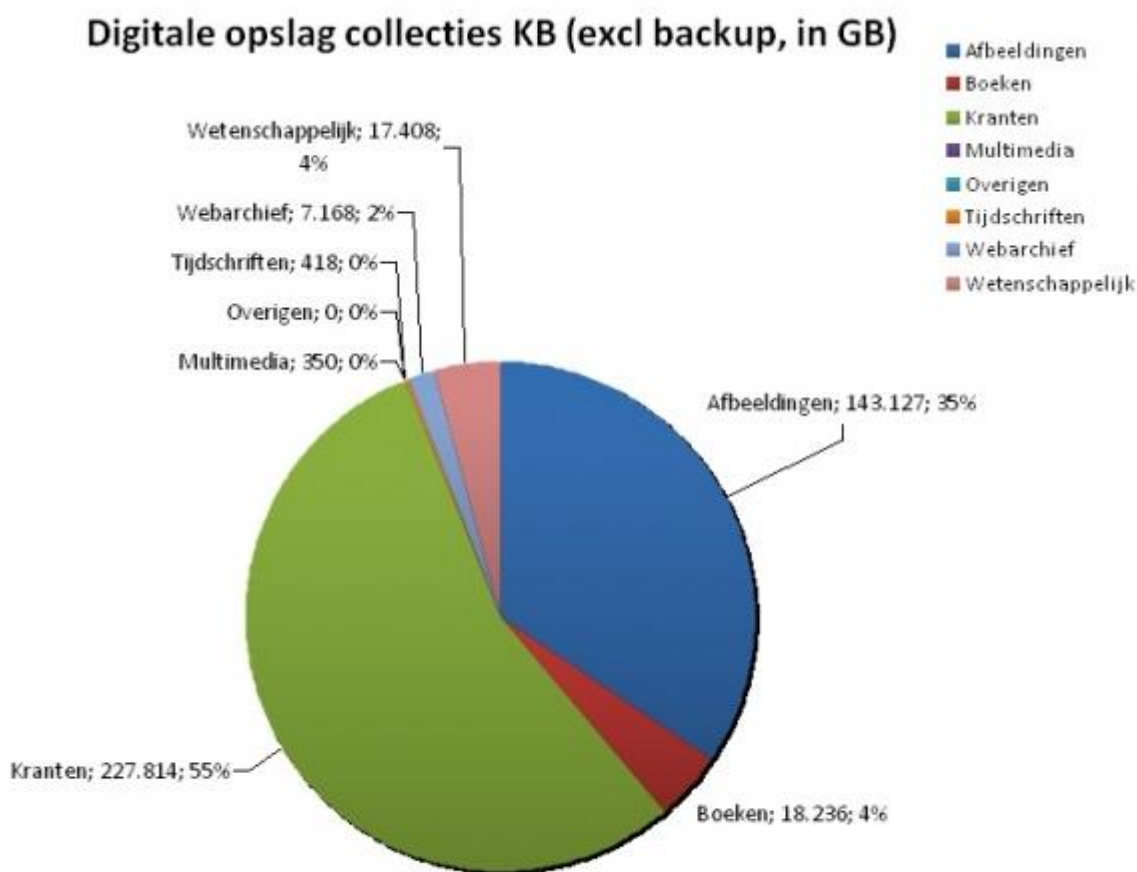


Figure 8: Storage of digital collections at the KB

³³ [Dutch captions, in English: images, books, newspapers, multimedia, others, journals, web archive, scientific \(= international e-Depot\)](#)

Part of the digital collections (eBooks, ePapers) will have DRM added by the KB. This is the only way to provide access to these collections. The percentage of digital material (born digital and digitized) will only increase in the future. So will the demand for some kind of DRM protection if we are to give access to this material.

DRM PROTECTED DATA

No figures are available, because DRM protection is not checked during ingest.

For access it will depend on the demands of the publishers and/or owners (or representatives of these owners). Data Types that are DRM protected:

- eBooks
- eJournals
- ePapers

DRM techniques and tools

During the Ingest phase, e-Journals or other types of materials are preferably accepted without DRM. But in reality, it depends on the publishers how the content is delivered.

For Access, the KB is looking into different techniques and/or tools. The library hasn't made a decision yet. It wants to introduce a service-oriented infrastructure to make sure techniques and tools can be changed over time. The demands of the publishers etc. will not be static.

At this moment the KB only uses IP address protection, only on the PC's on the library's premises can certain digital material (eBooks) be accessed. For other digital material, for example digitized material, a 'light' limitation is considered (warning that certain newspapers are considered having personal information, e.g. WWII newspapers). The protection is not yet implemented, but work is underway in a project.

THE APPROACH TO ENSURE THE LONG-TERM PRESERVATION OF DRM PROTECTED MATERIAL

- DRM is not checked during ingest, but publishers and owners are requested not to include it in their files.
- For Access, the KB is in the process of research and implementation to provide DRM (in the broadest sense of protection) to protect the material with copyright according to mainstream developments. As mentioned earlier, the library is not only looking into DRM protection of the files themselves (e.g. Adobe DRM), but will develop protection in all access services for our customers (password access, IP range access, access restriction according to certain metadata like publication date etc). XACML is used as a reference model. The KB intends to 'follow the market', even if these accepted solutions can be breached (e.g. Adobe DRM). The goal is to make the copyrighted material more accessible to our customers.

LIMITS

During Ingest, DRM can't check because of the complexity and sheer volume of the digital material, so the KB runs the risk of archiving DRM protected material in the e-Depot. If password protected and encrypted material is discovered, the publisher and/or owner are contacted to solve the problem.

Digital rights and DRM information will be stored in the metadata.

On policy level, there is a lot of contact with publishers and representatives of owners to make general agreements to comply with digital rights. A project is running to implement the technical requirements.

CHALLENGES

Concerning DRM preservation, the KB is now unaware if we have DRM protected material in our e-Depot. This can lead to problems later on.

But also the increase of formats and the implementation of format implementations (e. g. ePub, web formats) are of concern. This asks for regular updates of systems, policies and tools. It is recognized that the amount of file formats to be accepted must be limited. But that is difficult not having a legal framework. Work on detecting DRM in the ingest process needs to be undertaken.

4.3 THE DUTCH SURVEY

In 2012 the KB gave green light to a new project to bring about a solution for managing access to its digital content for its customers (access to for example eBooks, e-journals, digitized books according to legal possibilities). One of the first things undertaken in the project was to contact and/or visit other institutions that were confronted with similar challenges.

The questions focused on six main topics concerning access management:

1. Technical limitations (e.g. DRM, passwords; which technics are used?)
2. Rights metadata (how is this type of metadata stored, managed and used; is there a datamodel?)
3. What systems do you use? How are your infrastructure and processes constructed?
4. Do you let others harvest and use your metadata? Do you use technical limitations in that area or another method?
5. Is there a system present in which you store documents (e.g. all agreements) concerning your digital collections and relevant processes (e.g. diligent search, privacy complaints or copy right claims)?
6. Do you have any advice for us?

The following institutions and libraries were contacted: Nationaal Archief; Stadsarchief Amsterdam; National Archives of the UK, Beeld & Geluid, Eye Filmmuseum, Stanford Library, Rijksmuseum Amsterdam, Bibliotheek.nl, Centraal Boekhuis, Deutsche Nationalbibliothek, National Library of Denmark, Stadsarchief Antwerp.

All the answers were recorded in one document. The first pages consisted of the most relevant findings for the project:

1. **No 100% protection.** Not one single technique can offer full protection from unwanted usage. The implementation and use of any technical protection on content only shows the effort our organization is willing to make to protect content according to the rights of others. To give a Dutch example: no matter how good the lock, any bike can be stolen, even if the best and most expensive lock is used.
2. **Maintain customer friendliness.** Use a technique or method that offers protection, but don't lose sight on the usability for the customers. In general the highest protection offers the lowest usability. A balance must be found here. For example the best way to protect content is to not show it to the customers. Or a situation where content is only available on a computer on the premises of our library with no internet or USB port and a camera aimed at the computer checking the customer's actions. On the other side of the scale is the option to make content available on the internet with only the request to withhold from sharing it with others, which would make it very easy to circumvent the rights of the owners of the content (e.g. authors). It's very important to find a sustainable balance (both for the customers as for the rights holders) in making digital content available.

3. **Keep the technical design simple.** Keep the variations in type of roles, processes and rights as simple as possible. Don't give the external parties (e.g. publishers and other rights holders) and the internal parties a lot of choice, just a certain number of variations out of which they can choose one that offers the best fit. This could mean that you'll be implementing a variation that offers less than in theory might be possible. But less makes it more manageable and affordable. Start simple and slowly expand in a controlled manner.
4. **Use standard tools.** It is more efficient to use different kinds of viewers, players and readers parallel if needed in case of different filetypes, than to invest in one tool that can do it all. Because of the nature of these tools they have to be updated on a regular basis or even be replaced by a new better tool (because the old tool could be hacked). These actions should be executed fast, and off-the-shelf products have an advantage in this case compared to tailored products. Keep in mind that it is also possible to give the content to an user in combination with an agreement ('you can only use the material for these specified purposes', or when it is sensitive material, in case of the digitized Nazi newspapers, give a warning in advance).
5. **Most used technical limitations.** All or nothing; streaming; watermarking; remote access (e.g. Citrix); Adobe DRM; limited time to view the file (often shorter than in case of the paper version); single sign on mechanism; different roles have different rights; access by means of specific viewers or lower resolution (photographs).
6. **Managing rights.** Important to be able to change rights for a whole set of content, not one by one. Don't forget logging (which employee changed which value). Rights can be matched with an object or with a role or rights holder. Rights can be stored as 'hard code' or by means of rules (policies). Managing rights metadata should be made central. Keep the amount of rights metadata as low as possible to limit the maintenance burden.
7. **Metadata format.** Most organizations have developed their own formats. Official standards for rights management, e.g. ODRL are largely unknown.
8. **Rights levels.** Most organizations use a hierarchical system of rights. The lowest level is leading. For example: all newspapers till 1940 are freely accessible on the internet. Except the newspapers containing anti-Semitic content. Users have to agree to a warning before access. Or newspapers pages containing pictures of a certain photographer are only accessible on the premises (and not on the internet).
9. **Processes are crucial.** It is very important to rethink the needed process steps in organizing the necessary activities (who is responsible for which action) in maintaining well-functioning access management.
10. **Harvesting of metadata.** Some institutions have their metadatasets open to harvest to anyone. Others only allow harvesting by third parties who have a password. To get a password these third parties have to contact a Datadesk / Accessdesk. Not only does that protect the metadatasets to some point, but it also gives the institution information on who is accessing their metadata and for what purpose. In case of updates or changes it also gives the opportunities to contact these parties.
11. **Administration system (agreements).** None of the institutions interviewed on this subject have an administration system or something similar. Agreements and contracts are stored on

paper and most work processes are not described in detail. Most of them do store e.g. the agreement number in the metadata of the content.

12. **Complexity.** None of the solutions of the institutions interviewed have the same amount of complexity that the KB is needing. The KB has a large amount of different kind of collections, filetypes, and third parties (rights holders) in combination with a strong ambition to make its collections and metadatasets as available as possible to its customers.

4.4 APARSEN DRM SURVEY

A further component in this review of DRM use and preservation was an online survey carried out as part of the work package. The purpose of the survey was to find out how the participants handle DRM-protected materials and the associated rights. One of the primary targets was to discover how the community deals with archiving such objects and what is being done to protect the associated digital rights - both at present and, of course, in the future. Although the work package is focused primarily on memory institutions, other institutions such as research facilities and data centres have been deliberately included in the designated community in the hope of obtaining additional new ideas on, and approaches to, handling DRM.

The questionnaire was developed jointly within the work package. The survey was carried out using the web-based survey tool surveymonkey.com³⁴; the questionnaire was originally made available as a static Word file for an initial overview and is also appended to this deliverable.

The survey was carried out over a period of roughly five weeks from mid-September to mid-October 2013. It was coordinated with the study in APARSEN work package 35 (Data Policies and Governance), i.e. joint invitations were issued and a joint flyer (see annex 2) distributed at iPres 2013 in Lisbon. Appropriate mailing lists were used for the e-mail distribution; these are also in the annex (see annex 3).

The analysis and the results of this study are presented in the following section and, together with the user scenarios and the Dutch studies already presented, they provide the basic framework for identifying the best practices and recommendations which constitute the conclusion of this deliverable.

³⁴ www.surveymonkey.com

5 RESULTS OF THE APARSEN DRM SURVEY

5.1 SURVEY RESPONDENTS

Of the 18 respondents, approximately half have declared that they wish their responses to the survey to be treated as confidential or be anonymised. The three organisations which have not placed such restrictions on the publication of their results are: DANS (Data Archiving and Networked Services), Finnish Social Science Data Archive, Koninklijke Bibliotheek, National Library of The Netherlands, German National Library.

Other organisations did not make a declaration stating that they wished to share the information provided in the survey and have, therefore, also been treated as confidential responses.

In analysing the responses received and to understand the results of the survey in the context of the organisations participating these range from; national libraries (n=9), scientific research organisations (n=3), archives (n=2), universities (n=2), one publisher and one project response. Geographically, 16 of the respondents were from Europe (Austria, Denmark, Finland, France, Germany, Italy, Netherlands, Switzerland and UK) with two were from the USA.

5.2 THE LEGISLATIVE FRAMEWORK

Digital rights can be determined by the legal regulations by which an organisation is bound. The survey results showed that this is generally ensured through adherence to legislation in place at a national level.

Legal deposit legislation in place in various countries determines the format and delivery of digital publications. Organisations ensure that in accordance with copyright law, the related rights of legal protection are applied where required. These national regulations related to copyright law also apply to publishers where contracts are also in place to deal with rights issues. Legal deposit ensures that relevant organisations receive a copy of every publication appearing in a country including online publications. Legal mandates are also in place to harvest the web on a regular basis. Some organisations across Europe ensure compliance with copyright law based on authors' rights. The Finnish Social Science Data Archive complies with the law related to higher education and scientific research.

Legislation related to legal deposit and copyright responsibilities can also include the right to make materials available to customers. Although no specific legal regulations related to digital rights are in place in the Netherlands, DANS complies with all the relevant legal frameworks in place at the national level e.g. copyright, personal data protection. In other cases, however, limited legislative requirements apply to organisations.

Where organisations generate and fund the production of data, although this is not subject to legal controls such as copyright, there is a growing interest in the availability of data that has been produced using public funds.

5.3 THE DATA STOCK

Most organisations have dealt with the preservation of digital material since their inception. Others have done so as they have developed and in the case of archives or national libraries, they have added to their collections through purchase, donation or creation of digital objects. About half of the organisations surveyed have dealt with the preservation of digital objects in the last ten years with most of the others having done so for a longer period of time. The material held is usually historic and so retrospective preservation may be an issue for such organisations. In this instance they may inherit digital objects for which preservation actions may not have been undertaken in the way in which the organisation receiving the objects would do so now. For example, DANS which has existed since 2005, curate data from previous organisations with their oldest datasets originating from the 1960's. For projects producing digital objects, they may only be preserved from the start of the project.

Data which is archived varies enormously depending on the organisation. This ranges from research datasets (theses, reports, papers, articles etc.), to relevant publications which may be received under a

legal deposit scheme. Other types of data include electronic journals, eBooks, archives of websites, audio materials and digital surrogates.

Wide ranges of file formats are used, for example, TIFF for the scanned books, MP3 and WAV for sound and PDF and ePub for eBooks. Others are listed below:

JPEG, JPEG2000, MP4, PDF, XML, HTM, MOV, Digi-Beta, AVI, MFX, METS, TXT, hOCR, ARC, various database formats.

Some organisations like DANS and DNB are able to specify preferred formats for archiving. See report³⁵ for further details. In practice it may be that when deposited data is received by archives all file formats are accepted as long as the digital object is able to be opened and verified for its content. In some cases, only a small number of formats are allowed for long-term preservation: XML, HTML, TXT, RTF, PDF, SPSS Portable text format, SPSS Syntax text format, TIFF, JPG, and MPEG. Where scientific datasets are concerned, specialised formats are used e.g. NEXUS.

Of the 18 respondents, 61% have dealt with DRM protected material, whilst 39% do not consider DRM to be an issue for their organisation.

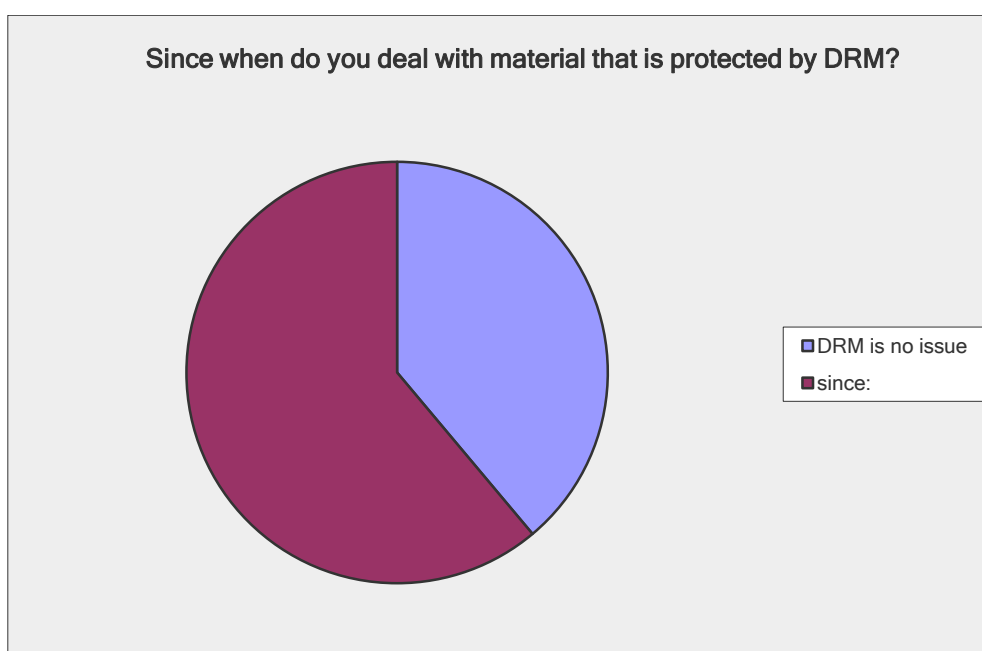


Figure 9: Since when do you deal with material that is protected by DRM? (n=18)

Of those responses (n=11) providing a date from which they have dealt with material which is protected by DRM, 83% have done so in the last ten years of which just over half have only dealt with this type of digital content in the last five years. Only two organisations providing timescales have been dealing with DRM protected material for over ten years.

5.4 THE APPROACH OF DEALING WITH DRM PROTECTED MATERIAL

Potentially all digital material, which varies enormously, could have a DRM mechanism and ideally data related to DRM should be collected automatically during relevant business processes along specific points of the lifecycle of a digital object.

³⁵ <http://www.dans.knaw.nl/sites/default/files/file/EASY/DANS%20preferred%20formats%20UK%20DEF.pdf>

As part of its mission, DANS supports the Open Access principle, while being aware of the fact that not all data can be freely available and without limitations at all times. Nevertheless the aim is that 'all datasets that are curated have a user licence, preferably an Open Access Licence'. Even so, it is important that research data that are not yet available or only available to a limited degree are archived in a sustained manner. Therefore, DANS applies the principle, 'open if possible, restricted if necessary'.

As already mentioned, various file formats are used in relation to DRM which range considerably. Some examples are: PDF or e-Pub formats for eBooks; MP3 or WAV (or similar i.e. FLAC etc.) for audio books; APK (Android Packaging Format) and IPA (iOS package format) for applications. Video games are in any number of formats depending on the platform (physically most are distributed on optical media), i.e. PlayStation, Xbox, Wii, PC, etc. Certain file formats can be used to ensure DRM, such as encrypted formats, e.g. ZIP, or Adobe Media Server to ensure that video can be streamed but not downloaded. In some cases DRM may be file format ignorant.

There are a number of approaches to preserving DRM protected material and these depend on the object as well as the legislation applicable to the object. Protected material may be preserved according to the license. A number of alternatives are possible:

- closed during an embargo period,
- requiring potential data users to justify their use,
- asking the relevant DRM provider to allow access under specific usage scenarios,
- using deposit licences in combination with conditions of use; and, fully encrypted storage in an archive.

This will depend on the conditions the owner or depositor sets or copyright and other legislation applicable.

In some cases, policies in relation to the preservation of DRM protected material are still being implemented and awaiting approval from senior management. In practice, where approaches have been implemented, DRM involves the maintenance, addition or improvement of data related to the material which will ensure provenance, rights, structural, technical and descriptive metadata.

DNB uses tools to detect DRM on digital objects during the ingest process. Previously, the detection was carried out manually by random sampling. With DRM measures in place, these may be considered as a potential risk in fulfilling an organisation's legal obligation. The DNB's approach for online publication is that material which is detected to contain DRM is refused for deposit. The publisher or the delivering institution is given the opportunity to remove the DRM protection as the organisation receiving the material does not modify the publication themselves.

Where digital objects can be received DRM free this would negate the need for the preservation of the DRM element, for example, publishers are asked to provide eBooks and audio books in this way or else they are not accepted as preservation cannot be assured. Another example is audio-visual material, where the producer who is depositing the material, for legal deposit reasons, is asked to provide a 'clean' version of the material i.e. with no DRM applied.

There is no particular approach in preserving DRM protected material collected during web archiving. This is due to the fact that a mass approach to collecting this material is applied and, therefore, no specific consideration is feasible. This is also the case for applications. For video games, DRM protection can be circumvented during media migration.

Another view, however, is that in principle, DRM material is not accepted for preservation purposes and only applies when providing customers or users with access to the material.

Of the nine respondents, 67% stated that there are both differences and similarities between online and offline material. Overall, DRM protected material is generally not directly accessible on-line as it is easier to provide material in this way where there is no DRM in place. In one example online material with DRM which can be seen in the reading rooms and is not currently predestined to enter the digital repository. Off-line material is mostly of a more sensitive nature, meaning that access has to be more restricted and controlled.

Having provided survey responses in relation to digitised material; there is an exhaustive range of other policies covering the preservation of non-digital material.

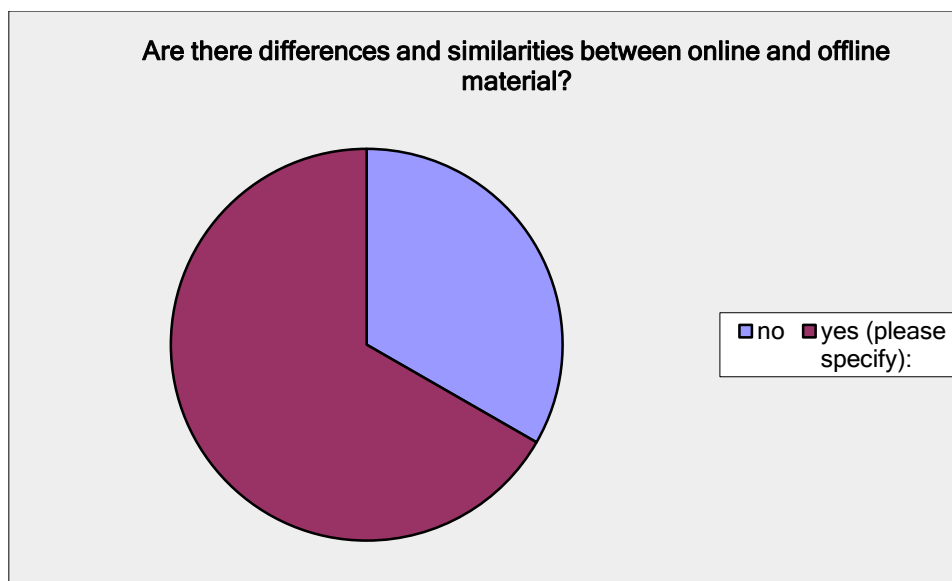


Figure 10: Differences between online and offline material (n=9)

In the survey we asked whether organisations had plans in place for analysing their existing data stock for a DRM mechanism. Of the ten respondents more than half (60%) stated that they had no plans for a DRM mechanism for existing digital objects, however 40% stated plans were in place.

Further information was provided by six respondents with three of those stating that DRM mechanisms are in place. DANS has mechanisms which analyse and adjust DRM for research datasets as a structural activity within their organisation. In another example, automatic detection as a DRM mechanism was carried out manually by random sampling. Therefore, objects already archived could have unrecognized DRM associated with them. Currently, evaluations are underway to upgrade or replace repository software. More recently, in the context of any required data migration, the migrated data passes through diagnostic tools which detect DRM in place. Two respondents are in the process of putting these mechanisms in place with one organisation stating that they were in the process of modifying and expanding their access categories as well as introducing federated access. The DNB states that this work is scheduled from 2014 onwards. In a case for video games, DRM is in place for some digital objects, which is applied case by case with no DRM in place for other material currently, as other tasks have higher priority at present.

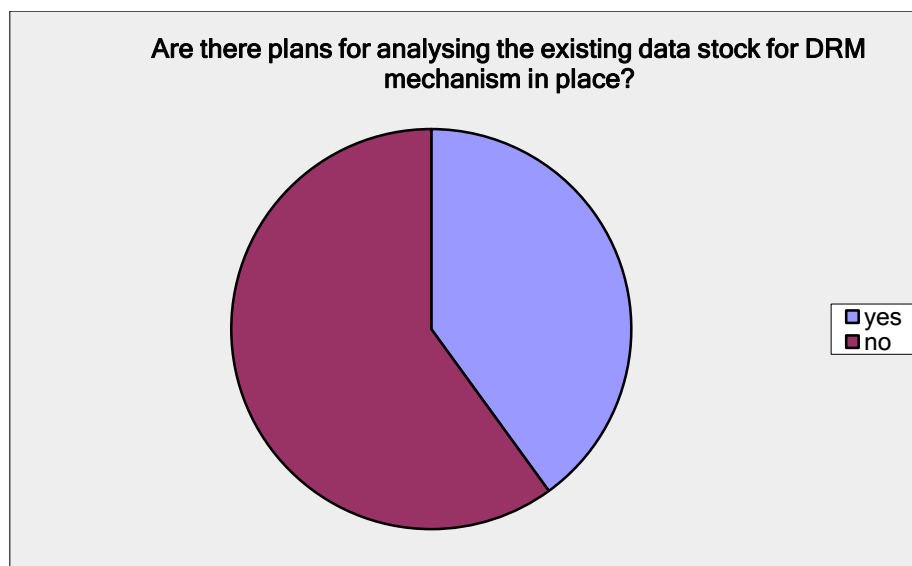


Figure 11: Plans for analyzing the existing data stock (n=10)

5.5 THE APPROACH OF DEALING WITH DIGITAL RIGHTS

Of the 13 respondents to this question 62% stated that they processed and preserved rights information at object level, with 38% stating that they did not.

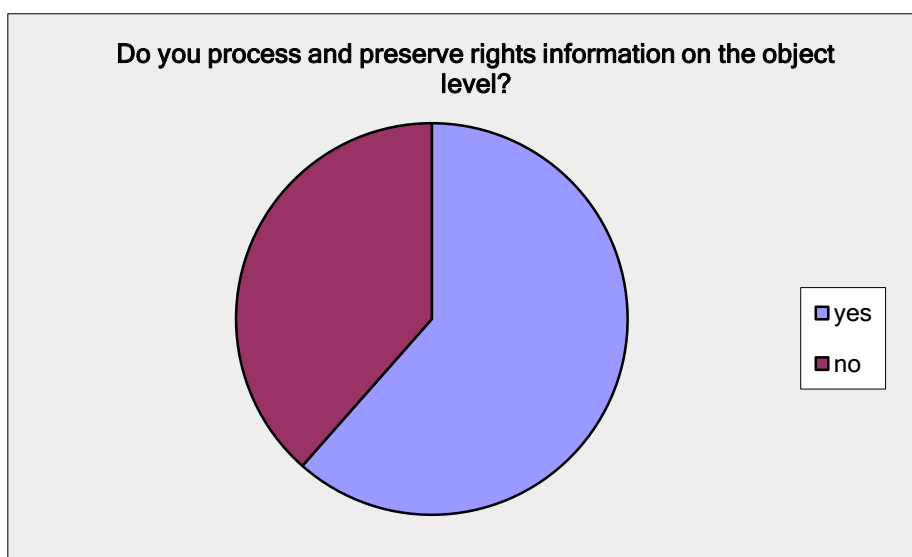


Figure 12: Rights information on the object level (n=13)

Ten responses were received when asked what organisations did to preserve digital rights and/or DRM information. Within DANS access and use is normally assigned at the level of the data type, but is also assignable for every individual object. This relates to the respective legislation that is valid at the time of access. The information is recorded in a data field at publication level and is part of the catalogue metadata.

In another organisation, rights and DRM information is part of the metadata for objects that are catalogued with another stating that this information should be prescribed in the catalogue record for each object. Similarly, digital rights (deposit licence or additions) are always kept with the data in the same archive package in the online data archive. One respondent stated that they had recorded rights information for the last 8 - 9 years, according to their business rules at collection level to restrict access at certain levels. They also have statements describing rights at object level with the aim of

having all links as persistent as possible. Articles can contain a copyright statement, or an appropriate creative commons licence and the original agreements kept and implemented. DRM related information can also be extracted and preserved in PREMIS metadata if available. However, technical DRM protection mechanisms are not preserved. In another example, an organisation kept experiment data and metadata private to the Principal Investigator and the experiment team for three years. After this time the data and metadata becomes publicly available where the data is publicly funded, with commercial users owning their data exclusively. In terms of licensing, these are an integral part of the metadata of the dataset.

When asked whether there was a metadata standard, like PREMIS or METSRights, in use to process and preserve digital rights, 13 responses were received with just over half (53%) stating that standards were in place.

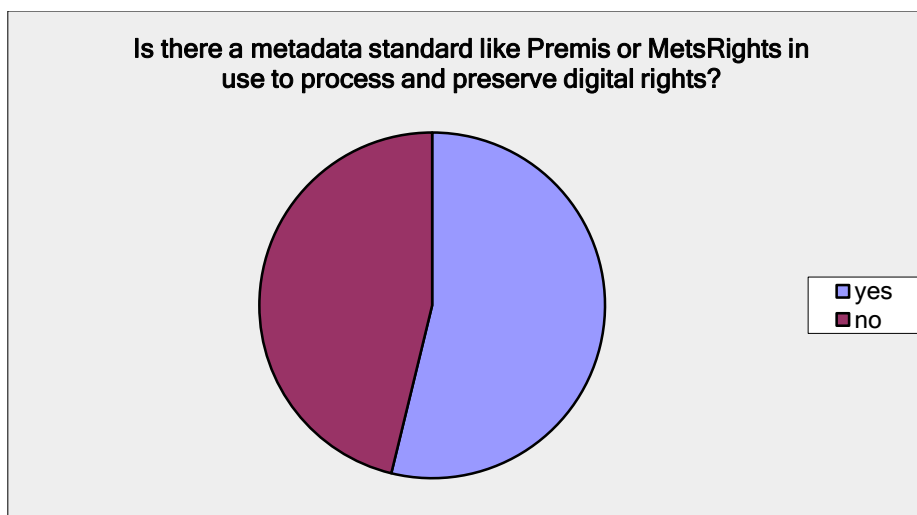


Figure 13: Use of Metadata standards (n=13)

The comments also received in response to this question from six respondents show that two organisations currently use PREMIS, and that for another a number of locally defined fields are used with the PREMIS rights part ready for use. Dublin Core and partially DDI (Data Documentation Initiative) are also used in another instance with metadata elements being used to describe research datasets which are inspired by the Dublin Core data element set.

5.6 DIGITAL RIGHTS AND ACCESS

One of the respondents, the DNB, states that the access mechanisms provided via a portal ensure compliance of digital rights of collected publications and apply to all digital holdings. Therefore, the user can only access the publication within the reading rooms of the organisation on specific types of computers. To restrict copying these 'read only' computers have the following limitations:

- no connection to the internet
- no drives for writeable medias (Floppy Disk, CD-ROM)
- no USB port
- access only via specific viewer/player software.

The access to digital holdings via the portal requires the registration and authentication of the user. Systems or processes are implemented to ensure that the agreements are adhered to. The installed viewer software allows only the printing of a publication. Thus the user cannot save the file under another location, and in particular not on an external storage medium. On every reading computer and printing station, the user gets a warning that they are also obligated to comply with copyright

legislation. For printing that means that only 15% of the pages of a publication are allowed to be printed.

In another organisation, national collections materials are as a general rule only available on-site in the reading room. Video games etc. are presently made available on the original data media and original hardware. Certain materials, for example, eBooks and audiobooks are presently unavailable but will be made available on-site on dedicated hardware that does not allow copying or data transfer. Web material can be accessed both on-site and off-site through a web interface, but only by researchers at the level of Ph.D. or higher (as per national regulations on copyright and personal data protection). All activity in the web archive is registered.

To ensure digital rights compliance on access, mostly the work is done before the decision on archiving is made: depositors can specify conditions on data use, and users must submit to the conditions set by the depositor and the archive. This requires that copyright has been cleared, there are no legal impediments to archiving and the original purpose of the data collection does not prevent archiving. Generally, data are disseminated for research and teaching only. Permissions departments also handle rights requests, and work with relevant bodies to ensure copyright management.

Providing access only on-site seems like a popular option for the organisations surveyed as access to archived material is only available in reading rooms, according to licence agreements. Access to other material is only available after login by registered users. Policies in relation to copyright, permissions and licensing are made public, however, where rights are an issue these are not disclosed publicly. Rights and credits are also displayed beside each digital object and certain file types are used to ensure access but prevent downloading with branding incorporated into the object. As in another case, software components are designed to respect access restrictions at the object level. Material with digital rights is generally reachable only in the research library and is not reproducible. For DANS, conditions of use are compulsory for all access categories. For data files with restricted access, explicit permission for access has to be given, either by DANS or the rights holder. There are sometimes additional access conditions.

80% of the software tools used for providing access to digital objects are developed in-house, with 50% purchased as commercial tools and a further 45% open source tools being used (multiple answers were accepted). In-house tools include operational databases as well as a number of command-line tools for data processing. Commercial tools include Cold Fusion, Adobe Media Server, Aquabrowser, Digitool (Ex Libris), EZProxy, IOPscience platform, SPSS (IBM predictive analytics software), Office Suite and EBSCO (research database). Open source tools include Linked Open Data access for datasets that have an open access licence as well as Acrobat reader (for PDF), Wayback machine, Bookviewer and Audioplayer.

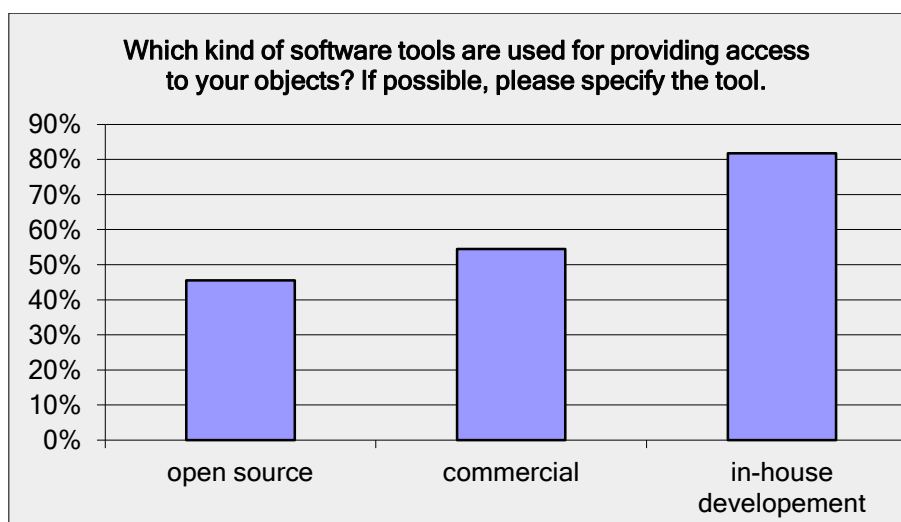


Figure 14: Kinds of software tools used (n=11)

Rights metadata is updated when rights situations change, for example, licences are monitored and when an embargo period ends the licence has to be updated from 'closed' to 'open'. In terms of the catalogue system of one of the organisations, rights metadata is managed and updated accordingly. Other organisations have rights described for each dataset in a database in which the rights metadata is edited and updated with information pushed to all relevant access points (website mainly). In other cases updating the catalogue records automatically proceeds to modifications for access. For changes to rights situations DRM metadata files are updated and the metadata feed publishes a change notice for the digital object identifier (DOI). Other updates are carried out manually by data archivists in the online data archive. One respondent stated that a changes to rights metadata rarely happened as it would involve updating agreements.

Three categories for access were identified as open access (also remotely), limited access and no access to the public. Multiple categories were selected by respondents. Access varies across these categories with all respondents stating that they provided limited access to their digital records. Open access (also remotely) was provided by 85% of organisations although 77% stated that no access was provided for the public. This would tend to suggest that given the diverse range of digital material held by the respondents access would be dependent upon the type of material held, hence the varying access categories selected.

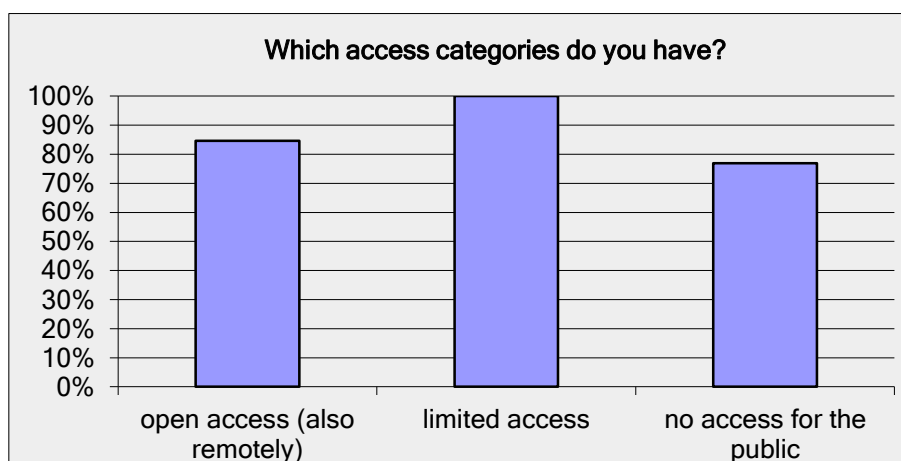


Figure 15: Access categories (n=13)

Where access is allowed options vary and multiple options are in place, with 69% stating that access is provided remotely with user authentication and 62% only providing access in-house. A further 46% state that other types of access is allowed which is diverse and varies from fully closed to automatic access with explicit permission of the rights holders. Other types of access include: open access without authentication; remotely without user authentication; open access without proper user authentication. Access to restricted objects is via application procedures where datasets and relevant documentation are then emailed and sent on CD-ROM to customers. Open access data is remotely accessible, and does not require authentication.

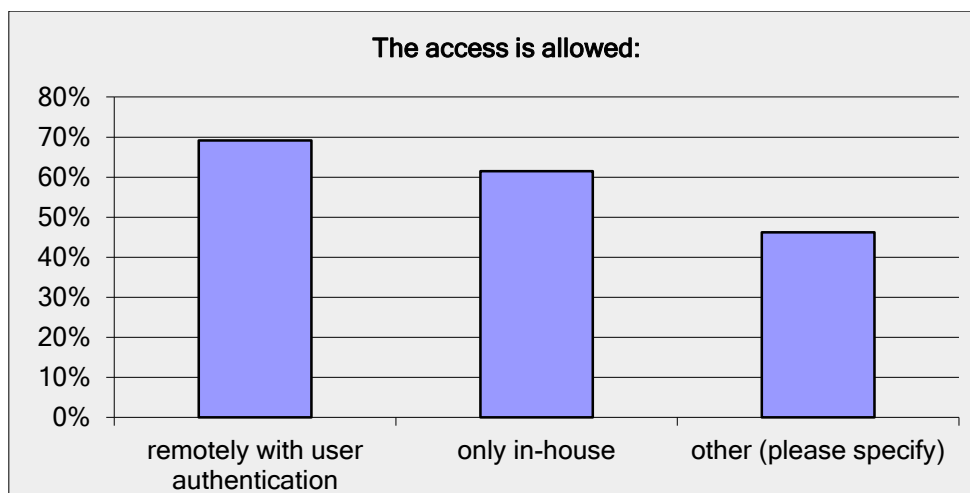


Figure 16: Access allowed (n=13)

Access configuration at object level was carried out by 82% of the 11 respondents, although one respondent stated that DRM as a technology was not really relevant as rights concerning access and use of data are enforced through working practices and systems, not through object-level technology.

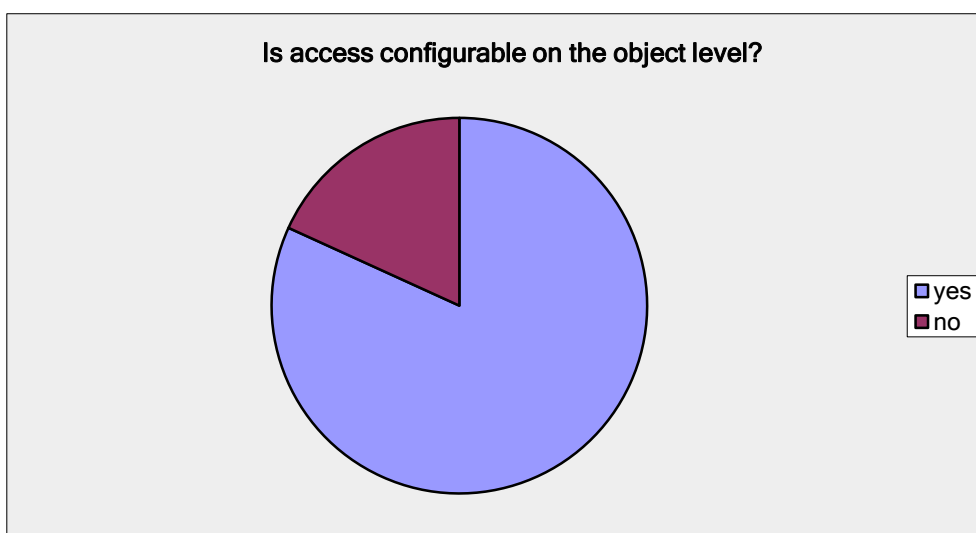


Figure 17: Configurable access (n=11)

5.7 THE APPROACH OF DEALING WITH DRM PROTECTED MATERIAL AS A DATA HOLDER

79% of the organisations (out of 18 responses) are data holders.

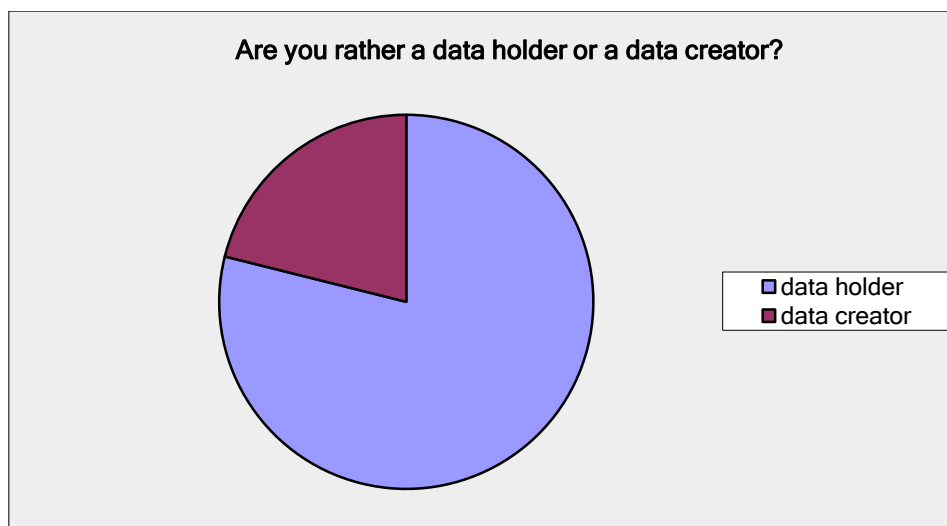


Figure 18: Data holder or data creator (n=18)

The origin of digital objects varies considerably and is based on the remit for the organisation. From within the research community across various subject areas, e.g. social science and humanities, the origins of digital objects may be individual researchers, research groups, research institutions, national and international survey data. Digital objects may originate from authors undertaking university research as well as from publishers and research projects.

The survey responses would suggest that usually no DRM technologies are used by the producers of digital objects. Where DRM protected material is provided, the material may not be accepted by the repository. On the other hand, DRM in the form of usage licences is applied and monitored in close co-operation between the depositor and the archive. In other instances, a wide range of copy protection schemes are in place. Other examples include: PDF document restrictions (password protection and print, copy restrictions), Adobe DRM (mostly from publishers) and encrypted ZIP containers. Although PDF may not be considered a DRM mechanism per se, PDFs restrict the editing of documents.

Some organisations like the DNB have specific ingest agreements with the producers of material which they collect. The arrangement could be in the form of a user licence as a mandatory part of the data transfer process. In some cases the process is formalised (formats agreed on etc.), but usually these agreements are not at a technical level. In other instances, for example, eBooks and audiobooks, these are provided for ingest by producers and distributors as DRM-free versions. More detailed agreements can stipulate that depositors deliver their material without any DRM mechanism and also fulfil the specification for packaging and transferring transfer packages and providing their material via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). A general agreement may be in place regarding storage and user access as a minimum. On the other hand, such agreements are not in place in other organisations.

60% of respondents (n=10) stated that the detection of DRM and digital rights information is a part of their ingest process.

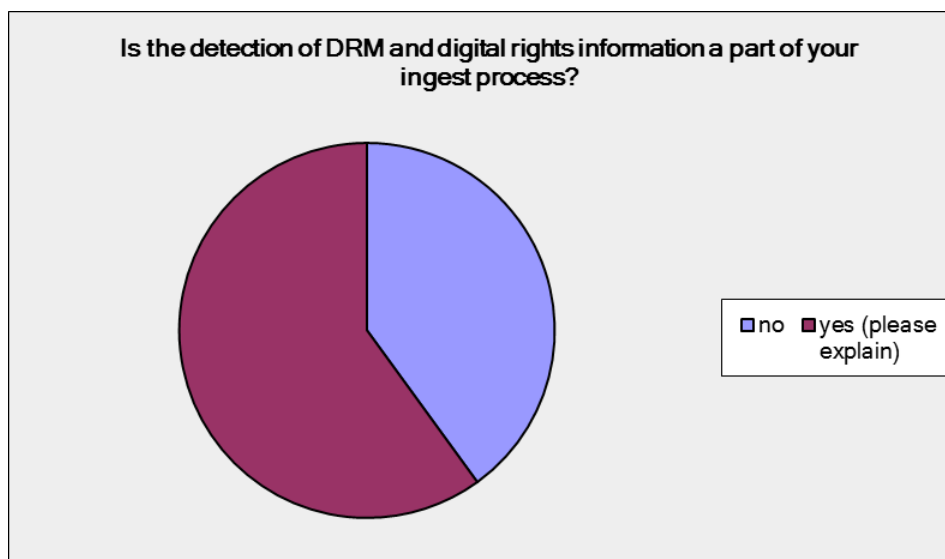


Figure 19: Detection of DRM and rights information (n=10)

As already mentioned, the detection of DRM and digital rights information may be a part of an organisations ingest process and may be included within any agreements (e.g. deposit agreements) with depositors of digit material. Automatic detection of DRM and digital rights can be undertaken using open-source tools. In the case of encrypted ZIP containers, there are regular unpacking routine reports on whether DRM protection is in place. For some time now the automatic generation of technical metadata using metadata tools has been a recognised and established component of the ingest process. The DNB has long been using the File Information Tool Set (FITS) as a framework for using an entire tool set. This framework provides access to a whole range of tools including the JSTOR/Harvard Object Validation Environment (JHOVE) tool, the Digital Record Object Identification (DROID) tool and the NLNZ Metadata Extractor. Use of a tool set widens file format support and reduces the risk of errors in the identification and validation of the file format. Some of the above tools (e.g. JHOVE) also permit the recognition of document restrictions such as password-protected PDF files. In another example, deposit agreements are developed to include DRM information; however, some organisations do not accept DRM controlled material.

60% of respondents (n=10) do not accept DRM protected material. Although some material cannot be provided without DRM (i.e. console games, apps, etc.), other material is ingested in bulk with no means of assessing the DRM status (i.e. bulk web harvesting).

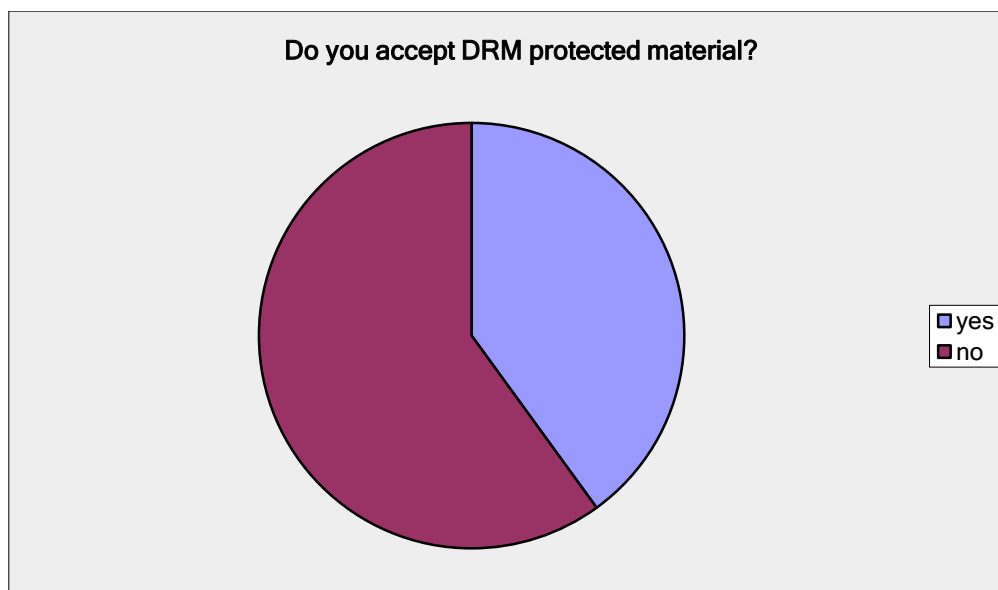


Figure 20: Do you accept DRM protected material? (n=10)

When looking at specific access agreements made with producers of material half of respondents (n=9) stated that these types of agreements were in place. For example, one respondent stated that as per their national regulations all material should be made available on site. In the case of eBooks and audiobooks agreements with suppliers access is not available until it is possible to do so using dedicated hardware that does not allow copying or data transfer. Access agreements would be particularly relevant in the case of sensitive material or could provide access only on-site for personal use in the case of preserved articles. One organisation which strives for open access can provide other types of access as required, for example; closed, open after embargo period, open only after permission, etc. In terms of material deposited, which is bound by legal mandate, depositors must confirm that they are entitled to deliver the publication as a deposit copy with the rights necessary for the legal mandate. Beyond the exceptions for use of copyright legislation, the rights holder may grant the following rights: access for registered users via internet or unlimited access. Agreements related to access can also be part of a deposit agreement or part of the licence. In another case, although depositors can specify conditions on data use or set their data in embargo for a limited time, specific agreements on access are not in place.

5.8 THE APPROACH OF DEALING WITH DRM PROTECTED MATERIAL AS A DATA CREATOR

Generally, when assessing the kind of DRM mechanism used to protect the material of an organisation as a data creator, the organisation is in fact also a data holder and so the same DRM mechanism as previously mentioned would apply to the material held as well as created. The mechanisms include a simple embargo where data is released after a defined timescale. Another example is that metadata is used to assess the mechanism applicable whereby services are built around the DRM to be applied. A publishing tool is used that automatically acts on the metadata and streaming allows viewing only but not downloading. Web pages are also designed using the underlying DRM metadata. Some controls are at item level which involves human interaction in order to determine the requirements for certain items and allows for the relevant actions as required.

5.9 SUMMARY

Overall, due to the survey response rate being low, a low confidence is placed on any analysis or summation of the results of the survey. The low response rate may also justify the conclusion that this is a relatively new area of research which requires further attention in the future. Therefore, as a summary, some key points have been highlighted below:

- 18 respondents overall, of which half were from national libraries, others were from scientific research organizations, archives, universities, one publisher and one project
- DRM issues were relevant for two-thirds of the respondents, in terms of:
 - rights of legal protection of digital content
 - copyright legislation based on authors' rights
 - right to make materials available to customers
- Implementation of DRM techniques is difficult due to a wide spectrum of different file formats and access issues related to the different number of services enabling the availability of digital content
- ideally data related to DRM should be collected automatically during relevant business processes e.g. detect DRM on digital objects during the ingest process
- Interestingly, 60% stated that they had no plans for a DRM mechanism for their existing (already archived) digital objects

6 RECOMMENDATIONS

The recommendations on the handling of DRM protected material in this section are based on observations and experiences (best practices) of the Work Package participants and on the results of the studies that were presented in this paper. The observations and experiences were comprehensively presented in section 4.2 (User Scenarios). This chapter ends with a synoptic list that can be understood as a package of recommended measures.

Restrictively, it needs to be added that there are only few truly reliable practical experiences beyond prototypical experiments with the execution of preservation actions on DRM protected materials. Because DRM - as part of the content - emerged on the market only a couple of years ago, there was little need to migrate or emulate this content to prevent it from obsolescence. However, the authors of this paper are convinced that the consideration of the compiled recommendations will facilitate the long term preservation of DRM protected materials and the protection of associated rights.

The compiled recommendations are most of all of prophylactic in nature. Under “prophylactic measures”, we will in the following understand measures that are taken before the actual archiving process, during or at least shortly after the ingest process. The goal of these measures is to recognize potential threats for the execution of future preservation actions early and, if possible, to remove them with current means.

6.1 GENERAL RECOMMENDATIONS

a) Keep the technical design simple

Keep the variations in type of roles, processes and rights as simple as possible. Don't give external parties (e.g. publishers and other rights holders) and internal parties a lot of choices. Give a limited number of variations out of which they can choose one that offers the best fit. This could mean that you will be implementing a variation that offers less than that which in theory might be possible. But less makes it more manageable and affordable. Start simple and slowly expand in a controlled manner (see section 4.3).

The system should be fully scalable and flexible. This could be achieved through standardisation of processes, with all DRM components linking to common data held in centralised repositories and machine-readable databases. This automated system should be balanced against business processes (see section 4.2.2).

b) Use standard tools

It is more efficient to use different kinds of viewers, players and readers in parallel if needed in case of different file types, than to invest in one tool that can do it all. Because of the nature of these tools they have to be updated on a regular basis or even be replaced by a new better tool (because the old tool could be hacked). These actions should be executed fast, and off-the-shelf products have an advantage in this case compared to tailored products (see section 4.3).

Standard tools are mostly loosely-coupled. That enables the replacement of components without needing to redesign the others (see section 4.2.2).

c) DRM and Rights Policy

The study that was presented in section 4.4 showcased as a best practice the definition of an institutional DRM and Rights Policy. The policy defines how DRM protected materials and their associated rights are treated. The policy should also define which DRM variants or restrictions are accepted or not. Already the process of discussing and defining such a policy creates awareness on all levels and introduces transparency. When published, the Rights Policy establishes confidence for publishers and content creators (rights holders) and can sensitize users to respect the rights of the digital objects that they use.

The Rights Policy should also contain rules for changing and adding usage rights for the purpose of auditing (see section 4.2.2). Usage rights definitions should be simplified and streamlined.

d) Collaboration between rights holders and archives

The DRM and Rights Policy mentioned as measure c) could be negotiated with a publisher or another content creator before they submit their content, if the resources of the preserving institution allow for individual arrangements and the benefits justify the effort. This could, for example, be the case for big publishing houses. If the preserving institution can guarantee appropriate DRM on the objects in their archive, then rights holders will be much more inclined to deposit the digital objects free of DRM. A good example for that is the agreement on digital publications between the KB, the Dutch Publishers Association and the International Association of STM Publishers (see section 4.2.4).

In addition to measure c) it could be also helpful to create awareness of the risks of DRM by training the content creators and publishers. This could be done by individual discussions, group seminars, webinars or presentations at relevant conferences or book fairs – perhaps also by referring to this report.

6.2 RECOMMENDATIONS FOR THE HANDLING OF DRM PROTECTED OBJECTS

e) DRM detection

As a basis for any further treatment, the detection of DRM mechanisms in archival objects is required. Such mechanisms can be detected with manual checks, either of each single object or as sample checks as part of the quality protection. The person responsible can, for example, check if the object can be deployed with the respective viewers / players. Potential access or usage restriction can thereby easily be found. If sample checks are conducted, it must be recognized that a certain amount of DRM protected objects will be ingested. When large volumes are ingested, it is preferable to use automated mass processing applications, i.e., software tools, for these checks. The Open Source *File Information Tool Set (FITS)*³⁶ deploys a range of recognized analysis tools like JHOVE. These tools provide, at least for common formats like PDF and Microsoft Word, an initial indication if DRM is used. The results of these tools can be used for risk assessment, for example by defining a LTPR or an Ingest Level (Schmitt & Hein, 2013).

f) Format Policy

If a list of preferred data file formats for long term preservation are specified in a format policy and this list is communicated to the data providers, the probability of having to deal with a variety of DRM mechanisms is reduced. The reduction of the file format variety is therefore beneficial for DRM handling and rights preservation because fewer tools need to be selected and handled.

g) Measures when DRM is detected

The detection of DRM is only sensible if a pre-defined measure or at least any kind of reaction follows suit. On the basis of the LTPR concepts presented in section 3, the following measures are conceivable:

LTPR = no risk

The data object does not contain any DRM, or, respectively, the contained DRM mechanisms like watermarks do not harm the execution of long term preservation actions. Consequently, the object can be ingested into the long-term archive.

LTPR = medium

A data object with an associated DRM mechanism should not be archived without further analysis. It is recommendable to request a version of the object from the data provider that is free of DRM. If this should not be possible, the conversion into a format or a data carrier that is free of DRM can be considered. If the legal circumstances allow for it, the “digital-to-analogue conversion” could be an option, even if a lossy one. If such measures towards normalization are of greater complexity and require a more thorough preparation, it is recommendable to archive the object, but to record the kind of DRM mechanisms as part of the technical metadata (see measure h). Because it is possible to use objects with a medium LTPR with current hardware and software, the objects should be normalized as soon as possible after ingest.

LTPR = high

³⁶ <http://code.google.com/p/fits/>

Because objects of this category can even currently only be used with restrictions and will certainly result in restrictions during normalization of preservation actions, these objects should not be archived and DRM free versions should be requested from the content providers.

The Ingest Level Concept that is in use at the German National Library, for example, leads to rejection of all objects with any kind of DRM. It is, however, not always an option to reject DRM protected objects, respectively, to request DRM free versions, especially when the producer cannot be identified anymore. Furthermore, not every content provider is immediately willing to provide its objects without DRM to the preservation institution.

In these cases, it can only be attempted to create awareness for the problem on the side of the producer / content provider. If there is a legal mandate, the preservation institution can use it as an argument. Also the guarantee that the rights will be protected via an institutional access management, so that no disadvantages result from DRM free objects for the content provider, can assist the argumentation. It will, however, imply additional effort for the preservation institution elsewhere, namely in the implementation of such an access management (see section 4.2.2).

If the request for DRM free versions turns out unsuccessful, the measures h) and i) remain.

h) Documentation and Archiving of DRM

If there is no alternative to archiving the object with DRM protection, it is recommended to document it as detailed as possible in the Data Management Functional Entity (see OAIS) at the data level. “As detailed as possible” means to provide all possible details concerning the DRM mechanism used, for example, the kind of usage restrictions.

The documentation in Data Management puts the preservation institution in the position to conduct certain measures later, for example, a normalization or later DRM detection as described in section d) of this chapter. Moreover, the capturing of DRM information in a database enables the creation of a comprehensive statistical basis that allows for reliable statements about the quality of the data holdings and for estimations about the portion of protected objects.

At this point, it would also be conceivable to renounce any further DRM specific measures and to limit the attention to bit stream preservation and to the protection of the DRM mechanism. However, from the point of view of the authors of this paper, this is not recommended. Especially on the example of DRM systems, it becomes obvious that the reproduction or emulation of all external dependencies, in particular of the individual backend components of a DRM system, will hardly be possible. Even the option “password protection with encryption” involved the danger that the password is lost sometime in the future. The password needs to be archived and kept accessible together with the preserved content. In the case of copy protection it needs to be taken into account that current hardware that can deal with the protection measures will most likely not be available in the future. So it is highly doubtful if, for example, a copy protected audio CD will be readable in a future device at all, independently of the robustness of the data carrier itself.

i) DRM removal

If the legislation allows it for memory institutions, the removal or bypassing of DRM protective measures during the ingest process could be a feasible step, e.g., as part of the normalization of archival content.

There are, however, a couple of critical points that need attention:

- The technical realization of this strategy needs a thorough examination of each of the data file type dependent DRM protection measures in order to identify or develop suitable tools. Therewith, it is a relatively laborious strategy. Sometimes it needs to be checked if these tools can, under consideration of national or European legislation, be legally acquired and used.
- Moreover, it needs to be clarified if the removal of DRM protection measures constitutes a migration (especially in terms of normalization). In particular, the question rises whether this touches upon the authenticity of the object. The quality checks need to ensure for every scenario that all significant properties are unchanged after the removal of DRM protection.

- The manipulation of content makes checksums unusable. This is critical if these checksums were meant to be used for the assessment of the data's integrity, especially if the author is no longer reachable to confirm the authenticity of the content.
- It needs to be taken into account that the removal of a password encryption is possible only to a limited extent. If the password encryption is robust and the length of the password is sufficient, it is almost impossible to crack a password with a *brute force* attack in justifiable time.
- The tools utilized for DRM removal need maintenance and support and, potentially, additions.
- The removal of DRM protection mechanisms can be CPU intensive and time consuming. Thereby, it influences directly the complete processing time of an object.
- The removal of password encryption does not necessarily create an object that is free of DRM. It is only the first step: If, for example, a PDF document can be accessed after password removal, it needs to be converted into a version that is free of DRM.

Even if a series of arguments seems to speak against the suggested approach, the APARSEN study has shown that the removal of DRM during migration is already applied for example for video games.

j) Analyzing the existing data stock

One of the findings of the APARSEN DRM survey is that 60% of the respondents have no concrete plan to analyze their already archived objects for risks that could come up with DRM mechanisms in the future. Unknown or undocumented DRM protection could be a problem which is not solvable. In the worst case, access to the object is forbidden by DRM protection and no one knows how the mechanism works or what requirements are necessary to gain access or to provide specific usage functionalities. Therefore it is advisable to analyze already archived objects or at least their generated technical metadata to detect any restrictions in time. The analysis and the following steps could be done for example by the help of the presented measures in this section.

6.3 RECOMMENDATIONS FOR THE PROTECTION OF DIGITAL RIGHTS

DRM mechanisms are known to have the purpose to protect the rights of the creator of a work. Consequently, the preservation institution assumes a special commitment when it requests DRM free objects of when it removes DRM from archived objects.

k) Detection of Rights Information

The associated rights are by no means always clear or documented with the archival object. In order that digital rights can be protected, they must in cases of doubt need to be detected and documented. As a first approach, preservation institutions can of course contact the content provider. Beyond that, tools like the *Public Domain Calculator* presented in section 4.1 can help to identify rights.

l) Documentation and Application of Rights

If the rights are known, it is necessary to document them appropriately. Here, recognized standards like the Rights Expression Language (REL) presented in section 2 should be used.

Independently of the selected implementation, it needs to be ensured that the access of archived objects is always organized according to applicable law. If the information needed to ensure this in the access systems originates from Data Management or any other system like the library's catalogue or an own Rights Management solution (see section 4.2.2), does not matter. One result of the study (see section 4.4) shows that the majority of respondents on our survey manage such rights information on object level. As it is shown below, it is sensible to manage some rights information additionally at collections level in order to minimize maintenance efforts.

The preservation of rights information needs to meet the same standards as the preservation of the archival content itself. That means that if rights information is stored in data bases, future access needs to be ensured, and the data model needs to be interpretable and usable in the future, too.

The system for the management and preservation of rights information needs to account for changes in the rights information (see section 2.8). If the rights owner, for example, withdraws some rights that were previously granted, the rights information needs to be updated accordingly. The update needs to be respected, of course, by the access function. From section 4.3: It is further important to be able to manage and change rights for a whole set of content, not only one by one. The logging of changes or audit trail should not be forgotten (which employee changed which value). Managing rights metadata should be made central for example by using a Rights Management System. Keep the amount of rights metadata as low as possible to limit the maintenance burden. The high-level principles described in the user scenario of the British Library require that the Data should be live, reliable and reusable (see section 4.2.2).

The complete and correct documentation of Rights (i.e. by the usage of a rights matrix) is the requirement for their application and safeguarding. BL's ideal Rights Management solution has shown that its access control component is thus able to evaluate, for any particular request, whether content should be delivered to the user. Delivery of content is according to the role of the user, location and the usage right being applied to the content, as well as the time at which the request is submitted (see section 4.2.2).

m) Information about digital Rights

The users are not always aware of the opportunity of rights infringements when using digital materials. Beyond the suggested DRM and Rights Policies, it can be helpful to display some information about copyright and the limits of fair use immediately before the user access the requested archival information. This will help to raise the user's awareness concerning digital rights.

n) Storage in the Archival Package

In addition to option h) (Documentation and Archiving of DRM), it is conceivable to store rights information within the Archival Package (see section 5.5). In Open Source Software products, for example, it is common use already to include usage licences like the GPL³⁷ as a text file into the software package. It is, however, recommendable not to use this information for evaluation during access and use, but to apply a procedure like it is described in g). The reasons for this become clear quickly, given that the package needs to be downloaded and unpacked each single time before the rights information become visible. To regulate access, however, the rights information is needed before the access package is submitted to the user. If rights information is stored and managed in a database system, it can be controlled and maintained more easily and efficiently and it can also be integrated into access functions more easily. Consequently, the inclusion into the Archival Package is sensible for the case that the system used in scenario h) is damaged or destroyed. Ideally, the rights information that is stored in the Archival Package can then be transferred automatically to the new or repaired system, using standardized RELs like PREMIS or METSRights.

³⁷ General Public Licence: <http://www.gnu.org/licenses/gpl.html>

6.4 SUMMARY

Id	Name	Question and Purpose
a	Keep the technical design simple	Less variations make rights, process and roles more manageable and affordable
b	Use standard tools	One tool for all formats and document types is hard to maintain and risky because of its complexity; in contrast: the usage of standard tools for a specific format is more robust against error and has only impact on the specific file format
c	DRM and Rights Policy	Transparency in dealing with DRM and digital rights
d	Collaboration between rights holders and archives	Make the content creators and publishers aware of the risk that DRM carries related to long-term preservation so that they are more inclined to deposit digital objects free of DRM.
e	DRM detection	Is an object protected with DRM measures? Precondition for further measures.
f	Format Policy	The specification of preferred file formats facilitates the handling of utilized tools.
g	Measures when DRM is detected	Adequate measures like further examinations, request of DRM free versions or rejection, dependent of the employed DRM mechanism
h	Documentation and Archiving of DRM	Archiving of DRM protected objects and documentation of utilized measures in Data Management. The purpose is to facilitate activities that are targeted towards DRM protection measures like normalization into a DRM free format or copying onto a generic data carrier.
i	DRM Removal	If the legal conditions allow for it, the removal of DRM protection measures in order to derive an object without DRM.
j	Analysing the existing data stock	Detect potential risks related to DRM in time. The longer the risk remains undetected, the greater the risk that the content is lost.
k	Detection of Rights	Which rights are associated with the archival content? Are the rights information supplied by the content provider or do they have to be determined?

l	Documentation and Application of Rights	The structured storage, management and preservation of rights information in a Rights Management System that is connected to the access function. Access and use are managed in accordance with the rights information in the Rights Management System.
m	Information about digital Rights	Raise awareness of the user about the handling of digital materials.
n	Storage in the Archival Package	The additional storage of rights information with help of metadata standards like REL in the Archival Package. It allows the re-creation of rights information in case the Data Management or another Rights Management Systems is damaged or destroyed.

6.5 POTENTIAL RESEARCH QUESTIONS

A **comprehensive ‘DRM solution’** includes management of digital IP policies and application of usage rights, business processes and technical components (both software and hardware) both to control access, track and maintain usage rights. Further research into the management of policies, business processes and technical solutions would provide a unified and comprehensive approach to DRM requirements faced by a number of organisations. In addition this approach should start the development of “open” **DRM systems/solutions for public institutions** according to the common requirements as stated by the British Library:

- Automated instead of manual processes, as far as possible
- DRM capability to support **copyright and IP policy** as well as access and re-use policies
- **Licences and contracts** need to be automated with machine-readable rights metadata by focussing on the terms of use

In this context it should not be forgotten that there is currently no industry-wide standard for DRM. This would, of course, be highly desirable especially with regard to compatibility, interoperability and long-term preservation.

Another research questions is whether restrictions on the frequency of use of an object, due to the application of digital rights, would equate **to restricting the implementation of LTP measures**. The question remains unanswered whether e.g. analysis tools for preparation or post-processing (e.g. quality assurance) constitute an incidence of use and therefore reduce the number of uses of the object. The proportion of DRM protected material will be decreasing as **data carriers** such as CD-ROMs and DVDs will decrease in volume. This does not mean, however, that the problem is solved. There is no evidence available which portions of DRM protected data carriers reside within library repositories these days. The decay of the data carriers poses a threat to long term preservation, and the longer the data carriers sit in the shelves, the more urgent it becomes to migrate the content from the carrier to mass digital storage systems. In order to safeguard all these materials for future use, firstly, a statistical analysis of the problem needs to be made, and secondly, a mechanism for data carrier migration albeit the DRM protection needs to be developed.

Furthermore new technologies like tablet PCs and portable eBook readers with new embedded techniques to protect digital rights make research in that field an **ongoing task**. This task would complement the evaluation of the developed DRM risk evaluation scale and also further the development and maintenance of the tools used for detecting DRM mechanisms.

7 CONCLUSION

One of the primary goals of this work package was to examine approaches for dealing with digital rights and protection mechanisms, which are known as DRM or DRM systems, in memory organizations and other related institutions. After a comprehensive orientation into the terminology of this subject, the problem of how to represent and administer digital rights and eventually of how to archive them was examined. Additionally it was investigated which DRM techniques were widely used and how the risk of archiving was assessed on protected objects.

Analysis of the study in section 5 revealed that the topic DRM had relevancy for approximately two-thirds of the polled participants. The study likewise showed that the problem definition itself is very complex and difficult in terms of how to implement DRM techniques due to a wide spectrum of different file formats. An attempt to make the complexity of this report manageable was undertaken in a classification of the DRM techniques in section 3, regardless of which file format.

Section 4 of this report takes a look back. A compilation was made of those initiatives and projects which have already grappled with this topic. Furthermore there was a description of the current concrete use scenarios amongst the work package participants. The application of best-practices, in addition to the results from the analysis of the two studies (Dutch survey, APARSEN survey), found its way into the third section of this report, that contains a catalogue of recommendations in dealing with DRM and digital rights.

Even when an overwhelming majority of input from the responses came from the area of the data holder, it should be clear that the statements made here, in particular the recommendations could be applied to data creators, and also across sectors.

Despite, for instance, content providers like Apple³⁸ having meanwhile sounded the alarm for the lack of interest in DRM in at least the area of music, it should be taken into account that dealing with DRM and the preservation of digital rights is a continuous responsibility and at the same time it remains a challenge. However, DRM is still in a decidedly growth period especially in the area of digital publications and the film industry (i.e. online video stores).

From a memory institution's point of view, a distinction must be made between the digital rights and the DRM techniques. For memory institutions, safeguarding the protective rights of their archived assets is essential, and therefore they either fall back to already existing mechanisms, for example their retrieval systems, or, as was demonstrated in the BL user scenario, it will be set on their own internal rights management system. This approach requires that the digital archive is a durable trusted archive and that the owner of the objects trusts the repository. At this point the financing of the model can be problematic, as the repository has to finance the archive infrastructure, but does not have the authority to provide access. In this situation public funds have to pay the archive infrastructure and the (commercial) publishers can exploit their assets without worrying about the durability of the assets. Even if the latter prefers a type of 'all-in-one' solution in the end, the demand will always be present to be able to process, manage, and archive rights and rights information properly within a system. For such solution to be an "open" DRM system or solution for public institutions it is important to use open standardized components and open metadata standards presented in section 2. Also essential is the investing in training and qualification, because only capable and competent personnel are able to operate a DRM system accordingly and take care about the preservation of the content and the associated digital rights.

Through the integration of proprietary rights control mechanisms as an integral component of digital objects, a new problem has arisen regarding long-term archiving. The main cause of this problem has been that access and restrictions of use could hinder the preservation of the object. From the long-term archiving perspective, it is rather seen as a problem of the long-term preservation of rights information and even of the capability to safeguard the represented rights in the future. The authors of this report hope that their proposed catalogue of recommendations in chapter 6 has provided type of 'first aid' support for this problem to all affected institutions.

³⁸ <http://www.apple.com/pr/library/2009/01/06Changes-Coming-to-the-iTunes-Store.html>

REFERENCES

- About European Civil Rights in Europe*. (2010, 12 22). Retrieved 10 02, 2013, from European Digital Rights: <http://www.edri.org/about>
- Angelopoulos, C., & Jasserand, C. (2011, 10). *Public Domain Calculato: Report and Documentation*. Retrieved 10 2013, from Europeana Connect: <http://outofcopyright.eu/research/Public%20Domain%20Calculator%20-%20Report%20and%20Documentation.pdf>
- CEPS DIGITAL FORUM. (2013, 06). *Copyright in the EU Digital Single Market*. Retrieved 10 2013, from http://aei.pitt.edu/42912/1/Copyright_in_the_EU_Digital_Single_Market_FINAL_e-version.pdf
- Commission, E. (2012, 12 18). *Communication from the Commission on content in the Digital Single Market*. Retrieved 11 25, 2013, from http://ec.europa.eu/internal_market/copyright/index_en.htm
- Consultative Committee for Space Data Systems. (January 2002). *Reference Model for an Open Archival Information System (OAIS)*. Abgerufen am 16. 01 2014 von <http://www.ccsds.org/documents/650x0b1.pdf>
- Corporation, M. (2004). *Architecture of Windows Media Rights Manager*. Retrieved 11 25, 2013, from <http://www.microsoft.com/windows/windowsmedia/howto/articles/drmarchitecture.aspx>
- Creative commons. (n.d.). *About*. Retrieved 10 2013, from <http://creativecommons.org/>
- England, Paul (Bellevue, W., DeTreville, John D. (Seattle, W., Lampson, & Butler W. (Cambridge, M. (2002). *Newsweek report on Microsoft Palladium which lists these patent owners as Palladium programmers - Digital Rights Management Operating System*. Retrieved 11 25, 2013, from <http://cryptome.org/ms-drm-os.htm>
- European Digital Rights. (2012). *European Digital Rights - Annual Report 2012*. Retrieved from edri.org: <http://www.edri.org/files/EDRI-yearly-report-12.pdf>
- Europeana Foundation. (n.d.). *Europeana Data Exchange Agreement*. Retrieved 10 2013, from europeana.eu: [pro.europeana.eu: pro.europeana.eu/documents/900548/8a403108-7050-407e-bd00-141c20082afd](http://pro.europeana.eu/documents/900548/8a403108-7050-407e-bd00-141c20082afd)
- Grimm, R., & Neubauer, C. (2004). *LWDRM - An Alternative Rights Management System*. Retrieved 11 25, 2013, from <http://waste.informatik.hu-berlin.de/Grassmuck/drm/Folien-Grimm-Neubauer-eng.pdf>
- Iannella, R. (2001). *D-Lib Magazine June 2001 Volume 7 Number 6 - Digital Rights Management (DRM) Architectures*. Retrieved 11 25, 2013, from D-Lib Magazine: <http://www.dlib.org/dlib/june01/iannella/06iannella.html>
- Iannella, R. (2002). *Open Digital Rights Language (ODRL) Version 1.1*. Retrieved 11 25, 2013, from W3C: <http://www.w3.org/TR/odrl/>
- INDICARE project. (2006, 04). *Consumer's guide to Digital Rights Management*. Retrieved 10 02, 2013, from indicare.org: <http://www.indicare.org/consumer-guide/>
- indicare.org. (2008, 09 04). *indicare.org*. Retrieved 10 03, 2013, from <http://www.indicare.org/tiki-page.php?pageName=ProjectDescription>
- Kuch, N. (2007). *Rechtmanagement multimedialer Assets*. Saarbrücken: VDM Verlag Dr. Müller e. K. und Lizenzgeber.
- Lavoie, B., & Gartner, R. (2005). *Preservation Metadata. DPC Technology watch report 2005*. Abgerufen am 16. 01 2014 von <http://www.dpconline.org/docs/reports/dpctw05-01.pdf>
- OCLC/RLG Working Group on Preservation Metadata. (June 2002). *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects*.
- Picot, A., & Thielmann, H. (2005). *Distribution und Schutz digitaler Medien durch Digital Rights Management*. Berlin Heidelberg New York: Springer.

- PREMIS Working Group. (May 2005). *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. Von <http://www.oclc.org/research/projects/pmwg/premis-final.pdf> abgerufen
- public domain calculation*. (n.d.). Retrieved 11 13, 2013, from Europeana Connect: <http://outofcopyright.eu/index.html>
- Rosenblatt, B. (2005, July 14). *Enterprise Digital Rights Managament*. Retrieved 11 25, 2013, from <http://www.giantstepsmts.com/Authentica-RMS%20Whitepaper.pdf>
- Rust, G. (2013, 04). *The LCC Rights Reference Model*. Retrieved 10 2013, from http://media.wix.com/ugd/bff7bc_739b7aa7f0d4b4b2c8e7929aa3f07868.pdf
- Schmitt, K., & Hein, S. (2013). *Risk Management for Digital Long-Term Preservation Services*. Retrieved 11 28, 2013, from IPRES 2013 : proceedings / of the 10th International Conference on Preservation of Digital Objects: http://purl.pt/24107/1/iPres2013_PDF/Risk%20Management%20for%20Digital%20Long-Term%20Preservation%20Services.pdf
- Software & Information Industry Association. (2001, 05 30). *The Digital Object Identifier: The Keystone for Digital Rights Management*. Retrieved 10 11, 2013, from SIIA Digital Rights Management Working Group: <http://www.siiia.net/estore/download/doi-01.pdf>
- Waters, D., & Garret, J. (1996). *Preserving Digital Information: Final Report of the Task Force on Archiving of Digital Information*. Von <ftp://ftp.rlg.org/pub/archtf/final-report.pdf> abgerufen
- WIPO. (n.d.). *World Intellectual Property Organisation (WIPO) - Website*. Retrieved 11 25, 2013, from <http://www.wipo.int>

ACRONYMS

AIP	Archival Information Package
ANSI	American National Standards Institute
DRM	Digital Rights Management
DoW	Description of Work
DWA	Digital Watermarking Alliance
FITS	File Information Tool Set
HTML	Hypertext Markup Language
IP	Internet Protocol
LMF	Local Media File
LTP	Long Term Preservation
LTPR	Long Term Preservation Risk
LWDRM	Light Weight DRM
MARC	MACHine-Readable Cataloging
MD	Metadata
METS	Metadata Encoding & Transmission Standard
MP3	MPEG Audio Layer 3
MPEG	Moving Picture Experts Group
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
ODRL	Open Digital Rights Language
OMA	Open Mobile Alliance
PREMIS	PREservation Metadata: Implementation Strategies
REL	Rights Expression Language
RLG	Research Libraries Group
SMF	Signed Media File
VcOE	Virutal Center of Excellence
WP	Work Package
XrML	eXtensible Rights Markup Language

ANNEX 1: QUESTIONNAIRE OF THE APARSEN WP 31 SURVEY

Digital Rights and Access Management survey (word version) for APARSEN project

Dear survey respondents,

Work package 31 “Digital Rights and Access Management” needs your help. Our goal is to find out how you deal with DRM protected digital objects and their associated digital rights. As a memory- and a research institution or a data centre we are strongly interested in what you do to preserving these kind of objects and what you undertake to guarantee digital rights now and, of course, in the future?

Please take some time to help us taking a good snapshot of the common approach of dealing with digital rights and DRM within our community.

For a common understanding of DRM the following definition is given:

Digital Rights Management (DRM) is a set of technologies that are used with the intention to control the access and use of digital content and devices [1].

DRM mechanisms can be implemented inside or outside the file that is being protected. The different mechanisms range from digital watermarking (over specific control mechanisms like the protection for viewing), copying, printing and altering that are often capabilities of the file format. To DRM, in a broad sense, this includes any access control technology outside the file. In the field of libraries, for example, the implementation of DRM can also be a part of the access system that stores information of digital rights in a separate metadata area and that is also responsible for guaranteeing these rights.

[1] http://en.wikipedia.org/wiki/Digital_rights_management

Survey questions

Basic information about the institution and the repository

1. Name of the institution:
2. Institutional Mission:
3. Legal regulation:

Keywords: legal mandate, national copyright act, digital rights in your country, specific national and international regulations to comply with

4. Contact of the questionnaire respondent:

Name

eMail

Role

Examples: ITstaff, librarian, researcher, management

The data stock

5. Since when do you deal with the preservation of digital objects?
6. What kind of data is archived in general?
7. Which file formats are used in general?
8. Since when do you deal with material that is protected by DRM? DRM is no issue or since...?

The approach of dealing with DRM protected material

9. Which kind of your data have or could have DRM mechanism?
10. Which file formats are used in regard to DRM?
11. Describe your approach of preserving DRM protected material.
12. Are there differences and similarities between online and offline material? No? If yes, please specify:
13. Are there plans for analysing the existing data stock for DRM mechanism in place?
If yes, please describe your approach, if no, please describe why not:

The approach of dealing with digital rights

14. Do you process and preserve rights information on the object level? Yes/no
15. What do you do to preserve digital rights and/or DRM information?
16. Is there a metadata standard like Premis or MetsRights in use to process and preserve digital rights? If yes, please specify:
17. What do you do to comply with digital rights - especially on the access side?
18. Do you use different viewers / players for different file formats to comply with digital rights? If yes, please specify:
19. Which kind of software tools are used for providing access to your objects? If possible, please specify the tool.
open source / commercial / in-house development
20. How do you update your rights metadata when the rights situation has changed?
21. Which access categories do you have?
open access (also remotely) / limited access / no access for the public
22. The access is allowed: remotely with user authentication / only in-house / other (please specify)
23. Is access configurable on the object level? Yes / no

Form of organization

24. Are you rather a data holder or a data creator? data holder / data creator

The approach of dealing with DRM protected material as a data holder

25. The origin of the digital objects is:
26. Which DRM technologies are used by your producers?
27. Do you have specific agreements on the ingest with the producers of your collected material?
28. Is the detection of DRM and digital rights information a part of your ingest process? no / yes (please explain)

29. Do you accept DRM protected material? yes / no / because: (please specify)

30. Do you have specific agreements related to the access with the producers of your collecting material?

The approach of dealing with DRM protected material as a data creator

31. What kind of DRM mechanism is used to protect your own material?

Declaration

32. Under what conditions would you be willing to share the information? no limitations / only anonymous / only confidential use with the partners in the APARSEN project

Thank you very much for your time and willingness to support us.

ANNEX 2: INVITATION FLYER



Alliance Permanent Access to the
Records of Science in Europe Network

APARSEN is currently conducting two separate but complementary surveys.

Your input is very much appreciated.

DIGITAL RIGHTS AND DRM:

<http://URL>

DATA POLICIES AND GOVERNANCE STRUCTURES:

<http://URL>

Please take some time to help us taking a good snapshot of the common
approach of dealing with digital rights and data policies.

Please complete both of these surveys by September 30th 2013



ANNEX 3: MAILING-LISTS FOR SURVEY DISSEMINATION

- pasig-discuss@mail.asis.org
- eudat@postit.csc.fi
- also at the EUDAT web site: <http://www.eudat.eu/news/aparsen-needs-your-help-online-surveys-30-september>
- rda-all@lists.rd-alliance.org
- sim4rdm@googlegroups.com
- Knowledge Exchange
- National Digital Library of Finland (sent to key people; reminder will be sent)
- TTA. TTA (Tutkimuksen tietoaaineistot) is a research data network in Finland
- DC-PRESERVATION@JISCMAIL.AC.UK
- DIGITAL-PRESERVATION@JISCMAIL.AC.UK
- diglib@infoserv.inist.fr
- DPC-DISCUSSION@JISCMAIL.AC.UK
- APARSEN@JISCMAIL.AC.UK
- ALLIANCE-ANNOUNCE@JISCMAIL.AC.UK
- research-dataman@jiscmail.ac.uk - Joy Davidson forwarded the email to this list
- all@list.scape-project.eu
- Archivliste archivliste@Lists.Uni-Marburg.DE
- CODATA International CODATA_International@kbx7.de
- DeMuseum demuseum@dhm.de
- digipres digipres@ala.org
- ERECS ERECs-L@LISTSERV.ALBANY.EDU
- Inetbib INETBIB@ub.uni-dortmund.de
- padiforum-l padiforum-l@nla.gov.au
- SWISS-Lib swiss-lib@switch.ch
- VDB Liste vdb-list@lists.hsu.hh.de
- WEB-ARCHIVE web-archive@cru.fr
- wiss-org wiss-org@bonn.iz-soz.de
- NCDD NCDD-DISCUSSIE@NIC.SURFNET.NL
- Archivi23 (list of Italian archivists). archivi23@lists.anaiveneto.org
- AIB-Cur (list of Italian librarians): www.aib.it/aib/aibcur/aibcur.htm3
- nestor (list of the German competence network for digital preservation): nestor@langzeitarchivierung.de
- FORSCHUNGSDATEN@LISTSERV.DFN.DE