Ensuring Perpetual Access:

Establishing a Federated Strategy on Perpetual Access and Hosting of Electronic Resources for Germany

Edited Final Report – February 2010

Prepared by:

Charles Beagrie Limited in association with Globale Informationstechnik GmbH

www.beagrie.com

A study funded by

The Alliance of German Science Organisations

PREFACE

of the “National Hosting Strategy“ working group of the Priority Initiative “Digital Information“ of the Alliance of German Science Organisations

Due to the availability of scientific information in digital form, usage of electronic sources has become a dominating factor in everyday life in many scientific fields. In this respect, the Alliance of German Science Organisations decided on the Priority Initiative “Digital Information” (http://www.allianzinitiative.de/en) in the year 2008.

The increasing acquisition of digital commercial content presents all scientific institutions with the challenge of making these contents permanently available via a suitable infrastructure. To accomplish this task in a coordinated effort at a national level seems imperative given the expected costs as well as the technical and organisational requirements of the task. It is the goal of the priority area “National Hosting Strategy” of the Alliance Priority Initiative “Digital Information” to initiate and support such action.

This study was conducted as basis for all further steps towards a national hosting strategy. It was financed jointly by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), the Fraunhofer-Gesellschaft, the Helmholtz Association, the Leibniz Association and the Max Planck Society. The study is intended to be the starting point from which to arrive at concrete ideas and activities related to a coordinated national hosting strategy. The intensive, often controversial but always constructive and continuing discussion now expands out of the working group into the public area to be continued there in the same manner. Invited for further discussion are all stakeholders with responsibility in the German science system to establish and finance sustainable structures for perpetual access as well as long-term preservation for electronic resources.

Bonn, February 2010

PREFACE

The members of the Alliance of German Science Organisations are the Alexander von Humboldt Foundation, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), the Fraunhofer-Gesellschaft, the German Academic Exchange Service (DAAD), the German Rectors' Conference (Hochschulrektorenkonferenz - HRK), the Helmholtz Association, the Leibniz Association, the Max Planck Society and the Wissenschaftsrat (German Council of Science and Humanities).

The Alliance members have agreed to coordinate their activities and to expand on the ideal of the innovative information environment by means of a Joint Priority Initiative “Digital Information” from 2008 to 2012 with the following goals to:

- guarantee the broadest possible access to digital publications, digital data and other source materials;
- utilise digital media to create the ideal conditions for the distribution and reception of publications related to German research;
- ensure the long-term availability of the digital media and contents that have been acquired from around the world and their integration in the digital research environment;
- support collaborative research by means of innovative information technologies.

This study has been commissioned by the Alliance in support of its goal to ensure the guaranteed long-term availability of the digital media and contents and has been supervised by the hosting working group of the Alliance membership.

We would like to thank all members of the Alliance hosting working group and individuals and organisations that were interviewed for the study and who generously gave their expert input and time to the development of this report.

CONTENTS

1. Executive Summary

1.1. Investment in Electronic Resources and its Impact

Research is critical to the health of Germany's economy and some 230 million euros are invested each year in scholarly resources for researchers in German Universities alone. E-journals and other scholarly commercial electronic publications are a critical and heavily used part of these resources particularly for researchers in science, technology, and medical fields. At the same time significant funds are also being invested in retro-digitisation of scholarly resources still in paper form such as books, journals or other materials such as photographs. The DFG's investment in retro-digitisation was 12 million euros in 2008 and has been between 25-50 million euros in total over the previous 5-10 years. These resources are also heavily used by researchers.

1.2. The Associated risks and Challenges

These are large public investments which need to be safeguarded and maintained and whose loss of access could severely impair research. Although there are many obvious benefits that accrue from publishing and accessing academic journals electronically, we need to be looking ahead if our growing collections of electronic research publications are still going to be accessible for researchers 5, 10, 50 or 100 years from now. High amongst the list of challenges, are ensuring that this material can continue to be accessed for the indefinite future (perpetual access) and be preserved (digital preservation). A particular challenge is that the owners of, and access platforms for, the files are now in the majority of cases the publishers, and publishers and their journal titles can merge, transfer or disappear over time. So we need new procedures, collaborations and sometimes organisations to ensure safe custody of these objects is maintained through time, technologies, and organisational changes.

The challenges for long-term access and preservation of retro-digitised materials are somewhat different to those for e-journals as institutions have greater direct control over all stages from creation through to use and access to the material. In principle they can also be re-digitised from originals. In practice however funding for digitisation is limited and the

quantity of non-digitised material is huge, so institutions and funders expect their investments in digitisation to be preserved and accessible.

1.3. Recommended Approaches for Hosting, Perpetual Access and Archiving

This study has been commissioned by the Alliance of German Science Organisations to help develop a strategy to address these challenges. In undertaking the study we were requested to focus on commercial e-journals and retro-digitised material. Evidence has been gathered via a wide-range of desk research and interviews with major relevant organisations in Germany and internationally and assessed in this report. **Our gap analysis showed the largest gaps in Germany to be provision for perpetual access for e-journals.** Taking into consideration the requirements identified by interviewees, potential solutions we have reviewed, and associated risks, we have recommended the following approaches:

Hosting for e-journals. **Access via publishers' servers** is the most common and well-tested approach for hosting of commercial e-journal content licensed by institutions in Germany. On the basis of work completed for the study, we believe it currently provides the most effective main solution for hosting and immediate access of this content in the majority of cases and is recommended to the Alliance. However, **some capacity to host e-journals within Germany will always be required in addition to this** when publishers cannot host appropriately and will need to be provided for.

Perpetual access for e-journals. This is much more immature in terms of development of appropriate services and successful business models. **In the short-term many libraries can exercise post-cancellation access via hosting on publishers' servers. We have noted this is not sufficiently robust as a single perpetual access solution long-term.** However given the immaturity of longer-term solutions this is the approach we would recommend for the immediate first horizon of the Strategy (perhaps years 1-3). However, some capacity to provide archiving and post-cancellation access for e-journals within Germany would be required in addition to this for when publishers cannot host appropriately for post-cancellation access or other reasons. **In working towards longer-term solutions we recommend two options are explored in parallel:**

- **Development of independent perpetual access capacity in Germany with international partners.** We believe the larger market, spread of risk, and broader combination of expertise and experience would make this our preferred option to developing a purely German solution which would have a higher risk profile.

- **Participation in Portico**, a US based not-for-profit archive. Possibly one institution in Germany could pilot this service over the first horizon of the Strategy (there are no German library participants currently). A dialogue could be established to discuss potential weaknesses from a German perspective and future directions.

Hosting for retro-digitisation materials. Our interviewees generally expressed a high level of confidence in existing infrastructure and solutions at national/regional level and local level. **Areas of current concern mentioned the needs of small and medium sized institutions. We have recommended that the Strategy considers the federated role that the Verbünde and regional library services could play in support for them. Another area of concern is around the sustainability of hosting. We have made a number of recommendations which may assist in addressing this, e.g. by helping to contain short-term or long-term costs.**

Archiving for retro-digitisation materials. There are now a number of established national/regional archives for retro-digitised material based around the two national digitisation centres and the national legal deposit system Kopal. Other services such as the Verbünde and regional library services are also developing or have deployed archive storage services for digitisation archive master images. **We recommend ongoing development based on these systems at national/regional level to maximise economies of scale. Discussions on forming a "virtual national library", or exploration of archiving solutions for e-journals within Germany might also contribute to this aim.**

These approaches are supported throughout the study by discussion of organisational, technical and financial issues and by separate confidential cost and budgetary data. Other potential solutions and approaches are also included and remain open for consideration. Extensive consultation is now required to confirm the proposed Strategy, and associated funding and governance. We have indicated a set of actions as immediate building blocks in the recommendations which can provide a foundation for future development.

2. Overview of the Study Report

2.1. Objectives

This study focuses on the "national hosting strategy" of the Alliance Priority Initiative: Digital Information. The study's objectives were:

- Analysing existing international approaches and current initiatives on a national level, both centrally and de-centrally organised;
- Taking into account and prioritising the specific practical requirements necessary for the effective and sustainable implementation of such a strategy;
- Detailing possible use cases, solutions, and scenarios for a strategy and including a sound recommendation on a roadmap towards implementing the strategy to be pursued by the contracting authority.

In undertaking the study we were requested to focus on e-journals and retro-digitisation but to indicate in the strategy how it could be extended to other electronic commercial publications (e-books and databases) in future stages.

2.2. Methodology

The study was conducted by consultants from Charles Beagrie Limited and Globale Informationstechnik GmbH between May and November 2009.

The study collated information and findings (either on a public or confidential basis) in four major areas: Content, Licensing, Perpetual Access, Preservation and Hosting; Technical Infrastructure and Standards; Business Models, Costs and Funding; Organisational and Policy Issues.

We reviewed information gathered from a survey of organisations in Germany and the investigation of international initiatives to identify strengths and weaknesses of the existing structures against stated objectives for the strategy. We drew together and identified core requirements from the survey of organisations in Germany and our gap analysis provided a draft requirements analysis, prioritisation, and series of related use cases.

We recognised that a strategy would be highly complex and present significant challenges and opportunities to which more than one approach could apply, and could require a phased implementation over a period of years. Therefore we defined and assessed a number of possible solutions, risk profile scenarios for approaches and timescales for realising the strategy with the clients which have been reflected in this final report and our final recommendations and roadmap.

2.3. Contents of the Study Report

3. Introduction. The introduction provides the background and methodology for the study. Of key importance to the remainder of the study is its discussion of terms such as archiving, perpetual access, and hosting and their definitions and relationships.

4. The International Context. The first stage of desk research and interviews collated information on 11 selected international initiatives and projects which might offer best practice, similar experience or services relevant to the development of a Federated Strategy on Perpetual Access and Hosting for Electronic Resources in Germany. We used our existing knowledge and the client's suggestions in the Invitation to Tender to select the most relevant projects and services in scale and impact for the study. The 11 interviews covered three continents: Australia and New Zealand (interviewees – CSIRO and National Library of New Zealand); Europe (interviewees - JISC Collections and the British Library in the UK and The Koninklijke Bibliotheek in the Netherlands); and North America (interviewees CLOCKSS/LOCKSS, HathiTrust, Los Alamos Laboratory e-Research Library, OhioLink, University of Toronto, and Portico).

5. The Current Position, and 6. Views on the Future Position in Germany. In parallel with the international interviews, desk research and interviews were used to complete a survey of current hosting services and projects and future requirements in Germany. We used the client's suggestions to select 28 German organisations for interview. The interviews within Germany used a structured questionnaire which aimed to capture data on the current position in Germany and in addition sought views via a series of open-ended questions on what the future position on hosting and perpetual access and the strategy for this should be.

7. Technical Infrastructure and Standards. The ITT for the strategy identifies a number of technical issues for the study to address. These issues are touched upon throughout our study but some needed more detailed introduction and explanation for the broader audience of this report. Of particular importance is the discussion of architectures and standards.

8. Gap Analysis. The purpose of the gap analysis is to identify discrepancies between current and ideal states for perpetual access and hosting for electronic resources in Germany. The results are used to formulate use cases and prioritise underlying requirements as a basis for building the Strategy. The analysis identifies a number of large gaps for perpetual access and preservation of e-journals and mainly low to medium gap levels for retro-digitised material in Germany.

9. Use Cases. The aim of these use cases is to explain reasons why and where action is needed, indicate potential solutions and the drivers for the further development of the Strategy. Seven use cases are outlined for e-journals and retro-digitisation materials.

10. Risk Scenarios, Potential Solutions, and Recommended Approaches. In this section of the study we outline potential solutions and different scenarios that address the use cases. The solutions outlined often overlap and none will provide a unique or complete answer to the use cases we have described. Different levels of risk, effort and need for innovation apply for each solution and potential combinations of them, so we have outlined a number of risk profile scenarios. These then lead into our recommended approaches.

11. Future Development of the Strategy. The final section sets out the initial building blocks and steps that should feed into the future development and implementation of the Strategy including consideration of other content types excluded from the initial implementation.

2.4. Summary of Report Recommendations

In considering the recommendations and approaches suggested in this study, there are a number which can be seen as initial building blocks towards the Strategy almost regardless of the final direction of travel. The consultation period required for developing the Strategy is likely to be significant as there is no single committee or funding source in Germany that can be engaged for support of the Strategy. In the interim, these building blocks can be pursued with relatively modest resources and extensions of existing work and contribute to the final

Strategy adopted. We would recommend that they might be prioritised for the years 1 and 2 of the implementation. These building blocks are highlighted in blue background in the summary table of recommendations below.

REC. NO.	DRAFT RECOMMENDATION	SECTION REFERENCES
1	Maintain an international dimension to the Strategy, evaluate potential international partnerships and service providers, maintain an oversight of emerging international best practice and trends.	4.5 7.5
2	Wide consultation and further consensus building amongst all stakeholders including publishers will be a critical element of further development and implementation of the Strategy given the complex organisational landscape in Germany.	5.3 6.3 8.5
3	Develop a federated organisational and funding model to maintain and develop the national licensing initiative. Include perpetual access and hosting in its scope. Explore if savings achieved from licensing collaboratively might contribute to funds for shared perpetual access and hosting requirements.	5.4 8.4 8.5
4	Develop management information for licensing of e-journals and requirements for perpetual access and hosting. This should facilitate analysis of the need and size of demand for federated action.	5.5 8.2.4 8.4
5	Future developments of the Strategy can be guided by a “content map” which provides a mechanism for distinguishing different types of databases and e-books and places commercial content into a context alongside open access materials.	5.7 11.4
6	The findings of the “Legal Framework (P6)” Alliance working group which will take into account any new legislation on digital preservation, should inform future updates to the Strategy.	6.3
7	Develop a service orientated reference architecture model and corresponding exchange formats and communication protocols to illustrate how different service components can contribute to the Alliance’s vision for an integrated information environment for the German science and research community.	7.2.2
8	Develop model licence and clauses for DFG national licences and seek to share, maintain and extend these within Germany.	8.2.1
9	Incorporate TRANSFER Code of Practice into model licences and agreements. Seek to influence future development and strengthening of the TRANSFER Code of Practice.	8.2.1

Rec. No.	Draft Recommendation	Section references
10	Develop and agree technical guidelines and requirements to accompany licence agreements. These should promote common standards e.g. use of the NLM DTD/schema.	8.2.1 8.3.2
11	Maintain and continue to develop DFG Digitisation Guidelines.	8.2.2
12	Facilitate rights identification and clearance for the digitisation of out-of-print and orphan works. If outcomes of pilot project are promising continue development of the ARROW registry.	8.2.2
13	Maintain and continue to develop investment in applied R&D in areas such as workflow systems, sustainability, and cost/benefit studies, for retro-digitisation, archiving, and hosting.	8.2.3 8.4
14	Invest in development/purchase of digital preservation infrastructures and tools, and services to maintain them.	8.2.3
15	Consider the federated role that Verbünde and regional library services could play in support of retro-digitisation storage, particularly for small and medium-sized institutions.	8.2.3 8.5
16	Develop funding, and a financial strategy and economic models to provide a framework for decision-making on support and prioritisation of hosting and perpetual access services.	8.2.4 8.4
17	Develop an archive content selection and development policy to scope required initial coverage and help prioritise content to be included as it develops.	8.2.4 8.4 8.5
18	Select or establish service provider(s) who can deliver implementation of your perpetual access rights and/or local hosting rights for e-journals.	8.2.4
19	Establish dialogue with publishers, service providers and other potential partners on technical and service level issues, coverage of journals, and approaches for small publishers.	8.2.4 8.5
20	Evaluate archiving practice and preservation and access services for supplementary materials to journal articles; and use of persistent identifier services for linking supplementary materials and research datasets associated with journal articles.	8.2.4 8.3.2 11.4
21	Develop and apply criteria and standards for hosting and perpetual access/preservation services and certification processes for them.	8.2.4 8.5
22	Decide on mechanisms for content acquisition and transfer for e-journals and whether to store material as acquired, or normalise to an internal format.	8.3.2

Rec. No.	Draft Recommendation	Section References
23	Design a search and browse interface that meets current standards and expectations in terms of ease of use and presentation of results.	8.3.3
24	Develop an authentication and authorisation infrastructure based on SAML (such as Shibboleth) that meets international standards and supports single-sign-on to a wide range of resources. Continue to support legacy mechanisms such as IP and username/password.	8.3.4
25	Implement mechanisms to maintain an archive of licence entitlements, administer rights management, and appropriate access control.	8.3.5
26	Establish a Technical Panel that can review and decide technical issues and develop appropriate guidelines, standards, and inter-operability.	8.3.5 8.4 8.5
27	The implementation of the Strategy should address the need for sustainability and affordability and its phasing take into consideration the budgetary cycles and commitments of potential funders.	8.4
28	Decisions made about business models and financial backing for the strategy, short through to long term, will impact upon the type of organisational model(s) that is practical to adopt for the implementation of the Strategy. The organisational model(s) should be in keeping with the type of archiving and hosting solution(s) adopted.	8.5
29	The organisation to implement the Strategy should aim to bridge interests, foster co-operation, and take into account the federal structure of Germany.	8.5
30	The Strategy should segregate different content types and consider appropriate policies and implementations for them.	11.4

3. INTRODUCTION

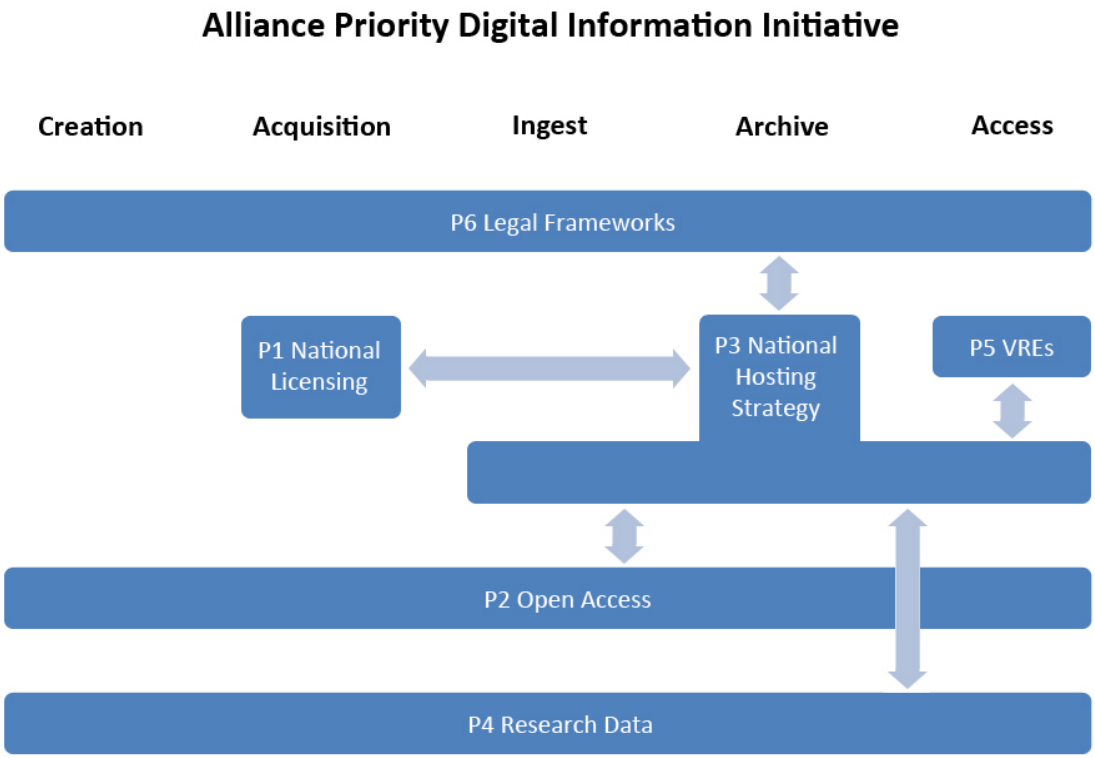
"Equipping scientists and scholars with the information infrastructure best suited to meeting their research needs is the guiding principle of this priority initiative...The increasing acquisition of commercial publications in digital form presents all scientific institutions with the challenge of making these contents permanently available via a suitable infrastructure. There is consensus that a national strategy is indispensable, both for reasons of cost and because of the technical and organisational requirements of the task." Excerpts from "Priority Initiative Digital Information " (Alliance of German Science Organisations 2008).

3.1. BACKGROUND

Germany has traditionally been one of the Organisation for Economic Co-operation and Development (OECD)'s top performers in science, technology and innovation. With a mature national innovation system, including a number of large, well-established research institutions and firms, it has a large and growing share in total OECD high- and medium-high-technology exports, and is the fourth most intensive patenter in the OECD area (adjusted for population). However, its productivity performance has been slipping against the leading OECD countries in recent years. To encourage a reversal of this trend, Germany has introduced a wide range of policies to support innovation. The federal government's High-Tech Strategy launched in 2006 is a national strategy which encompasses all ministries. It sets out strategies for 17 "future fields" and aims at translating ideas from basic technologies as rapidly as possible into marketable products, services and processes. Under the Initiative for Excellence, Germany is also providing project funding to support graduate schools, "excellence clusters" and frontier research at universities (OECD 2008).

Access to information in digital form provides huge benefits to German researchers and research productivity. Increasing and guaranteeing efficient access to digital information of all kinds such as journals, theses, conference proceedings and other "grey literature", digital images, and research data sets, are essential requirements for research and innovation in Germany.

The Alliance of German Science Organisations has therefore recently established a priority initiative for digital information. The Alliance partners are committed to increasing Germany's competitiveness in world-class research by concentrating skills and resources, and improving the co-ordination of their activities. The digital information initiative is focusing on six complementary and related major priorities: 1. national licensing; 2. open access; 3. a national hosting strategy; 4. primary research data; 5. virtual research environments (VREs); and 6. legal frameworks (Alliance 2008). Alliance working groups are focussing efforts on each of these different areas.



Relationship of the "National Hosting Strategy" to the other Alliance Digital Information Priority Initiatives and Stages in Scholarly Information Workflow.

3.2. Aims and Objectives of the Study

This study focuses on the "national hosting strategy" of the Alliance Priority Initiative and is steered by one of the Alliance working groups. The need for a strategy has been driven by many factors primarily relating to changes in scholarly publishing of journals. Scholarly publishing is increasingly based on electronic publications. Access to these publications is

frequently licensed (i.e. access is rented from a publisher or supplier) rather than based on ownership of a local copy for preservation and access (as exists with the traditional print journal). Libraries are increasingly cancelling their print subscriptions and moving to e-only access. As this trend is progressing, finding resolutions to the archiving of, and continuing access to, scholarly e-journals has assumed even greater urgency. Significant efforts have been made to explore options for e-journal archiving which can provide greater assurance of continued access into the future. Alongside these requirements for archiving and continuing access, hosting is often required by research organisations for accessing specialist digital publications, or accessing and archiving the results of open access publishing or digitisation projects.

The goal of the strategy is to establish and operate an efficient infrastructure for the storage and long-term preservation of digital documents, which can guarantee perpetual access to licensed commercial publications and retro-digitised library materials. The study's objectives were to analyse the status quo regarding the establishment of a strategy and infrastructure. This included:

- Analysing current existing initiatives on a national level and existing international approaches, both centrally and de-centrally organised;
- Taking into account the specific practical requirements necessary for the effective and sustainable implementation of such a strategy;
- Detailing possible scenarios for a strategy and including a sound recommendation on the strategy to be pursued by the contracting authority.

In undertaking the study we were requested to focus on e-journals and retro-digitisation but to indicate in the strategy how it could be extended to other commercial publications (e-books and databases) in future stages.

3.3. Methodology

The study was conducted by consultants from Charles Beagrie Limited and Globale Informationstechnik GmbH between May and November 2009 as a series of tasks: initiating the project with the clients; information gathering from desk research and interviews; preparing a gap analysis, requirements analysis and use cases; developing options,

recommendations and a roadmap; reflection and feedback by the clients at key decision points; and submission of a final report.

The list of Interviewees was agreed with the clients at the project initiation meeting and a pro-forma interview questionnaire was used to ensure consistent coverage. We collated information and findings (either on a public or confidential basis) in four major areas:

- Content, Licensing, Perpetual Access, Preservation and Hosting;
- Technical Infrastructure and Standards;
- Business Models, Costs and Funding;
- Organisational and Policy Issues.

We reviewed information gathered from a survey of organisations in Germany and the investigation of international initiatives to identify strengths and weaknesses of the existing structures against stated objectives for the strategy. We drew together and identified core requirements from the survey of organisations in Germany and our gap analysis provided a draft requirements analysis and series of related use cases.

We recognised that a strategy would be highly complex and present significant challenges and opportunities to which more than one approach could apply, and could require a phased implementation over a period of years. Therefore we defined and assessed a number of possible scenarios for approaches and timescales for realising the strategy with the clients which have been reflected in this final report and our final recommendations and roadmap.

3.4. Discussion of a National Hosting Concept

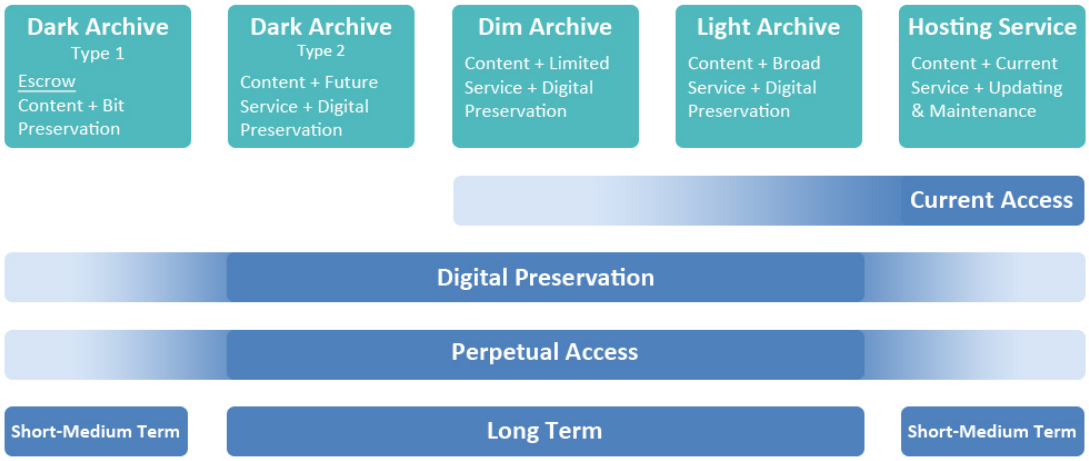
The Alliance Working Group recognises that an emerging strategy could be scoped in a variety of ways and be implemented by a range of potential solutions and service options. The concepts and terminology behind this strategy are complex and we have provided a glossary with suggested definitions and explanations of key terms such as “dark archive”, “dim archive”, “light archive”, “hosting”, “access”, and “digital preservation” as an appendix (section 13).

As outlined earlier in the Introduction, the strategy has to establish and operate an efficient infrastructure for the storage and long-term preservation of digital documents, which can

guarantee perpetual access to licensed commercial publications and retro-digitised library materials. The preservation and access roles could range from:

- "**dark archives**" services with no current access for users and future access dependent on the occurrence of specific events. Typically these dark archives can be divided into two main types: **type 1** - those only providing a form of escrow or "bit preservation" of content that is suitable as a short-medium term solution for guaranteeing access; and **type 2** - those providing the bit preservation of the content plus some degree of associated services for future access (this may include a "back-up" access service should a primary access service fail for any reason, and digital preservation planning and preservation action services such as file format migration which will keep the content accessible in the future). These provide for and can help guarantee long-term perpetual access; to
- "**dim archives**" providing bit preservation for the content plus digital preservation planning and actions for long-term perpetual access, and also limited current access (perhaps limited to on-site users or previous subscribers post-cancellation, etc); to
- "**light archives**". A crucial concept behind many archives is agreeing to restrict access over a period of time or until specific events occur in the future. At such points "dark archives" and "dim archives" may transform to "light archives" where a much broader range of access is permitted to users for specific content; and finally
- "**hosting services**" providing the point(s) of full current access to authorised users and ongoing maintenance, updating, and security of that content (including routine back-ups and disaster recovery planning). These may provide for and help guarantee short-medium term horizons for access but are not focussed on digital preservation or organised for the long-term and would need to partner with or be combined with other services to achieve or guarantee longer-term perpetual access.

  Various combinations of these options can be tailored to provide services for specific content, time horizons and functional requirements.



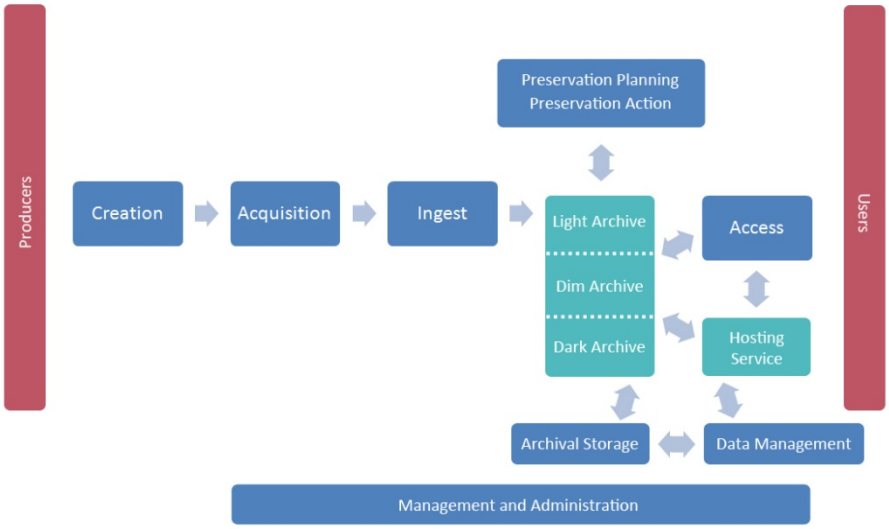
Concepts and Properties of Archives and Hosting in the Strategy and their Relationships ©Charles Beagrie Ltd 2009. CreativeCommons Attribution-Share Alike3.0

Key: solid colour represents core properties and fading colour represents weaker properties of archives and hosting services.

Similarly the strategy may be implemented by individual libraries and other organisations and/or national and regional consortia, within Germany; but could also encompass a range of different services and collaborations with publishers or others internationally; or various combinations of these.

For these reasons the terminology and title for the strategy and this report are important: we have suggested and the Alliance working group has agreed “ENSURING PERPETUAL ACCESS: ESTABLISHING A FEDERATED STRATEGY ON PERPETUAL ACCESS AND HOSTING OF ELECTRONIC RESOURCES FOR GERMANY” as the title that best captures and conveys the concept of a national hosting strategy for this study. We expand and build on this concept for the strategy in the report and refer to a Federated Strategy on Perpetual Access and Hosting from this point onwards.

3.5. Model used in the Study



Model used for discussion of the Federated Strategy on Perpetual Access and Hosting of Electronic Resources for Germany

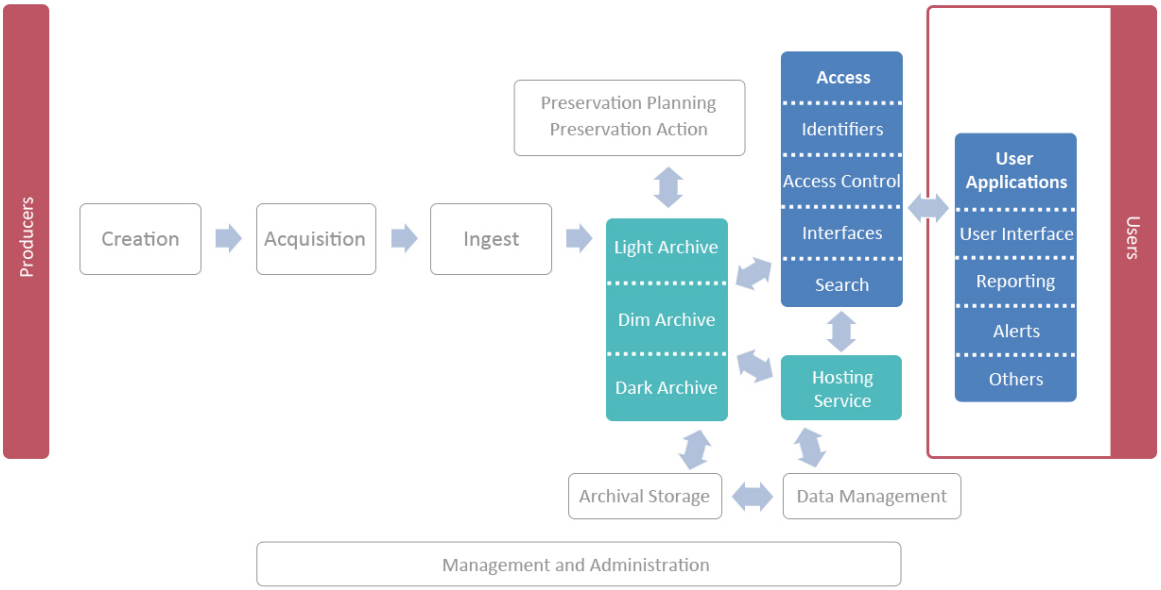
To illustrate these concepts, we have used a simple functional model in later sections of the study (see section 8), which may also be helpful for future development and explanation of the Strategy. It is not an "architecture" but a reference model to aid understanding and development of the Strategy. It is built on selections from the Keeping Research Data Safe Activity Model (Beagrie, Chruszcz, and Lavoie 2008) and its extensions to and adoptions of the Open Archival Information Systems (OAIS) Reference Model (CCSDS 2002), and its integration with the concepts of Archive and Hosting Service developed for this study.

The key functional components are:

- **Creation**. Activities involved in the process of creation of digital resources including the negotiation of rights and permissions, production and processing of the resource, description and documentation, customising software and data management, and generating products for future dissemination or preservation;
- **Acquisition**. Activities involved in the process of acquiring digital resources including applying a selection policy, negotiating submission agreements, and providing support for transferring the resource from the producer;

- **Ingest**. Activities involved in taking in and loading of digital resources including, receiving, reading, quality checking, cataloguing (including metadata, documentation, etc.) and reference linking to the point of insertion into the archive;
- **Archival Storage**. Activities involved in the storage and retrieval of Archival Information Packages (AIPs);
- **Data Management**. Activities involved in populating, maintaining, and accessing both descriptive information which identifies and documents holdings and administrative data used to manage them;
- **Preservation Planning and Preservation Action**. Activities for monitoring, providing recommendations, and taking action, to ensure that the information stored in the archive remains accessible over the long term, even if the original computing environment becomes obsolete. Preservation Action covers the process of performing actions on digital objects in order to ensure their continued accessibility. It includes evaluation and quality assurance of actions, and the acquisition or implementation of software to facilitate them;
- **Access**. Activities which make the information holdings and related services visible to Users.
- **Management and Administration**. Activities needed to control the operation of the other functional entities on a day-to-day and long-term basis.

In addition to the functional entities described above, a number of basic Common Services are available, such as operation system services, network services, and security services but these are omitted from the model for clarity. It is highly simplified but when required aspects of the model can be enlarged as needed to show specific applications and activities e.g.



The Study Model with the Access and User Domains expanded to show Applications ©Charles Beagrie Ltd 2009. CreativeCommons Attribution-Share Alike3.0

4. The International Context

4.1. Overview

The first stage of desk research and interviews collated information on 10 selected international initiatives and projects which might offer best practice, similar experience or services relevant to the development of a Federated Strategy on Perpetual Access and Hosting for Electronic Resources in Germany. We used our existing knowledge and the client's suggestions in the Invitation to Tender to select the most relevant projects and services in scale and impact for the study. The 10 interviews covered three continents: Australia and New Zealand (interviewees – CSIRO and National Library of New Zealand); Europe (interviewees - JISC Collections and the British Library in the UK and The Koninklijke Bibliotheek in the Netherlands); and North America (interviewees CLOCKSS/LOCKSS, HathiTrust, Los Alamos Laboratory e-Research Library, OhioLink, and Portico). In addition although not interviewed, information via email from University of Toronto was included.

Entries for each organisation have a summary of key findings and conclude with the interviewee's suggestions for a strategy for Germany.

4.2. Australia and New Zealand

4.2.1 CSIRO

Key findings: *content and licensing: until recently negotiated with publishers to purchase content and host locally, but have recently decided to move to remote access via publishers' servers. As part of their negotiations with publishers they are asking them for details of their long-term preservation arrangements if they are not part of Portico.*

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) is a prestigious, large, and distributed Australian national science organisation that supports research on social, environment and economic stability. The library provides a service to CSIRO staff and has a physical presence on 39 of the 56 sites as well as a virtual presence. They have a range of searchable, licensed content, including 7,000 e-journal titles, and are making a big push to provide more electronic resources and desktop delivery. Until recently CSIRO negotiated with publishers to purchase content and host locally, but have recently

decided to move to remote access. They feel that the notion of local hosting is passé, and cite the following reasons:

- the software they were using is no longer supported and it is not cost-effective to write their own;
- not all publishers will provide content to host, and so it has been split, which is not helpful to the users, and difficult to manage and serve;
- aggregators now have the functionality they need;
- hi-speed broadband networks mean it is no longer necessary to hold content locally;

They intend to dispose of their existing e-content as they migrate to remote services. However they are also not disposing of print collections until they are sure the security of the e-content they use. As part of their negotiations with publishers they are asking them for details of their long-term preservation arrangements if they are not part of Portico.

Interviewee's Suggestions for the Strategy: *The challenges for national hosting are co-operation and local politics. If individual universities go too far down their own path it is difficult to turn things around. University of Toronto is a good model for local hosting. National rights have to be sought from publishers in the national strategy.*

4.2.2 National Library of New Zealand

Key findings: *content and licensing: The National Library of New Zealand (NZNL) has a mandate for the sustainable long-term preservation of content. No major international publishers are based in New Zealand and legal deposit has a focus on locally produced material. EPIC a national licensing scheme for e-journals administered by NZNL provides current access but has no continuing access provisions; Technical: NZNL has procured the development of a preservation system with ExLibris and Sun Microsystems. This will also be marketed to other libraries.*

The National Library of New Zealand (NZNL) has a mandate for the sustainable long-term preservation of content and has a number of digital library projects, of which the most pertinent for the study is the National Digital Heritage Archive (NDHA). NDHA refers to the

technology, new business processes and other organisational changes the National Library has put in place to provide long-term preservation of, and ongoing access to, the digital heritage collections under their guardianship. At its core, the NDHA is a system of software applications that support a digital storehouse for the websites, sound and vision files, digital images and other born-digital and digitised items that make up the Library's growing digital heritage collections. An imperative strand of NDHA's work is archiving journal article material, much of it born digital. New Zealand has legal deposit of e-journals and receives the electronic 'published' format, rather than the 'source' format. They considered using an existing open source repository set-up, for example, DSpace, but have instead procured the development of a system with ExLibris and Sun Microsystems. Security, Scalability, Open standards, and Compliance were requirements for the system, although they argue that scalability, storage size etc are not the most important issues for national preservation activities, as storage systems are beginning to manage their own size. They are also not being prescriptive about format (unlike others such as Portico and HathiTrust) and take a more eclectic approach.

New Zealand also has a national licensing scheme, EPIC, which makes available 16,000 full-text titles and databases to all New Zealanders via public libraries. The National Library pays for these licences; however the content is not acquired and all access is lost if a subscription discontinues.

Interviewee's Suggestions for the Strategy: *There is a real imperative for national hosting strategies because few individual organisations have the mandate and/or resources to do what is needed. Institutions need to see the incentives for participating in a national strategy, such as those identified by the Blue Ribbon Task Force on Sustainable Digital Preservation and Access Dec 2008 (http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf). A study by Berkeley University (http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/) based on 1999/2000 information showed that 92/93% of all unique published material appears online, and that print publishing is also still increasing. The challenge therefore is keeping up with the growth of new publications which is impossible for each organisation. There is the potential for organisations like a national library to be central to a national hosting strategy, which should include local hosting of both metadata and objects.*

A successful strategy will need to find a way into the country's/institution's funding system and policy issues, and jurisdiction and governance issues will need to be resolved. Rights issues across the industry have to be solved. Leadership, and champions with vision drive and ability, are key success factors, as is the need to minimise the impact of any local competitiveness. It is important to focus on the business case for long-term preservation and the simplest solutions with long-term benefits, rather than the 'gimmicks of delivery'. Publishers will be integral to the delivery of the strategy and it will be necessary to think carefully about the role of publishers, and to gain their confidence. The delivery mechanism from publishers' needs to be relatively stable, with minimal changes. Hosting resources nationally needs to be operating with open standards, but not necessarily open resources, so accessible by local APIs with no proprietary lock-ins. This will be easier for the big publishers, and harder for the smaller.

One of the problems with e-journals in the context of national strategy is that there is no best solution anywhere, although the Royal Library in the Netherlands is a good example. This also acts as a primary back-up if the publisher's system fails. In terms of a national strategy this is a very interesting model. There are good working relations with the publishing community, a deposit agreement, a last bastion of material for the long-term, and back-up for publisher failure.

The model may be difficult to apply at a federal level in Germany but may work at a State level. Although an alternative model could be to provide state-level/regional/local nodes for logging into a national structure. At state-level issues of redundancy, co-operation, etc would need to be considered. The NZNL has a mandate for the long-term preservation of content. The strategy may need to distinguish between institutions with a preservation mandate and those that just want long-term access.

4.3. Europe: The UK and Netherlands

4.3.1 JISC Collections

Key findings: *content and licensing: operates a national licensing scheme for UK tertiary education based on a model licence for negotiation with publishers. The model licence has clauses for perpetual access and use of TRANSFER Code of Practice. Technical: provides*

technical guidelines for publishers. There is a UK Access Management Federation for single user name and password sign-on via Shibboleth. Business and organisational: JISC Collections has a financial strategy and economic models for licensing and hosting showing a range of options for publishers and JISC on hosting, promotion and funding. Interesting evolutions in the organisation over 13 year period and in funding models which have parallels for Germany.

Originally operating within the Joint Information Systems Committee (JISC) in the UK, JISC Collections was made into a separate limited company in 2006 that mutually trades with its members and is now a JISC funded service. It works on behalf of its members - all UK Higher Education (HE) and Further Education (FE) institutions and Research Councils (RCs) that receive direct funding from the UK HE and FE funding bodies. It has begun expanding membership to include affiliate members – organisations in the UK and overseas which qualify for membership through their engagement in life-long learning.

JISC Collections reflects over 13 years in evolution of national licensing in the UK: its origins lie in the three year National Pilot Site Licence Initiative begun in 1995 and the National Electronic Site Licensing Initiative commenced in 1998, which introduced the concept of a Managing Agent and piloted the use of a national model licence for use in collectively negotiating agreements with publishers.

Five Working Groups support JISC Collections in identifying and acquiring online resources for inclusion in the collections catalogue. The Working Groups are made up of representatives from the further and higher education community and the research councils.

The range of resources licensed for use by the agreements constitutes a national collection of online resources for education and research. It provides members with a collections catalogue of free and subscription-based online resources such as full text databases, e-books, digital images, e-journals, online film, learning materials and geospatial data.

The online resources in the collections catalogue are licensed from publishers, aggregators and content providers. Each institution decides which resources it wishes to subscribe to based on the needs of their users.

The negotiations for e-journals are managed under the National Electronic Licensing (NESLi) scheme run by JISC Collections. There are currently 4 model licences for datasets,

e-books, e-journals, and moving image and sound. Perpetual access clauses are included in the e-journals licence. Archival rights must be there in the e-journal licence (the only exception currently is JSTOR). JISC Collections has recently amended the NESLI licence perpetual access clause to say the publisher must deposit with a third party service and record that in the license. It has also built the TRANSFER Code of Practice (UKSG 2008) into the licence to address issues raised by transfer of titles between publishers. Key features of NESLi are as follows:

- Use of the Model NESLi2 Licence for Journals;
- A clearly defined list of publishers to seek agreements with, based on feedback from the community;
- An independent and experienced Negotiation Agent;
- Pre-defined criteria to assist the negotiation process;
- Order channels and access routes that support the negotiation process;

JISC Collections also represents the JISC in the Knowledge Exchange – a European partnership with includes DFG for Germany and has collaborated on multi-country licensing with Knowledge Exchange partners.

Most resources in the Collections are not hosted locally but accessed via publishers' servers. However some databases are hosted in JISC-supported Data Centres when they require added-value services not available from the publisher e.g. Ordnance Survey digital map data. JISC Collections works to three principle cost categories for hosting and promotion when acquiring electronic content for the JISC Collections catalogue. See the JISC Collections Financial Strategy (http://www.jisc-collections.ac.uk/about_collections/coll_financial_strat.aspx) and JISC Collections Economic Models (http://www.jisc-collections.ac.uk/about_collections/coll_financial_strat.aspx#a) for further details.

JISC Collections has purchased some digitised collections e.g. Early English Books Online (EEBO) in perpetuity. There can be difficult issues around costs and services post purchase for such collections. Once sold, a publisher may have little incentive to invest in the platform

if it has a one-off fee. However if there are ongoing access fees these can also be a barrier and reduce uptake amongst intended audience for whom it was purchased.

JISC has also made substantial investments in the digitisation of newspapers and other assets through the JISC digitisation programme and consequently these materials are also of interest to JISC Collections.

The JISC Guidelines for Publishers are a series of fact cards that explain how publishers can comply with the JISC Information Environment Standards and how this benefits both publishers and their customers (JISC, Not Dated a). JISC Collections ensures that all the online resources in their portfolio are compliant with the required access management and authentication standards of the JISC. JISC has established the UK Access Management Federation to allow institutions to manage access to online resources through single sign-on using freely available open source software such as Shibboleth. By separating the authentication (which will be controlled by the institution) from the authorisation (which is based on user attributes and controlled by the resource provider), it facilitates granular levels of access to resources and requires users to remember only one username and password (JISC, Not Dated b).

Two JISC funded projects related to this study were noted and further information gathered. PECAN is a short JISC-funded project to investigate how best to support libraries and their patrons through access to e-journal content post-cancellation. To improve provision of post-cancellation access, two facilities are being explored: a registry of entitlement (which has reliable information on the journal content that has been subscribed to by libraries via NESLi2) and a secure virtual archive (providing secure and robust access to back journal content). As such, the intention of this project is twofold. Firstly it aims to investigate the required policy and procedures needed to establish an accurate registry of subscription information. Secondly it aims to investigate and propose a candidate technical infrastructure for a central UK journal archive that would provide appropriate controlled access to licensed material, in a robust and secure manner (http://edina.ac.uk/projects/pecan/).

The second relevant JISC funded investigation is the PEPRS project. The purpose of the PEPRS project, jointly conducted by EDINA and the ISSN International Centre, Paris, is to scope, build and test a pilot for an E-journals Preservation Registry Service [PEPRS]. The

overall aim is to provide librarians and policy makers with information on provision for continuing access all scholarly work published in e-journals. The project will pilot an online facility that would let a range of stakeholders check the archival provision for e-journals and to identify the gaps in such provision (http://www.edina.ac.uk/peprs).

Interviewee's Suggestions for the Strategy: *Archival rights are complex and a challenge for all libraries and purchasing consortia. Some current solutions are aimed largely at disaster scenarios which may be very rare. A more likely scenario especially in economic downturns is the need for post-cancellation access. Post-cancellation access is often available via the publisher at "reasonable" or undefined cost in the licence. However at a minimum a third-party mechanism (or local hosting) is still needed to provide the guarantee of perpetual access which is not possible via the publisher alone. There is also another future risk if the publisher is the sole means of access, as libraries then have a weak negotiating position on costs for post-cancellation access –local hosting or third-party post-cancellation access options at least gives some bargaining power even if not used.*

Local loading/hosting is not done in the UK but does give greatest security for the content because you have the content but local hosting is expensive and hard to do well. Germany may need to consider further economies of scale if nationally it chooses this route – perhaps something European or building on existing partnerships with countries in the Knowledge Exchange.

4.3.2 The British Library

Key findings: *content and licensing: major e-journal mapping study completed for UK legal deposit with data on UK publishers of interest to a German strategy for same licensed material; technical: Developing a shared peer storage network consisting of 4 mirrored storage nodes and 6 access nodes with other UK legal deposit libraries. Practical experience of ingest challenges for e-journals. Selection of JPEG2000 for image storage and NLM DTD as preferred format for e-journals. Lead organisation for the PLANETS digital preservation project; business and organisational: Model of collaborative preservation network of deposit libraries. Formation of a Technical Advisory Panel for deposit of e-journals.*

The British Library is one of the World's largest research libraries, active within the UK and internationally. It is part of a network of 6 legal deposit libraries in the UK (British Library, National Library of Scotland, National Library of Wales, University of Cambridge, University of Oxford, and Trinity College Dublin).

Legal deposit legislation for electronic publications has been introduced in the UK, but specific regulations for different types of content remain to be enacted. In the interim, pilot projects are underway on e-journals with publishers and on Web-archiving. Electronic content is also licensed from publishers to support the British Library's document supply business. Major digitisation projects are underway in partnership with JISC and publishers. Current projects include the digitisation of:

- 20 million pages of 19th century literature [approximately 80,000 books];
- 1 million pages of historic newspapers in addition to the 3m already digitised;
- 4,000 hours of Archival Sound Recordings in addition to the 4,000 hours already digitised.

By 2011 they expect to have built a secure, resilient and scalable storage system with 4 mirrored storage nodes at the British Library (London and Boston Spa), the National Library of Scotland and the National Library of Wales. It is expected access to stored material will be possible from all 4 nodes and from an additional 3 "access nodes" at Trinity College Dublin, and the Universities of Cambridge and Oxford. This storage system will contain over 900,000 items and be 175TB in size. It will contain content from many sources, including digitised books and newspapers, newspapers and academic journals received digitally from publishers, digital sound and many forms of donated digital items. The Digital Library System will become the default repository for digital material acquired or created by the British Library.

The Digital Library System will include a rights management system that will enable researchers to access material whether they are in the Reading Rooms of any of the Legal Deposit Libraries (i.e. access nodes are located at all 6 UK legal deposit libraries), at a remote location (over the internet) or as members of institutions that have a relationship with the British Library. They will do this while respecting the rights of rights-holders and also have put in place the capability to charge for access, where appropriate.

The British Library has undertaken pilots with ingesting e-journals from different publishers. These have underlined the very considerable challenges involved and provided valuable practical experience. Key recommendations from the work have included:

- Submissions of e-journals should adhere to a limited set of formats because on the one hand, an adequate number of technical formats and mechanisms need to be supported so as not to pose an unreasonable barrier to deposit; but on the other hand, additional formats and mechanisms increase the cost of supporting legal deposit, inhibit interoperation, and increase the risk of loss of access to deposited material if obscure formats become inaccessible;
- Creation of a Technical Advisory panel. This Panel will need to be permanent because the technical standards for e-journal transmission, deposit and preservation will evolve over time and will need constant review.

Options for preservation will be selected on a case-by-case basis based on the balance between the up-front cost of normalisation versus the 'wait and see' approach of migration on request when future standards clearly emerge. Each individual project will have its own Digital Preservation Strategy based on a balance of benefits versus cost. An example of this approach has been the adoption of the JPEG2000 standard for digitised images. Originally the older TIFF standard was to be used, but it was discovered that the file storage requirements would have been at least doubled under that option. Understanding costs and benefits of services and for digital preservation has been a major area of work for the library in recent years. This has included the LIFE project which has influenced a number of preservation cost studies in the UK and elsewhere (Davies 2008) and a study on economic impact of the library's services. The British Library is also the lead partner for the European Union PLANETS digital preservation project.

The British Library, jointly with the US Library of Congress, announced in April 2006 that it supports the migration of electronic content to the National Library of Medicine (NLM) DTD standard, where practicable and is promoting its use for e-journal archiving.

An e-journal mapping study was commissioned by the British Library to inform the work of the Legal Deposit Advisory Panel in relation to the voluntary deposit of electronic journals (Rightscom 2008). It identified 5945 journal titles as falling within the definition of e-journals

published in the UK. Of these titles 3553 were published by only 8 publishers, with the balance published by 720 publishers. Of the 5945 titles, 530 were published by publishers with only a single title. It was noted that this polarisation in the size of businesses in to two classes of publisher (the very large and the very small) with substantially different characteristics, would have a substantial impact on the processes around the collection and preservation of their publications.

The study also found with metadata file formats, there was a polarisation between XML headers on the one hand and no metadata on the other, or metadata supplied via hosting services. Main content formats also showed very little use of SGML, with the majority of titles in PDF and XML full text, but a significant minority in PDF only and some in HTML only. Supplementary files are in a wide variety of formats, including spreadsheets, Word files, images, and various audio and video formats. Where DTDs were used, the majority of titles use proprietary DTDs internally, with a significant minority using the NLM DTD. For external delivery, NLM Publishing is the most used, with some NLM Archiving and some proprietary DTDs for delivery to some platforms, both publisher's own and third parties.

In terms of future trends, there was a consensus among the publishers on two main points: there is a generally-welcomed move to standardise DTDs on NLM or near-NLM DTDs; and there will be a very substantial expansion in the number and type of file formats used in the future, with much more multimedia in use, for example.

Interviewees Suggestions for the Strategy: *be very aware of the complexity around ingest of e-journals and challenges involved. Perhaps consider if there could be a role for Aggregators with established workflows with publishers to act as a service provider on the processing and dealing with long-tail of publishers. British Library has developed its own implementation for its peer network and preservation systems. It would be willing to discuss future collaboration as a strategy for Germany takes shape.*

4.3.3 The KB e-Depot

Key Findings: *content and licensing: A "dim" archive. Currently no support for post-cancellation access for licensees. It is aiming for coverage of all e-journals from the 20 largest commercial publishers. Also acting as a preservation service for content in 13 Dutch*

institutional repositories. Very large digitisation programmes at KB and digital masters archived; technical: Its DIAS system developed with IBM also used as core of the Kopal system in Germany. Contract with IBM expires in 2012 and future options being evaluated. Extensive R&D in digital preservation; business and organisational: Funded by Dutch Government but also exploring other funding models for additional funding streams in future. Current cost of e-Depot and digital preservation R&D about 4 million euro per annum. Concept of national (deposit) and international collections within the e-Depot. Bulk deposits covering both from e-journal publishers. Active in developing international and national collaborations and developing concept of a "safe places network".

The Koninklijke Bibliotheek (KB) is the national library of the Netherlands and operates e-Depot, its archive for the Dutch national deposit collection of electronic publications and other e-content (e.g. Dutch websites or master image files from KB digitisation projects). The e-Depot was established in 2003 and focused initially on Dutch material. Recognising the international nature of journal publishing (approximating 50-60% of all Science Technology and Medical publishing is based in the Netherlands) this has now been extended to international e-journals from the major publishers.

The KB intends to conclude archiving agreements for all the journals from 20 of the world's largest publishers and around 80% of open access publishers. Publishers wishing to make use of the services provided by e-Depot are required to conclude an archiving agreement with the KB and to deliver bulk content and specified metadata. The primary deposit file format is PDF.

The e-Depot could be considered a "dim archive". Generally, end-user access is restricted to on-site perusal at the KB for reasons of private research only and on-line access is denied. The archiving agreement specifies a number of trigger events when wider access would be permitted. Further work on agreed definitions of these trigger events is being undertaken this year. The e-Depot does not currently provide for post-cancellation continuing access by licensees of the content. Note however, full online access is already granted to publications by Open Access publishers such as Biomed Central.

The e-Depot also crawls and archives content from the 13 Dutch universities' institutional repositories; ensures sustainable archiving of the publications collected by NEREUS, the

partnership of European libraries in the field of economics in the context of their project NEEO (Network of European Economists Online); and archives the Directory of Open Access Journals (DOAJ).

A pilot project was undertaken to assess whether the e-Depot could be used for preservation of master image files from digitisation projects in all Dutch cultural institutions. This was not implemented as a service. The KB remains concerned at the storage required for digital image masters and associated costs given very low levels of likely future use.

When setting up the e-Depot in 2003, the Koninklijke Bibliotheek/National Library of the Netherlands (KB) created two separate organisational units: the e-Depot Department which is responsible for the operational management of the repository, and a separate Digital Preservation Department to carry out research and development. These research tasks include technology watch, (inter)national collaboration and the study of all preservation aspects in relation to the e-Depot, such as metadata, file format research, ingest control and the development of strategies for permanent access.

File format research and characterisation of digital objects are important areas of R&D. In addition, the KB develops strategies to render objects authentically in the future. At this point in time two strategies are seen as relevant: migration and emulation. Both strategies are being developed at the KB and in the context of the European project PLANETS. In 2007 a first modular emulator, Dioscuri, was developed in close collaboration with the Nationaal Archief of the Netherlands.

The e-Depot is an intrinsic part of the Dutch National Library, and therefore the Dutch Government is the major funder of both the e-Depot and the R&D efforts for digital preservation at the KB. However in the coming years the KB intends to develop a sustainable business model for the international e-Depot which will reflect both public and private responsibility for digital scholarly and cultural heritage. A number of discussions and projects are planned to advance this. Current cost of e-Depot and digital preservation R&D combined is around 4 million euro.

The KB has developed a specific workflow for archiving electronic publications. Elements of this workflow are: accept and pre-process; generate and resolve identifiers; search and retrieve publications; and identify, authenticate and authorise users. The technical heart of

the e-Depot system is IBM's DIAS (Digital Information and Archiving System). The DIAS solution conforms to the ISO Reference Open Archival Information System (OAIS) standard and supports physical and logical digital preservation.

The DIAS system also forms the core of the Kopal system used by the DNB and SUB Göttingen in Germany.

The contract with IBM for DIAS ends in 2012. KB is currently exploring options for the future with a larger group of 7 other national libraries: Czech Republic, Germany (DNB), Norway Portugal, Spain, Switzerland, and UK (British Library).

In the European arena a group of research institutes, national libraries and international publishers, coordinated by the KB, have joined together to establish the European Alliance for Permanent Access. Worldwide the KB also promotes the concept of a global "Safe Places Network". The Network is intended to include a limited number of institutions with certified digital archives which collaborate to ensure that information resources published by international publishers are permanently archived and continue to be available to future generations. The essence of this collaborative effort will be to share the responsibility for complete, world-wide coverage.

Interviewee's Suggestions for the Strategy: *Be aware of challenges and costs of ingest for e-journals. There is a need for the preservation community and publishers to define and develop generic ingest interfaces and then customised ingest interfaces for specific publishers and standards to underpin this. More could be done collaboratively between e-journal archives internationally. They believe current system of archiving internationally involves several different archives ingesting the same content from publishers and duplicating expensive processing. We should be aiming to collaborate more. Note may also see major new technical and business developments for the e-Depot over the next 2-3 years. The KB is already working with DNB and SUB Göttingen over Kopal. It would welcome ongoing engagement and discussion on collaboration with German partners as a strategy for Germany takes shape.*

4.4. North America

4.4.1 CLOCKSS

Key findings: *content & licensing: 14 publishers participated in the original trial and continue to work with CLOCKSS; the number of titles currently archived isn't clear: technical: a dark archive based on mature LOCKSS technology hosted by 11 libraries in 6 countries: business and organisational: a relatively new, not-for-profit dark archive service, but with a completely independent (from LOCKSS) organisational structure. The fact that it is a dark archive has encouraged large publishers such as Elsevier to sign up. Allowing access to the archive following a trigger event requires approval from a CLOCKSS board consisting of representatives from publishers and libraries. It aims to be a low cost operation and to reduce dependence on income from publisher and library annual contributions over the next few years by establishing an endowment.*

CLOCKSS (Controlled LOCKSS) was launched in 2006 and moved last year from a trial to a full service. A not-for-profit collaboration between libraries and publishers, it is a dark archive based on the LOCKSS software (see section below on LOCKSS) in which a limited number of libraries take on an archival role on behalf of a broader community. They describe it as a "private LOCKSS network". CLOCKSS is now a standalone not-for-profit company owned by librarians and publishers. The Archive ingests and saves either source or presentation files, as the publisher chooses. Allowing access to the archive requires approval from a CLOCKSS board consisting of representatives from publishers and libraries. When content is not available from any publisher, the content is copied from the archive, and made freely available on the Web (currently on servers at Stanford University and EDINA in the UK). There are 14 participating publishers. As a "dark archive" they have been able to recruit major publishers such as Elsevier, Taylor & Francis, Wiley-Blackwell. 11 libraries act as archive nodes in 6 different countries. CLOCKSS have had discussions with Humboldt Universität zu Berlin about their becoming the 12th Archive Node.

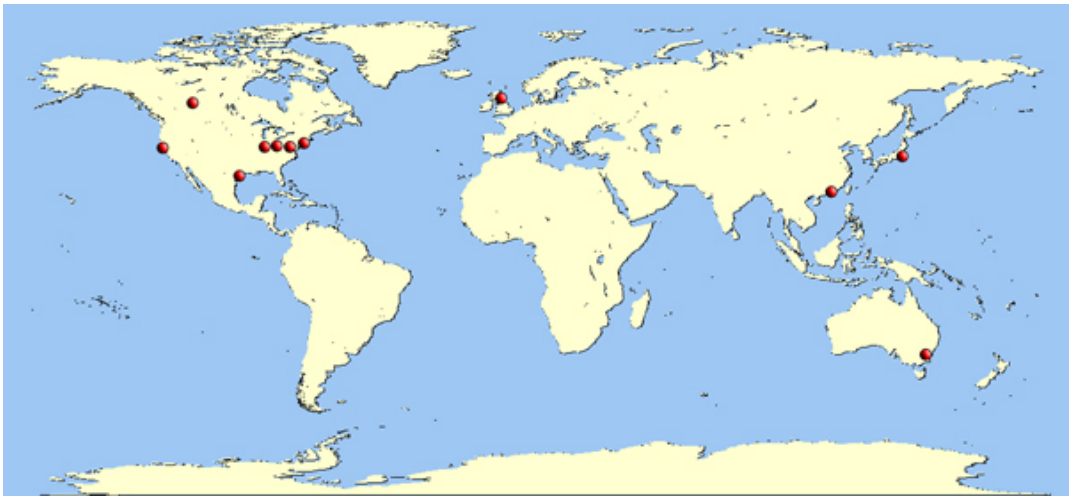


Figure: Locations of Current CLOCKSS Nodes (base map ©David Hall. CreativeCommons Attribution-Share Alike3.0)

Access to preserved content is only sanctioned when it is no longer available from any publisher; the Board of Directors then agree to making it freely available to everyone (not just previous subscribers) under a creative commons licence.

CLOCKSS is supported financially by publishers and libraries, but the Board is working to raise an endowment in order to make the service less dependent on subscription income. Supporting libraries are currently asked to contribute on a sliding scale of between $450 (315 Euros) and $15,000 (10,512 Euros) per annum depending on the size of their library materials budget.

CLOCKSS could be considered as a possible solution to the problem of long term preservation, but it doesn't provide post-cancellation access.

Interviewee's Suggestions for the Strategy: *the strong suggestion from the LOCKSS/CLOCKSS team is for a private LOCKSS network in Germany. CLOCKSS is, in effect, a private LOCKSS network.*

4.4.2 LOCKSS

Key findings: *content & licensing: LOCKSS now archives material from around 400 publishers (primarily small to medium-sized publishers); technical: a relatively mature open source software and community which uses a distributed model of digital preservation analogous to libraries collecting paper journal issues. It is a "dim archive" (meaning content*

can be revealed whenever required e.g. for post-cancellation access), has been in operation for 11 years, and has been self-sufficient for the last two. It is a storage system so any access services will need to be developed/integrated locally: business and organisational: A distinction is made between public and private LOCKSS networks; a private LOCKSS network has been recommended for the German libraries.

LOCKSS (Lots of Copies Keep Stuff Safe) enables libraries to collect, store, preserve and provide access to their own local copies of authorised subscription and open access content. The LOCKSS application is open source software. LOCKSS participating libraries harvest the web-published files of e-journals (e-books, etc.) from participating publishers. Access is triggered whenever (and for whatever reason) the material cannot be viewed on the publisher's (or intermediary's) servers. This includes, for example, short term network problems. The highly distributed nature of this approach aims to ensure that there is sufficient replication to safeguard content despite any potential disasters which might befall individual LOCKSS institutions. The LOCKSS infrastructure is being used to preserve a wide variety of materials including e-books, images, web sites etc. LOCKSS already has a presence in Germany with half a dozen sites with LOCKSS boxes and Prof. Michael Seadle of Humboldt Universität zu Berlin sitting on their LOCKSS Alliance Technical Policy Committee. The LOCKSS Program is eleven years old and has been self sufficient for over 2 years.

Users of LOCKSS have the option of joining the LOCKSS alliance which has a scale of fees, depending on the size and nature of the institution. U.S. academic library fees currently range from $1,080 (757 Euros) to $10,800 (7569 Euros) per annum. Alliance members "*are able to collect and preserve premium content not available to the general LOCKSS community. Participant fees support ongoing technical development as well as regular monitoring and tuning of LOCKSS Alliance boxes.*" In the UK a two year pilot has led to a LOCKSS UK Alliance with membership fees and a locally based support service. Current LOCKSS UK Alliance fees range from £1,800 (2,082 Euros) to £5,000 (5,782 Euros) plus VAT per annum. Charges and other information are shown on the JISC website: http://www.jisc-collections.ac.uk/catalogue/lockss

LOCKSS is currently being used to preserve content in two distinct types of environments. The public LOCKSS network holds and preserves material of general interest to a wide community. Approximately 400 publishers have given permission for their titles to be preserved by this service. Private LOCKSS networks hold and preserve materials of interest to targeted communities.

The greatest weakness of the open (public) LOCKSS system is that the bigger publishers (e.g. Elsevier) seem wary about the possible loss of control of their content implied by its distribution among many LOCKSS boxes, and so far haven't licensed their content to LOCKSS. But a "private LOCKSS network", building a National System using LOCKSS technology, may get around this problem. The Alliance for German Science Organisations could reach their own agreements with publishers regarding hosting their material on a German Private LOCKSS network.

Interviewee's Suggestions for the Strategy: *LOCKSS are recommending a private LOCKSS network (PLN) as the solution to preservation and continuing access in Germany. PLNs are self-governing and responsible for managing the legal, business and social infrastructure. Libraries running a Private LOCKSS Network own and control the content just as in the public LOCKSS Network. Challenges may include obtaining intellectual property rights to desired content and securing funds to build and maintain Hosting Services.*

4.4.3 HathiTrust

Key findings: *content and licensing: HathiTrust is a direct response to a need among libraries engaged in large-scale digitisation to ensure the preservation of their content over the long term; technical: Have developed a rights database for all items; business and organisational: an example of a highly co-operative and informal collaboration centred around a common need; funded by the 25 participating universities. The vision and drive of one or two people and a record of successful collaboration between universities was hugely important in its establishment.*

HathiTrust is a collaboration involving 25 universities, including the Committee on Institutional Cooperation (CIC) and the University of California system. HathiTrust was conceived to provide a repository, the fundament purpose of which is to preserve the output

of the participating universities' large-scale digitisation programmes (currently linked to the Google initiative) over the long-term, and that also allows shared access to the content. The repository contains over 3.5 million volumes (109 terabytes) as of 22 July 2009, none of which is licensed content. They hope to have 20 million volumes by 2012, and aspire to recruit more collaborative partners both nationally and internationally. They see themselves as one of numerous strategies to preserve digital scholarly content and emphasise the collaborative potential of these strategies rather than seeing them as being in competition with, or in opposition to, one another.

Ingest of content into the archives requires a 'Shared Digital Repository Digital Assets Agreement' . The depositor grants the repository rights of access for either 'perpetuity' or for five years. In return the repository will provide access services. If the repository is dissolved this license is revoked. The depositor must also sign a 'Permissions Agreement' to authorise copies of content, and which stipulates who owns copyright to the content, and HathiTrust maintains a rights database for all items. In order to provide persistent and high availability storage for deposited files their technology concentrates on creating a minimum of two synchronised versions of high-availability clustered storage with wide geographic separation, as well as an encrypted tape backup. Each separate storage system is equipped with mechanisms to provide mirrored management and access functionality, and of the storage or tape instances is physically secure and only accessible to specified personnel.

The informal and highly co-operative collaboration is funded by participating universities based on the amount of content they anticipate depositing over the next five years, although the University of Michigan underwrites much of the operating cost. To ensure some longevity to this voluntary endeavour partners commit to five years of participation. The incentives for participating in HathiTrust are the long-term preservation of digitised content, access to a core academic corpus that allows individual institutions to play to their strengths, and the provision of a more robust and stable infrastructure than any one institution could provide. The vision and drive of one or two people was hugely important in its establishment and success, as well as a successful record of collaboration between universities with similar needs.

Interviewee's Suggestions for the Strategy: *The key issues for a successful national strategy are collaboration and incentives, given potential competing interests. For example, an incentive for a university not wanting to participate because it wants to retain its uniqueness and specialisms is that it will have access to content they don't have.*

While national hosting will present significant technical issues these are not insurmountable, although the problems will be greater for hosting data than there are for books and journals.

4.4.4 Los Alamos National Laboratory Research Library

Key findings: *business and organisational: having been a pioneer in local hosting of commercial full-text content they have recently decided to stop local hosting having taken account of improvements in publishers' management of e-journals, new archiving options, resources and strategy.*

Los Alamos National Laboratory (LANL) is a premier national security research institution, delivering scientific and engineering solutions and working to realise a capabilities-based approach to science for the needs of U.S. national security. Their primary responsibility is ensuring the safety, security, and reliability of the nation's nuclear deterrent. As well as supporting the core security mission, their work advances several scientific disciplines. The Research Library offers extensive scientific and technical resources and services in support of LANL research, including 8,500 e-journal titles and 30Tb of stored content.

They were early adopters of local hosting of full text content and in many cases the agreements they secured for locally loading electronic content were the first agreements of their kind for the publishers. However, local hosting required a significant commitment of technical resources to keep the system running smoothly, and following a review of the market and their own strategic priorities they have decided that they cannot service their customers, sustain a huge and constantly growing set of commercial data, and match or exceed what commercial/national efforts are already doing better. At the end of July 2009 they are retiring their full-text repository. In the future their technical resources will be focused on dataset curation and on building new kinds of information services.

Interviewee's Suggestions for the Strategy: *None offered.*

4.4.5 OhioLINK

Key findings: *content and licensing: Prime example of negotiating consortial licences for commercial content that allow local hosting and ownership of content so it can be indexed and preserved in their own way. Although payment for a title may come from selected institutions, the licence is for all partners; business and organisational: an example of a state-wide highly co-operative collaborative network of 88 college and university libraries, underpinned by state legislation. The amount paid by each library varies depending on level of participation based on full-time equivalent (f.t.e.) student count, apart from a small membership fee.*

OhioLINK (the Ohio Library and Information Network) is a consortium of 88 Ohio college and university libraries, and the State Library of Ohio. Formed over twenty years ago as a consortium under state legislation initially to provide a state-wide electronic catalogue system, it launched the Electronic Journal Center in 1998. This is a collection of 8,777 full-text research journals, and more than 6.5 million articles (8 terabytes). Together with remote vendors, the EJC provides member libraries access to 12,000 electronic journals. OhioLINK is a prime example of successful consortia negotiations based on a standard licence allowing local hosting and ownership of content, permitting them to index and preserve content in their own way. They collaborate with OCLC OhioNet which is more likely to negotiate deals for smaller specialist titles, usually accessed remotely. Continuing access, for whatever reason, is achieved by local loading and ownership of content, as permitted by the licensing agreement. The loaded, accessible version of a journal is the primary copy, and this is backed-up; there is no additional dark archive.

In addition to electronic journals, OhioLINK provides access to more than 100 electronic research databases, primarily by remote access, and has a growing collection of some 40,000 e-books. OhioLINK also have their own repository – the Digital Resource Commons + Digital Media Centre - providing a robust, state-wide platform for sharing instructional, research, historic and creative materials which remain branded to the local institution.

The collaboration is highly co-operative, and service and content are paid for by participating libraries and OhioLINK (through its State funding). All institutions benefit from deals with

publishers, even if they are not contributing to the payment. A key licensing issue is that although payment for a title may come from selected institutions, the licence is for all partners and smaller institutions should be able to use the content free of charge. Such is the history of collaboration and co-operation, and the knowledge that good deals are being negotiated, that this does not create any tension. The amount paid by each library varies depending on level of participation based on full time equivalent student count, apart from a small membership fee. Price can be a 'deal breaker' in licence negotiations, and they aim to negotiate deep discounts on print versions of titles as well, although fewer libraries are now taking print. The current economic climate has impacted on budgets, and many of the members finance electronic content by cancelling print subscriptions.

> "... OhioLINK has more than quadrupled Ohio higher education's journal buying power. The $26.7 million invested in OhioLINK Electronic Journal Center statewide licenses ... would have cost at least $120 million if purchased individually just by our universities and a much larger amount if applied across all smaller colleges" 2008 Annual Report

Interviewee's Suggestions for the Strategy: *be aware that licensing issues are more complex at a national level than at a state level.*

4.4.6 Portico

Key findings: *content & licensing: as of 04/08/2009 the Portico website states there are currently 75 e-journal publishers and 10,324 e-journal titles committed of which 9,077 [88%] have a post cancellation access agreement; 7,502 e-journal titles are currently loaded in the archive. They also have 6,488 committed e-book titles, all of which have post-cancellation access agreements; only a small fraction of the e-books are currently loaded. 491 libraries participate in Portico, 157 of these outside the USA: business and organisational: a third party, not-for-profit, subscription service designed primarily as a dark archive but also provide option for publishers to nominate Portico as one mechanism for post-cancellation access. Portico is part of Ithaka, who are also responsible for JSTOR. Portico specifically recommend that the long term archiving function be based on existing services and should*

be separated from the issue of continuing access to previously subscribed material. Portico can provide authenticated access to triggered material if and when required.

Portico, launched in 2005, is designed specifically as a third party service for scholarly literature published in electronic form, including e-journals, e-books and digitised historical collections. Portico is a part of Ithaka (which is also responsible for JSTOR). It provides insurance to libraries that the e-journal and other content they have subscribed to will be preserved for the long-term. Portico only provides access to the e-journals they have preserved after specified 'trigger events'. Portico preserves publisher source and presentation files and delivers content in renditions appropriate to the current technology. Portico becomes a delivery mechanism in the event of a trigger event. In addition, if a publisher has designated Portico as such, it can also serve as a potential mechanism for post cancellation access. They anticipate significant growth in the preservation of e-books and large historic digital collections (e.g. newspapers).

Portico is funded by subscriptions from libraries and publishers. The annual subscription for a library is approximately 1% of a library's materials budget with 5% discount on the price for single institution for consortia.

Some argue that Portico's title selection and governance has a somewhat US-centric bias. It should be noted that, following a trigger event, Portico only makes preserved material available to the subscriber.

Interviewee's Suggestions for the Strategy: *Portico's recommendation is that, where possible, a National Strategy for Electronic Resources rely on existing preservation services, and that preservation and access systems be separated to ensure data security. A national approach to hosting electronic scholarly materials would, in Portico's view, be useful only if existing, well-established access and preservation systems relied upon by researchers from around the world are deemed inadequate to support the research community or other needs unique to Germany. If reliance upon existing systems is possible, this may enable access and preservation needs to be addressed at substantially lower cost than a stand-alone national solution. Portico believes that it has the experience and mature technologies necessary to address the preservation aspects of the requirements referred to in the study.*

4.4.7 University of Toronto and the Scholar's Portal

Key Findings: *business and organisational: maintaining its strategy of local hosting to serve libraries in Ontario Province.*

University of Toronto was not a formal interviewee for this study but was approached by email following the LANL and Ohiolink interviews as it is a well-known proponent of local hosting. It confirmed it is still very much a believer in local loading and are now adding e-books to go with the e-journals and e-abstracts & indexes, and data files, maps, institutional repositories, government documents etc. to get closer to offering a complete package. What was once a University of Toronto initiative has now evolved into a portal called ScholarsPortal (http://www.scholarsportal.info/) that serves the 21 universities in the province of Ontario. It believes that this local loading will continue and actually grow.

4.5. Conclusions

The international interviews have identified a number of interesting parallels and potential models for a Federated Strategy on Perpetual Access and Hosting for Germany. We have divided these into three sections: lessons that could be learnt from the international interviews; future trends; and a look ahead identifying possible best practice or opportunities that might be relevant in a German context.

4.5.1 Lessons that could be learnt

The Access-Preservation Spectrum of Services

For e-journals, an emerging service spectrum can be seen between services who are focussed on current access (hosting) services (LANL, OhioLink, CSIRO, University of Toronto) but are rather less focussed on long-term archiving, and those whose primary purpose is to focus on the long-term archiving and digital preservation (dark or dim archives) with publishers servers seen as the primary access for current and sometimes post-cancellation access (CLOCKSS, LOCKSS, Portico). Other dim archives are the national libraries (BL, KB, NLNZ) with a strong long-term archiving and digital preservation remit but with constrained current access services. The national libraries may also have a major focus on their national publications and a secondary but frequently significant focus on publications from elsewhere. JISC Collections is a national purchasing consortium: its economic models

allow for a range of options depending on the nature of the content and publisher for primary access (hosting) to be via the publisher's servers or a JISC supported Data Centre. These services point to the strategic choices that need to be made in terms of intended roles in current access hosting and long-term archiving that will be raised by a Federated Strategy on Perpetual Access and Hosting and the implications that will follow.

Intensive Processing of e-journal articles at Acquisition or Later on Request

Another important divide and set of strategic choices can be seen between some services that invest heavily in up-front but expensive normalisation and validation of formats and metadata at ingest/acquisition (e.g. BL, KB, Portico) and those that take the content "as is" from the publisher but would deal with any preservation actions on request at a later date (e.g. CLOCKSS, LOCKSS). Again these choices will need to be made within a Federated Strategy on Perpetual Access and Hosting for Germany and will affect the costs and also the nature of service(s) offered.

Preservation of Outputs from Retro-Digitisation

For retro-digitisation, we can see the local solutions for preservation of digital image masters implemented by major institutions such as the BL, KB, and NLNZ. A wider network of partners engaged in mass digitisation, collaborate within HathiTrust to preserve their digital image masters: though its services are largely based on University of Michigan. Services for preserving digital image masters appear less complex than for preservation and access of e-journals. Perhaps this is because mass digitisation is occurring in institutions with large-scale storage infrastructure and digital image masters are normally generated in-house and under the control of the same institution. Digitisation and archiving in smaller institutions are not really covered by our interviews but it is interesting to note that the pilot at the KB on archiving digital masters from other institutions in the Netherlands could not be supported beyond the pilot stage.

4.5.2 Future Trends?

A Trend away from Local Hosting?

A significant finding from the interviews was that a number of reasons including cost, improvements to publishers' services, ejournal archiving services, and network connections seem to be behind the imminent withdrawal from current local hosting services of e-journals

by some prominent former proponents such as LANL and CSIRO. However the picture is complex. Although examples of local hosting of commercial e-publications at scale seem to be becoming rarer, OhioLink and the University of Toronto (the Scholar's Portal) remain committed to local hosting and are examples of successful regional hosting services.

Standards

Interviewees pointed to a number of standards including JPEG 2000 for digital images and the National Library of Medicine (NLM) Document Type Definition (DTD) for e-journal articles in XML as of emerging importance.

4.5.3 A Look Ahead

Organisational Models

Both centralised and de-centralised and a number of hybrid models for organisation and delivery of services can be seen in the interviews. We would highlight some political and organisational similarities between the UK and Germany and perhaps the example of the legal deposit library network in the UK as an interesting model. Several interviewees (CLOCKSS/LOCKSS and NLNZ) also suggested that a peer network might suit the political and organisational framework in Germany.

Potential Services and Collaborations

As well as being potential models, several interviewees could contribute as potential services. Portico is an e-journal (and increasingly an e-book and digitised historical print collection) archiving and continuing access service based in the USA but open to international subscribers including those in Germany. Portico could therefore be evaluated as a service provider within the Strategy for Germany. Similarly CLOCKSS, although based in the USA, has several archiving nodes internationally (including one under discussion in Germany). It is focussed solely on long-term archiving without continuing access but could be a potential service provider in the Strategy. LOCKSS as a peer to peer archiving technology, LOCKSS Alliances for its support, and its implementation as a private peer network might also contribute.

As well as the direct service providers, many of the interviewees are also potential partners for German institutions if the strategy incorporates international as well as national collaboration in its framework. Existing international partnerships involving German partners such as the Knowledge Exchange and the Alliance for Permanent Access could also potentially facilitate international collaboration as part of the Strategy.

Best practice and other observations that could be relevant in the context of a Federated Strategy on Perpetual Access and Hosting for Germany

The description and analysis of a selection of international services and organisations above is intended to provide an international perspective and potential "benchmarks" for development of a Federated Strategy on Perpetual Access and Hosting for Germany.

Germany will have its own unique requirements and circumstances (as covered in the next sections 5 and 6). However the challenges and opportunities faced within Germany and by German organisations are likely to be shared with many of the countries and organisations interviewed in this section. Scholarly publishing of e-journals and even retro-digitisation (impact of Google Books project) are increasingly global in scale. Similarly, trends in information technology and business models for information are international and impact on many countries. We therefore recommend that the Alliance considers an international dimension to the Strategy that evaluates potential international partnerships and service providers and maintains oversight of emerging best practice and trends internationally.

Recommendation 1: Maintain an international dimension to the Strategy, evaluate potential international partnerships and service providers, and maintain an oversight of emerging best practice and trends internationally.

A summary table of best practice and other observations from the interviews that could be relevant in the context of a Federated Strategy on Perpetual Access and Hosting for Germany is presented below. This table together with the draft text of this section on international initiatives was shared with a cross-section of our German interviewees as part

of a workshop in Frankfurt on 25 September. The best practices and observations in the table were discussed with attendees at the workshop. Their views on priorities and importance of individual items in the table to a strategy for Germany were sought. Items given highest priority/seen as of highest significance for the strategy by attendees are highlighted in **bold** type in the table.

	International best practice and observations that potentially could apply to a Federated Strategy on Perpetual Access and Hosting for Germany
Content and Licensing	• Inclusion of TRANSFER Code of Practice and requirement for this in licences to address orphaning of journal access by title transfers. • **Use of Model Licences with standard wording for National Licences including preservation, hosting, and preservation clauses.** • Technical Guidelines and Requirements to accompany licences. • HathiTrust agreements for shared digitisation project output preservation service. • Archive selection criteria of top 20 publishers at KB. • **Post-cancellation access features of LOCKSS and Portico (88%).**
Technical	• Use of Peer to Peer Networks. • **Portico/CLOCKSS/LOCKSS/KB as archive and in some cases perpetual access (post-cancellation) options.** • NLM Journal Archiving and Interchange Tag Suite DTD/schema. • JPEG 2000 and TIFF as image archive formats. • The big publisher/small publisher divide and different ingest challenges at scale and heterogeneity. Aggregator experience of this. • Developing shared standards and ingest interfaces and customise. • **Use of publishers' platforms or regional services for local hosting (use of publishers' platforms more prevalent).**
Business and Organisational	• **Network of preservation and access nodes in UK deposit libraries**. • Cost/benefit studies in UK and making case for investment. • The time needed to build up the long-term trusted collaborations covered in interviews. The need for leadership to establish and grow. • LOCKSS Alliance/UK LOCKSS Alliance to support open source. • **Potential relative costs of hosting services and archives at different scales and functional requirements.** • **JISC Collections financial strategy + economic models**. • **Emerging nature of e-journal archive services and rapid change in content coverage and range of services offered**.

5. The Current Position in Germany

5.1. Introduction

In parallel with the international interviews, desk research and interviews were used to complete a survey of current hosting services and projects and future requirements in Germany. This was structured to capture and analyse current information on the status quo and future requirements within a Federated Strategy on Perpetual Access and Hosting. We used the client's suggestions to select 28 German organisations for interview and a structured questionnaire to guide the interview process. Some organisations were interviewed together where appropriate.

Detailed summary information from interviews for individual organisations is provided in a confidential appendix. Information on perpetual access and hosting is presented in section 8. A synthesis of other outcomes of the desk research and interviews is presented in this section focussing on:

- overview of the German context;
- organisational landscape;
- licensing of commercial content;
- specific types of content (commercial e-journals, retro-digitisation, and databases and e-books).

5.2. Overview

Germany, with 82 million inhabitants has the largest population among the member states of the European Union. The Federated Strategy on Perpetual Access and Hosting therefore is addressing the needs of the largest nation state in the European Union, with all the complexity and significance that involves.

Germany is a federal parliamentary republic of sixteen states (Länder). The German political system operates under a framework laid out in the 1949 constitutional document known as the Grundgesetz (Basic Law). This political framework provides for the national federation and the states to work within a mutual checks-and-balance system but also to practice mutual co-operation and consideration with the individual state governments participating

directly in the decisions of the national state or federation. Legislative power is divided between the national federation and the state level.

Federal legislative power is vested in the parliament consisting of the Bundestag (Federal Diet) elected through direct elections; and the Bundesrat (Federal Council), whose members represent the governments of the sixteen federal states and are members of the state cabinets.

The Basic Law presumes that all legislative power remains at the state level unless otherwise designated by the Basic Law itself. The states therefore have major activities and funding roles within Germany. For example, the responsibility for educational oversight in Germany lies primarily with the state governments.

Besides the responsibility of the state governments, the national government's responsibility for research and education is executed by its Ministry for Education and Research (Bundesministerium für Bildung und Forschung, BMBF).

The Federal government represented by BMBF, and the state governments may also agree to jointly fund some national bodies. Examples include the Deutsche Forschungsgemeinschaft (DFG) a central, self-governing research funding organisation that promotes research at universities and other publicly financed research institutions in Germany.

The library and information sector reflects this landscape of national and state institutions and collaborative organisations and initiatives.

This multi-faceted political, legislative, and funding landscape with its system of checks-and-balances, independence and collaboration provides the framework within which the Federated Strategy on Perpetual Access and Hosting for Germany must operate. There are already some examples of successful collaborations on which the Strategy can build. All interviewees within Germany spoke very favourably of the DFG national licensing initiative and the benefits this has brought for researchers and citizens within Germany. Similarly the existing participation and collaboration of the nine major associations and organisations in the Alliance of German Science Organisations is seen as a very positive development.

5.3. The Organisational Landscape in Germany

Our interviews for the study involved 28 organisations and were intended to provide a representative selection of the major organisations, associations and content likely to be relevant to the study.

In addition to the major political and funding divisions, the national and state governments mentioned above, the following were critical groupings and organisations covered by the interviews and desk research:

5.3.1 Research Funding Bodies (BMBF and DFG)

BMBF supports basic research and the respective organisations working in this field (in conjunction with the states); promotes key technologies; and promotes preventive research in the fields of the environment, climate, energy, ecology and health, the promotion of marine and polar research, research and development to improve working conditions, research into education and training, and research in the field of the humanities and the social sciences.

In addition, BMBF provides and administers funding for its e-science initiative that aims to "develop virtual knowledge environments in which users can access comprehensive databases, visualisations, and scientific information of all kinds in a dynamic manner" [6]. This is the main focus of the D-Grid [7] consortium. However, this has not resulted in a significant integration between D-Grid projects and infrastructures with the scientific information publishing, archiving, and digital libraries sphere. There are, however, some projects relevant in themselves, as well as some that contain their own relevant components.

Finally, BMBF has recently signed an agreement called "Pact for Research and Innovation" with the major German research organisations and has increased their overall funding budget by an annual 5% for the years 2011 till 2015. The ministry expects the research organisations to deal independently with the necessary issues related to a scientific information infrastructure from this funding and to suggest possible solutions and strategies if more is required.

DFG is a central, self-governing research funding organisation and a mechanism for the national government and 16 state governments to collaborate and provide funding for science and research. DFG funds research, advises government, and fosters relations

between academic and commercial sectors. It promotes research at universities and other publicly financed research institutions in Germany and as part of this it has funded to date 6 digital preservation research projects and studies. Beside research projects DFG also promotes projects and initiatives which serve to improve the conditions of the nationwide library network and information systems and these include the national licences for e-content, funding for retro-digitisation projects, and the Special Subject Collections scheme (all discussed further below).

As part of their action plan for 2015, electronic publishing and joined up repositories are high on their agenda. Since 2004 access to a significant number of databases and journal archives has been achieved through national licences with DFG funding. Since 2004 about 100 million euros have been spend by DFG for the funding of national licences. Its digitisation budget was around 12 million Euros in 2008 and over 5-10 years 25-50 million Euros have been invested in digitisation. The content digitised or licensed through DFG funds is accessed through/held by the libraries/institutions that obtained DFG funding and the governing statutes of DFG require that funding is project oriented i.e. it is typically for research projects or initial development work of 1-12 years duration. Furthermore, DFG funding must not be a substitute for the core funding for the information infrastructure through the universities and research institutions.

5.3.2 National Research Societies/Associations

280 research institutes are organised in 4 major German research societies: Fraunhofer Society, Helmholtz Association, Leibniz Association, and the Max Planck Society (MPG). Some of the 4 major research societies negotiate consortial licences centrally for their constituent institutes as well as the individual institutes negotiating for specialised unique content themselves at a local level. They participate in national or other licensing consortia when terms allow this. The Max Planck Society operates a central Digital Library (MPDL) and since 2003 has moved towards an e-only policy for licensing and since that date has put rights for local hosting into its licensing agreements with publishers. Individual MPG institutes also purchase licences individually. The Helmholtz Association and Fraunhofer Society also have systems for some central consortial procurement and also procurement locally by individual institutes. The Leibniz Association is more diverse but the three

specialist national libraries within its membership (TIB, ZB MED and ZBW) have formed the Leibniz Library Network for Research Information and have established a cooperative licensing office between them.

5.3.3 The German National Library

The National Library of Germany (DNB) was established in the early twentieth century and has the task of collecting, documenting and archiving all German and German-language publications from 1913 onwards. There is a legal deposit law for German publications and following a change in the law in 2006, this now covers all German publishers including open access publishers and retro-digitised material. The DNB hosts the office for Nestor (the German Network of expertise in Digital long-term preservation) and operates the Kopal digital preservation system for legal deposit publications with SUB Göttingen, GWDG and IBM.

5.3.4 National Special Subject Collections and Specialist Libraries.

A central funding programme was established in 1948 by DFG, called the “Special Subject Collection Field Program“, to overcome the problem that the libraries could not individually afford to collect all the scholarly literature universities needed. These special collections and libraries have either a disciplinary focus or specialise in specific countries, or regions and cultural areas. There are currently 121 collection fields that are completed every year and include everything that is published within these fields of interest, not only from Germany but worldwide, managed by around 75 libraries. They are obliged by the terms of their funding to collect relevant material and to provide it for interlibrary loan. This distributed national research library is formed from 23 state and university libraries, as well as three central libraries (TIB, ZB MED and ZBW), and some smaller special libraries for very specific disciplines.

5.3.5 State and Regional Libraries and the Libraries of Universities and other Research and Educational Institutions

There are around 25 regional deposit libraries and the federal states have individual deposit laws for their states. The state libraries can hold very large collections and two national competence centres for digitisation established since 1997 are based at the Bavarian State Library and the State and University Library of Göttingen respectively. In November 2009 Exlibris and the Bavarian State Library announced they have concluded a strategic

partnership on the use of Rosetta for digital long-term preservation. This solution could potentially be extended to the other academic libraries in the Bavarian Library Network. There are 104 universities, 189 universities of applied sciences, and 101 private colleges across Germany. The major public funding for universities higher education is via the state governments. Both state libraries and universities can fulfil roles at the national level as negotiators for DFG national licences or holders of national special subject collections. They also play major roles at regional and state level such as often hosting or leading regional purchasing consortia.

5.3.6 Regional Library Purchasing Consortia and Library Co-operatives (Bibliotheks-Verbünde)

Five major regional negotiating consortia with regional funding programmes and eight smaller ones negotiate consortial licences for electronic content on behalf of their members.

In addition, the German, Austrian and Swiss Consortia (GASCO) was started in 2000 as a pan-consortia consortium to achieve a better information flow between the participating regional consortia of Germany as well as the two national Consortia of Austria and Switzerland.

Finally, there are a number of major regional library co-operatives (Bibliotheks-Verbünde) and centres that provide shared library services to their members or local region and sometimes extend these to a national level.

5.3.7 Super-Computer Centres and Specialist Support Centres

There are close partnerships between several libraries and super computer centres at regional level or with a number of specialist information management or information supply centres at regional and national level (e.g. the Fachinformationszentrum, FIZ). They provide expertise and facilities for mass storage and bit preservation, information management, and hosting services for various libraries and others.

Recommendation 2: Wide consultation and further consensus building amongst all stakeholders including publishers will be a critical element of further development and implementation of the Strategy given the complex organisational landscape in Germany.

5.4. Overview of Licensing in Germany

Sylvia Weber, one of the negotiators for DFG national licences, has recently written an overview of the experience on national licensing in Germany over the past 5 years, which she has made available for the study (Weber 2009). This section is based on an abridged and edited version of her article with some additional material from the study interviews and desk research.

Licensing in Germany is complex and based on the way in which higher education and research is structured and many and varied historical factors which evolved for licensing print publications. These arrangements have moved over to the digital world where unique licences are negotiated for specialised subject content.

There is the potential for each individual organisation to have four different levels of licences for electronic content to deal with depending on their circumstances. These are namely:

- Local Licences;
- Consortial Licences;
- National Licences;
- Multinational Licences.

National licences for the whole of Germany are negotiated on behalf of DFG and 'multinational' licences as they are currently called are negotiated by a working group from the four European countries; Germany, UK, Denmark and the Netherlands through the "Knowledge Exchange Initiative".

Each university or research institution will have a different portfolio of licences e.g. the University Library of Frankfurt holds a portfolio of 193 licences that are split as follows:-

- 30.5% Local Licences
- 50.7% Consortial Licences
- 18.6% National Licences
- 0.5% Multinational Licences.

To add to the complexity of the situation for each type of license there are different types of contracts, different terms and conditions the organisation has to deal with.

Locally licensed content is funded from individual organisational budgets as will contributions to certain types of consortially licensed material. Consortial licences negotiated for regions are funded by the individual Regional States and/or participating libraries each year. Similarly consortial licences for the research societies such as MPG are funded by the society for all its members. National licences are funded by the Federal Government each year through DFG.

The collective acquisition budgets of all the German universities equates to about 230 million Euros per year, 5% of which is invested in digital media. Taking all the funding programmes together, Germany currently spends up to 40 million Euros per annum for digital scholarly content. It should be noted however that many German licences for journals are still based on historic print spending by the library, with publishers providing print copies and access to electronic versions and additional electronic titles for a small percentage fee on top of this. The e-spend may therefore appear artificially low because of this and lower than some other countries with more e-only journal licences.

National licences have a real challenge particularly to bring together the requirements of the various scholarly organisations with the policies and procedures of the national and international publishers. DFG funds both negotiation and review activities regarding national licences and has established strict policies and procedures for these activities. To establish a target for contracts and proposals, DFG publishes criteria for the National Licences that have to be agreed and signed off by the publishers, otherwise the proposals are rejected.

In addition to the negotiations and review activities DFG also funds two central services that are supplied by the Universities of Göttingen and Frankfurt for all the participating institutions:

- a central tool for the cooperative registration of participants and the conversion of metadata (www.nationallizenen.de) in Göttingen; and
- a server for user statistics that are collected from 54 publishers and that contains the usage of 531 participating libraries, developed and maintained in Frankfurt.

Two types of National Licences are funded by DFG:

a) **National Licences "Classics"**

In 2004 the backfile program ("The Classics") was the first funding program to be established. The funding covers databases, journal archives, completely digitised collections and e-books. The content reaches up to the first backfile year that is not covered by the current content licences, at least up to 2002. The licence has to be purchased with a one time payment and all access fees (at least for the next 10 years) shall be included. The backfiles are funded 100% by DFG and are supplied to the universities free of charge. Within the last five years DFG has purchased over 100 products covering all fields of science and the humanities.

b) **National Licences "Complete"**

After some years of funding the demands for current content became more urgent, so DFG started a pilot scheme for current content National Licences ("Complete"), in 2008 which runs for three years to 2010. A working group was established to define the licensing model and the funding scenario, as it was clear from the beginning, that current content licences would be too expensive for 100% funding. So far the complete journal collections of 12 publishers in greatest demand by libraries have been funded and licensed. The working group chose an opt-in model. As a consequence, access to the current collection is only open for those libraries who have paid a certain fee – but the older backfile years, the journals archives, are open to all authorised users of the National Licences. Between the current content and the archive a "moving wall" is defined: after the end of the current year, the content will be added to the archive and be opened to all German users, e.g. in 2008 the year 2007 will be added to the archive and in 2009 the year 2008 will be added. As a consequence, the participating libraries have access to the complete collection of the publisher from the first year of publication up to the current year.

The pricing model for these journal collections is based on subscriptions. That means libraries have to pay for their existing holdings and only a few cancellations are allowed within the three years term. The DFG instead pays for the additional access fees that are raised for the access to the complete collections of the publishers. For institutions without subscriptions a flat fee is negotiated, that is paid 100%by the DFG.

For the four main research societies a flat fee is negotiated too, that is paid 85% by the societies and 15% by DFG. In addition, DFG pays the one time fees for the journal archives as well. So all together, DFG's funding volume of this program is 18.3 million Euros for the years 2008 to 2010.

Recommendation 3: Develop a federated organisational and funding model to maintain and develop the national licensing initiative. Include perpetual access and hosting in its scope. Explore where savings achieved from licensing collaboratively might contribute to funds for shared perpetual access and hosting requirements.

5.5. Commercially Licensed Academic e-journals

5.5.1 Methodology

To provide a view on the current position for commercially licensing academic e-journals, the interview questionnaires were developed to capture information on current licensing and hosting (or preservation depending on the organisational mission) for commercially licensed e-journals. Given the very large number of journal titles available and also that many are bundled into licence packages by publishers, the questions focussed on which publishers were licensed rather than their individual titles.

A table of 52 publishers was prepared by the study team in consultation with the clients. Interviewees were asked to insert ticks against publishers whose content they licensed access to. If they had multiple roles e.g. for a regional library consortium and their institution, separate questionnaires were used for each.

Interviewees were also given the opportunity to suggest other additional publishers licensed/hosted whose archiving is important to their organisation.

There were also asked if they hosted any e-journals locally (either directly or via a service provider on their behalf). If hosted locally they were requested to add extent/completeness of coverage (range of years, issues etc) in free-text comments.

Similarly preservation and archiving institutions interviewed were asked to indicate which of the publishers they preserved e-journals for.

5.5.2 Results

Results for preservation and archiving institutions and the selected publishers they preserve e-journals for are presented in a table on page 97 and in a confidential appendix . Results for local hosting of e-journals from selected publishers are presented on pages 98-99.

The licensing results for individual institutions include e-journals licensed at all levels multi-national, national, regional and institutional were included in a confidential appendix to this report.

A number of observations can be made:

- Our core list of publishers appears approximately to cover the shared requirements of most of the interviewees. 52 out of 52 publishers in the core list are subscribed to by at least one of the interviewed Consortia, University, State, and Specialist Libraries, or selected research organisations. Only a few publishers might have been omitted from the original list. There are a few shared additional suggestions, mostly resulting from national licences which could be added in common. The regional consortia subscribe to 19 out of the 52 publishers listed but individual members of the consortia are likely to subscribe to many more via other sources or as individual institutions (as can been seen in the first table of interviewed University, State, and Specialist Libraries).
- There are also some substantial differences between these three groups particularly in terms of the importance of additional small publishers to them. It is notable that the Helmholtz institutes provided a list of 74 additional publishers whose archiving is important to them in addition to our core list of 52 publishers.
- If selecting or setting up services for perpetual access and local hosting informed by these tables, some selection will be required. This could be necessary even within the 52 publishers listed in the tables and certainly within any extended consideration of those listed combined with suggested additions by interviewees. The overall level of institutional subscription to different publishers may help in this process. Although arguably potentially all should be in the scope of the strategy different approaches and prioritisation may be required (i.e. different approaches to large publishers

subscribed to many different institutions within Germany; and smaller publishers who are important in very specific fields and important to specialist researchers);

- Only the DFG national licences and the MPDL licences currently seem to have the scale and market power to strongly influence the wording and terms for local hosting or perpetual access of licences with publishers. Most other interviewees noted they although publishers licences normally include clauses for perpetual access (often including supplying a copy to the licensee if a subscription ceased), they are highly variable and they have had little success in seeking to amend them during negotiations;
- There are some publishers currently who will not grant local hosting rights e.g. Nature Publishing Group and other special cases such as licences with aggregators where perpetual access clauses are not available. Exceptions and how they fit in the strategy will need to be considered;
- A final observation is that establishing demand for licensing and hence for perpetual access or hosting requirements in licences has been very difficult (both for the consultants and the interviewees) even in a relatively simple form. However this information and maintaining it is crucial for any national hosting strategy to be developed and implemented.

Recommendation 4: Develop management information for licensing of e-journals and requirements for perpetual access and hosting. This should facilitate analysis of the need and size of demand for federated action.

5.6. Retro- Digitisation Programmes

5.6.1 Methodology

Retro-digitisation was covered by interview questions on the current position for digitisation within their institutions. Interviewees were asked to provide information on major digitisation projects that they had or fund, their storage volumes, standards, metadata schemas, and file

formats used. Interviewees also had the opportunity to suggest relevant print or web resources for further desk research.

In addition for the future, they were asked what growth is estimated over the next five years in their storage figures and what they would want to see implemented in a Federated Strategy on Perpetual Access and Hosting for Germany.

It was noted that a major DFG study had been undertaken by the Bavarian State Library and the Armed Forces University on "Organisational and Business Models for the Long-Term Preservation of Digital Objects' focussing on digitisation projects and their long-term preservation. This is still in confidential draft but we have aimed to minimise duplication with this work. Its final report should be complementary to this study.

5.6.2 Results

It should be noted that overall retro-digitisation seemed to be an area of lower concern in terms of perpetual access and hosting for our interviewees. This lower level of concern seems to have two aspects in the interviews:

- where it is extremely important to research e.g. retro-digitisation of historic German literature, the nature of the control of the end-to-end processes within key digitisers reduces areas of concern substantially (i.e. they create material so can specify standards); also that the work done to date to develop standards and infrastructure to handle retro-digitisation within Germany is very advanced (the two national competence centres in digitisation were established in 1997 and have been in operation for over a decade) and there was a high level of confidence in its capacity;
- a lower importance for some disciplines particularly in science, technology, and medicine on the topic was reflected in the responses from some interviewees.

The only major areas of concern emerging from the interviews surrounded:

- sustainability of services (given most funding for digitisation is project-based); and secondly
- the position for smaller institutions whose infrastructure and staffing was inevitably more limited.

Some confirmation of these emerging concerns over sustainability and the position for small institutions comes from the desk research.

The issue of sustainability for outputs from retro-digitisation is a concern shared in all the major countries engaged in mass digitisation and efforts are being made to share emerging experience and best practice internationally. A recent study on the issue includes includes one example (DigiZeitschriften) from Germany (Maron et al 2009 p45-52).

A Nestor sponsored study of digitisation in German Museums identified concerns over infrastructure and staffing in small institutions to maintain outputs and services for them over the long-term (Witthaut et al 2005). Similar conclusions emerged from a recent DFG funded study of digitisation projects (BSB/Armed Forces Universities forthcoming).

A synthesis of the other main findings (excluding perpetual access and hosting) is presented as follows:

Creation

- The DFG have funded a comprehensive set of Digitisation Guidelines which were referenced and used by interviewees;
- Two national competence centres for digitisation are established at the State and University Library at Göttingen and the Bavarian State Library and have been in operation since 1997. These have substantial capacity for mass digitisation projects. Partnerships with nearby super-computer centres at GWDG and the Leibniz Super-Computer Centre respectively give access to relevant expertise and technical infrastructure for mass storage at currently marginal or no additional cost;
- The national competence centres, SUB Göttingen together with BSB, had a major role in writing the DFG Digitisation Guidelines so these reflect their local standards, metadata schema, and file formats;
- BSB is working with Google books (public private partnership) on mass digitisation of German books. The Munich Digitisation Centre has carried out more than 100 digitisation projects, funded by the DFG, the European Union and the Bavarian state. The current amount preserved is 134 Terabytes or 147,792,501 files (digital master

and access files). Expected future growth is 100 Terabytes plus per annum as the output from the Google partnership content starts to be included;

- The Göttingen Digitisation Centre currently has around 25 Tb of digitisation files and have purchased more storage for GWDG as part of project budgets. Over the next 5 years their storage requirements may grow to approx.150 TB depending on the archiving strategy and redundancy of objects stored. They are using TIFF uncompressed as the file format for storing master archive images and METS, with transfer into ZVDD metadata schema;
- At KOBV digitisation storage is 50 Tb of disk space currently and they anticipate 1-2 petabytes of storage will be required in the next 3 years as they are involved in some very large digitisation projects e.g. with Berlin State Library. They maintain online presentation files and a digitisation master archive file. There is a Hierarchical File Store with TIFF masters on tape and jpeg access files on fast disk access respectively;
- HBZ digitisation projects include: The University Library of Bonn - about 310,000 objects (digitised pages), 1.2 TB “Archival Master Files” (TIFF), 0.4 TB “Access Files” (JPG and PDF); Lippische Landesbibliothek (regional library) - about 35,000 objects (digitised pages), 0.3 TB “Archival Master Files” (TIFF), 0.1 TB “Access Files” (JPG and PDF); Lehrbucher (textbooks) – currently in planning stage. Planned scale is about 600,000 objects (digitised pages), 2.5 TB “Archival master Files” (TIFF), 1.2. TB “Access Files” (XML fulltext, JPG, and PDF). The aim is to to enable students to access the information in digital form.
- The Max Planck Digital Library together with the publisher “Zeitschrift für Naturforschung” is digitising the publisher’s journal backfiles of articles in three corpora published between 1946 and 2001, which amounts to approximately 30.000 articles and about 0.5 TB of data. MPDL plan to increase the number of retro-digitised collections, which will result in an increase of about 0.3 Tb annually for scientific articles. Due to an ongoing digitisation project of photographs they expect an increase in storage space for digitised photographs of about 5 Tb annually over the next five years.

- Major public funding for retro-digitisation has come from DFG. The DFG digitisation budget was around 12 million euros in 2008 and over 5-10 years 25-50 million euros have been invested in digitisation.

Acquisition

- Since 2007 DNB has collaborated with the Publisher Association and the collecting society (VG Wort) on a model that aimed to facilitate the digitisation of out-of-print and orphan works whilst respecting the relevant copyright protections. This work lead to the development of the ARROW project (Accessible registries of rights information and orphan works towards Europeana) which is a collaborative effort to identify copyright holders of out-of-print works, to create European registers of orphan works and also to develop models for integrated access to charged and free digital content.

Ingest

- The Mass Digitisation Centres have established major workflow systems for example the Bavarian State Library Digitisation Centre output is processed via a semi-automated workflow ("ZEND" framework for digitisation), which covers all processes from preparation, scanning, indexing, storage and preservation up to access and reuse of the archived digital master files under certain circumstances (e.g. document delivery of high resolution images).
- National legal deposit law for German electronic publications covers all German publishers including open access publishers and retro-digitised material.

Additional discussion on hosting and preservation of retro-digitisation outputs is included in section 8.

5.7. Databases and e-books

5.7.1 Methodology

The main focus of the current review was to consider a Federated Strategy on Perpetual Access and Hosting for electronic journals and outputs from retro-digitisation. However questions were included in the questionnaire to solicit views from interviewees on commercial databases and e-books they licensed and their future needs. It was recognised

all of these types of electronic resources are already being licensed (including at national level) or purchased outright and delivered to users but in many cases are less established than the e-journals market (i.e. e-books), or covering current and frequently updated reference material or secondary products in the case of databases (e.g. legal and medical databases; bibliographic and abstracts databases).

5.7.2 Results

Interviewees identified that databases cover a very broad field and that the future strategy would need to distinguish them. In simple terms a database is a collection of information in electronic format that is organised in a logical manner so that it can be easily accessed, managed and updated. They can be classified in various ways one being according to type of content: bibliographic, full-text, numeric, images or sound. Another way would be according to subject matter.

The interviewees referred to their acquisition under licence of current scientific, medical and business databases such as BIOSIS Previews, CAB Abstracts, Web of Science, SciFinder, INSPEC and Lexis-Nexis which are bibliographic secondary sources produced by some of the world's largest commercial publishers. They cover a wide time frame, extensive numbers of sources of various types and are generally now cross-linked to primary full-text sources such as journals and books to give the user a one-stop shop experience. Some such as SciFinder include the searching of chemical structures. Numeric/financial databases such as Frost and Sullivan and Amadeus are used in the business sector as are materials property databases such as Cindas LLC in the materials science/engineering areas. In general these types of databases are not purchased outright but are leased because of the constant need for updating of information, often on a daily basis. The commercial publishers are very protective of their proprietary rights with respect to these databases given the costs and human effort of producing them. Interviewees felt users would want access to the most current version of these resources from the publisher and perpetual access or local hosting of them would not be a priority or a concern for the strategy.

Collections of primary full-text sources such as letters, reports, books, newspapers, etc on a single theme are often called databases or online collections. The interviewees noted examples of these e.g. from the Alexander Street Press and CengageLearning/Gale. They

tend to be closed collections, time-limited and on a subject specific area. One of the DFG National Deals is an example of such a collection - 'Testaments to the Holocaust'. They have proprietary structures and packaging but are much simpler than e-journal collections. They are composed of scanned images which are in a particular sequence and there are links to and from the metadata. Interviewees noted these historic digitised collections have a more fixed and archival nature and may be a much higher priority for consideration in future versions of the strategy.

The main characteristic of databases versus journal collections is that the former contains a wide range of material of different types performing different functions for users. No one database is really like another and its structure and content is dependent on specific subject needs. A number of interviewees raised the issue of costs arising from this when considering local hosting. Databases illustrate the need to consider both the content and associated the features/services needed to access them in an electronic environment.

The interviewees mentioned examples of e-book collections that had been acquired and overall quantities licensed. Again interviewees stressed that the label “e-book” was one of convenience covering many different types of resource, some of which would have value for perpetual access, whilst others were works of reference constantly updated where the latest version on the publisher’s server would be required by users. Many interviewees also noted the emerging and uncertain nature of the e-books market and business models for it and felt it would be too early to develop a strategy for them. The need to consider open access content such as e-theses that might be considered under e-book formats was also raised.

A conclusion that can be drawn from these comments is that future developments of the Strategy may need to be guided by a “content map” which provides a mechanism for distinguishing different types of databases and e-books and places commercial content into a context alongside open access materials. This would allow the strategy to focus on those materials where perpetual access or local hosting (e.g. to provide specialised services not available from a publisher) are seen as priorities. It would also allow work conducted for the Federated Strategy on Perpetual Access and Hosting to be placed in context of related work on open access.

Recommendation 5: Future developments of the Strategy can be guided by a "content map" which provides a mechanism for distinguishing different types of databases and e-books and places commercial content into a context alongside open access materials.

5.8. Preservation and Hosting in Germany

5.8.1 Method

Interviewees were asked a series of questions on their long-term preservation or perpetual access and hosting agreements/arrangements for commercially licensed content and retro-digitised material. Interviewees indicated a range of projects and services and documentation that were included in the desk research.

In addition interviewees who were responsible for licensing content were asked their degree of confidence in current arrangements for perpetual access and its importance to their institutions.

5.8.2 Results

The Print Archive

It is important to recognise that the commercial e-journals market is still in transition with parallel print and electronic versions of journals. Equally we have parallel print and electronic archives and access services encountered in the interviews.

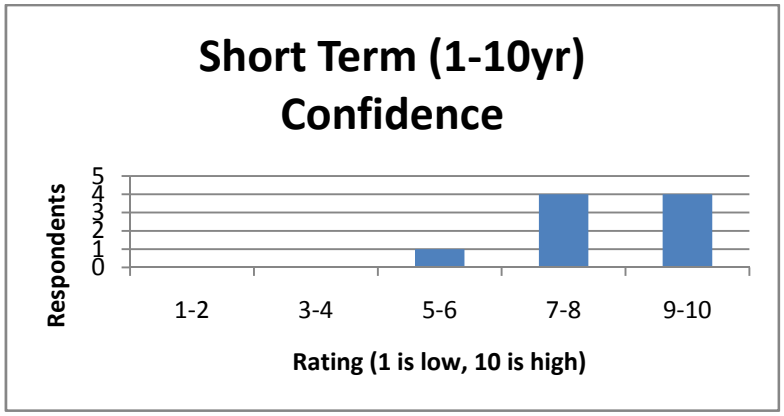
A significant number of interviewees stressed that they still rely on the existence of the print version of an e-journal in an archiving library (a library with a preservation remit) somewhere in Germany and its access via inter-library loan for their perpetual access to commercially licensed e-journals. The additional VAT cost of e-journals (not applied to print) and the uncertainty over electronic archives is inhibiting moves towards e-only policies. However many interviewees are moving in this direction and some notably the Max Planck Society are already e-only when licensing commercial content.

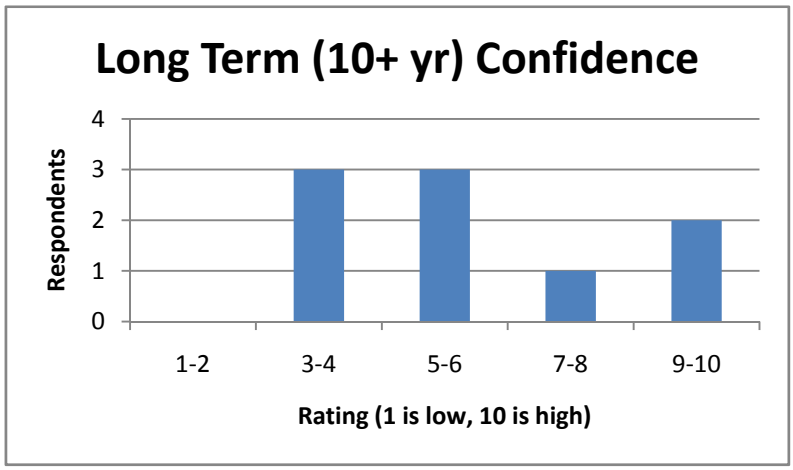
Some interviewees suggested that the existence of a print archive is perhaps a misleading comfort when contents of print and electronic versions begin to diverge and bundled titles in licence packages give access to far more titles in electronic form than are subscribed to in print.

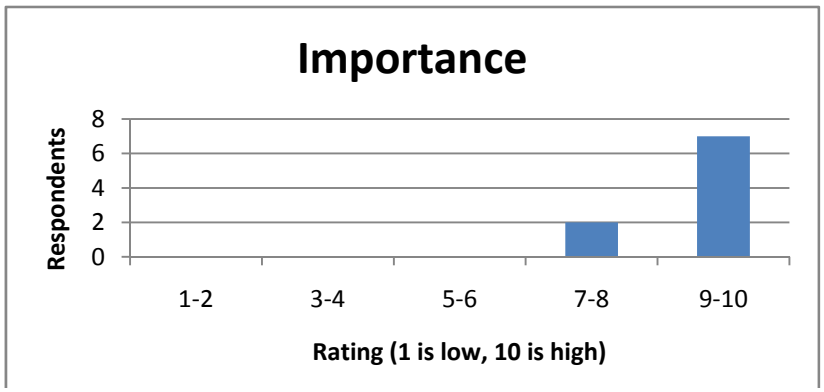
All however recognised the continuing growth of e-publication and desired to see the evolution of trustworthy arrangements for perpetual access.

Level of confidence in current arrangements for perpetual access for electronic resources and their importance to German Institutions interviewed

Interviewees who license content were asked to score on a scale of 1-10 (where 1 is low and 10 is high) their confidence in current arrangements for perpetual access and their importance to their institution. They also had the opportunity to comment on their scores in free-text. Their scores are shown in the figures below.







The majority were confident of arrangements for the short-term (next 1-10 years) with scores ranging from 8-10 in the responses. Free text comments outlined reasons for this as being variously the existence of a print copy, agreements for access in national licences for normally the next 10 years to the publishers servers, or confidence in local digital preservation expertise (for some of those hosting locally).

The position reversed when looking at the long-term, with the majority showing very low confidence in current arrangements for the long-term. This reflects both the uncertainty they felt over the publishing environment (transfer of titles between publishers, publisher mergers or other commercial changes), and the Current options for and evolving nature of digital preservation and sustainable services for the long-term.

These concerns and uncertainty over the future outlook are very significant when interviewees are also indicating that these arrangements are of very high importance to their institutions (scores between 8-10).

German and International Collections and Responsibilities

Although it was only remarked on explicitly by a few interviewees the source of material and responsibility for it was also implicit when considering preservation or hosting for either commercial content or retro-digitisation. What is published electronically in Germany and what German organisations have responsibility (or opportunity) to preserve and host permeated the interviews.

The German publishing landscape and the landscape for publishing science, medicine and technology (STM) internationally has undergone profound changes in recent years. There are no entirely German publishers today in the STM area. The smaller ones have been acquired and merged into the larger publishers. And the larger ones have all gone

international and have branches or affiliates internationally. For example Thieme is now active in the USA and has taken over Enke, Hippokrates, Parey and others. Springer has taken over Birkhäuser, Steinkopff, Kluwer, Physica, Urban & Vogel, Spektrum, and VDI. There are now 44 German publishers that Springer has acquired and more than 70 publishers in total they have acquired worldwide. There are publishers still publishing e-journals and other electronic publications within Germany but the vast majority of e-journals and a significant number of databases and e-books being licensed by German institutions are acquired from outside Germany itself.

The effect of this can be seen in the tables on page 97 and in the confidential appendix where very few of the e-journals listed in the table are covered by legal deposit in Germany.

The position differs with other types of content. In retro-digitisation we can see activity almost entirely within the control of German institutions and development of infrastructure for its hosting and preservation. Similarly there was a strong emphasis on open access materials in most of the institutions interviewed in Germany and hosting infrastructure and preservation activity for the long-term are emerging and were mentioned in the interviews.

6. Views on The Future Position in Germany

6.1. Introduction

The interviews within Germany used a structured questionnaire which aimed to capture data on the current position in Germany and in addition sought views via a series of open-ended questions on what the future position on hosting and perpetual access and the strategy for this should be. Responses from individual organisations interviewed were captured as free-text and the draft interview notes sent to the interviewees for corrections or additions. Their responses on the future are synthesised under "Interviewee's suggestions for the Strategy" in a confidential appendix. The following requirements are an analysis and synthesis of requirements and suggestions identified by German interviewees and additional input from the workshop discussions with interviewees. They have been grouped and mapped against our four core areas for the study analysis. It should be noted some aspects of this process are subjective. For example many interviewees expressed their requirements directly in the interview but interviewees also made observations on the strategy which indirectly imply a set of requirements which also need to be extracted to ensure the picture is complete. This was taken into account at a Workshop in Frankfurt with a cross-section of the interviewees on the 25th September 2009 when the requirements were further refined and prioritised, resulting in the table below.

6.2. Requirements and suggestions identified by German interviewees.

Relevant to areas:	**Content, Licensing, Preservation and Hosting**	**Technical Infrastructure & Standards**	**Business Models, Costs and Funding**	**Organisational Models and Policy**
Requirements				
A) Licensing and legislation				
HIGH PRORITY				
i) Raise awareness of what needs to be in licences to facilitate strategy. Differentiate archive and hosting. Develop national licences/ standardisation of licence terms/ models licences.	*			
ii) Access to journals "orphaned" by transfers between publishers.	*			
iii) Allow for materials to be in common, non-proprietary platform for long term access.	*	*		

Relevant to areas:	Content, Licensing, Preservation and Hosting	Technical Infrastructure & Standards	Business Models, Costs and Funding	Organisational Models and Policy
LOWER PRORITY				
iv) Identify material licensed under "national" DFG initiative and State consortia licensed material available nationally for subscribers and define policy on how both relate to the strategy.			*	
v) Include both licensed and non-licensed born digital material.	*			*
B) Control of trigger events				
HIGH PRORITY				
i) Allow for continuing access if subscription ceases	*	*	*	*
ii) Provide continuing access at short notice – have no time delays or uncertainty - independent of publisher veto	*	*		*
C) Working with publishers				
HIGH PRORITY				
i) Standardise metadata output	*	*		
ii)Supply of metadata for authorised applications	*	*		
LOWER PRORITY				
iii) Clear strategy to influence publishers views on continuing access [Include in contract or licence]	*		*	*
D) Coverage in Archive/Hosting Service				
HIGH PRORITY				
i) QA at ingest, ensuring no missing issues and no corrupt content	*	*		
ii) Include small specialist publishers as well as large [Include all disciplines]	*			*
E) Technical and Delivery				
HIGH PRORITY				
i) Define national technical standards (inc. metadata and files) and infrastructure, access mechanisms and authorised users. The working group expressed common understanding, that an infrastructure would be a distributed one, as service oriented architectures (SOA)	*	*		
ii) 24/7 availability or not dependent upon usage profile and cost.		*	*	
iii) Linking: to research data; from bibliographic links. Resolver required.		*		*
iv) Interoperability capability and accessible		*		
v) Open interfaces; open standards (not always open resources)		*		
vi) Clearly defined services / user driven		*		*
vii) More than one service centre		*		*
viii) Metadata data policy - full text from publishers to maintain common archive [Iteration of another requirement - Aiii]	*	*		
ix) Reliable DRM system	*	*		
x)Bit preservation' role separate from publishers' access versions		*		*

Relevant to areas:	Content, Licensing, Preservation and Hosting	Technical Infrastructure & Standards	Business Models, Costs and Funding	Organisational Models and Policy
F) Business Model for Perpetual Access and Hosting				
HIGH PRORITY				
i) Procure service as a competitive process			*	*
ii) Acknowledge difference between disciplines			*	*
iii) National hosts to be bound by SLAs, and can be changed if required			*	
iv) Identify funding model - range of national / state sources?			*	*
v) Address archiving as well as hosting. Dark archive can be separate from continuing access hosting		*	*	
vi) Address archiving as well as hosting. Dark archive can be separate from continuing access hosting		*	*	
vii) Disaster recovery for docs with archival rights - Explore German/European equivalent to Portico or make a national level agreement with Portico		*	*	*
viii) Consortium to hold material in multiple locations for redundancy.		*	*	*
LOWER PRORITY				
ix) Locally host nationally licensed content - but only in the context that the working group thought it a presumption - a goal.			*	
x)Address the culture of project funding but not long term funding			*	
xi) Exploit existing systems to provide distributed solution [This is a presumption that all agreed upon at the workshop and therefore classified as "lower priority".]		*	*	
G) Representation and Governance				
HIGH PRORITY				
i) Take into account federal structure and the nature of central and de-centralised functions.			*	*
ii) Composition of Board and Advisory groups - representative of organisations				*
H) Trustworthy organisations				
HIGH PRORITY				
i) Robust organisation essential		*	*	*
ii) Critical mass of staff with range of expertise		*		*
iii) Certification based on DINI or Nestor work. A need for long-term hosting and preservation definition.		*		*

Synthesis of requirements and suggestions identified by German interviewees and in the Frankfurt Workshop

6.3. Conclusions

It should be noted that the initial list of requirements were based on open-ended responses to the interview questionnaire at the initial data gathering stage of the study. Interviewees responded on the basis of their current knowledge and viewpoints: they did not have access to views expressed by others or further work in this study. The Alliance and the consultants conducted further consultation in the Frankfurt workshop which allowed us to further review, validate and prioritise these requirements in discussion with key user groups.

There are 38 different requirements that have been identified in the table. This is a large number. The Frankfurt workshop agreed broad assignment of requirements into higher or lower priority groups but further prioritisation may be needed by the Alliance as it implements the study.

In addition to these requirements a number of interviewees mentioned a current review of copyright legislation in Germany and its potential impact on digital preservation. The Legal Framework Alliance Working Group is responding to and monitoring this review and should feed any outcomes into this Strategy.

Recommendation 2: Wide consultation and further consensus building amongst all stakeholders including publishers will be a critical element of further development and implementation of the Strategy given the complex organisational landscape in Germany.

Recommendation 6: The findings of the "Legal Framework (P6)" Alliance working group which will take into account any new legislation on digital preservation, should inform future updates to the Strategy.

7. Technical Infrastructure and Standards

7.1. Purpose and Scope

The ITT for the strategy identifies a number of technical issues for the study to address:

- implementation of various authentication methods e.g. IP-based, password-based, support for remote access, Shibboleth;
- concepts for Digital Rights Management;
- assessment of scope and scale of any service e.g. metadata, full text, update scope and frequency;
- definition of suitable object structures and identification of “significant properties”;
- degree of standardisation e.g. consistent conversion to PDF/A, and definition of appropriate metadata;
- evaluation of various organisational concepts in particular centralised versus networked/distributed, taking mirroring and disaster scenarios into account.

These issues are touched upon throughout our study but some may need more detailed introduction and explanation for the broader audience of this report. The purpose of this section is to provide that broader context and explanation of selected technical issues or functions that may arise in any proposed infrastructure and to act as a guide to discussion of potential technical requirements and gaps provided in section 8.3. It is not intended to be a comprehensive discussion of all potentially relevant standards but it does seek to point to work of greatest relevance to this study and other reports providing broader surveys.

7.2. Infrastructure Issues

7.2.1 Introduction

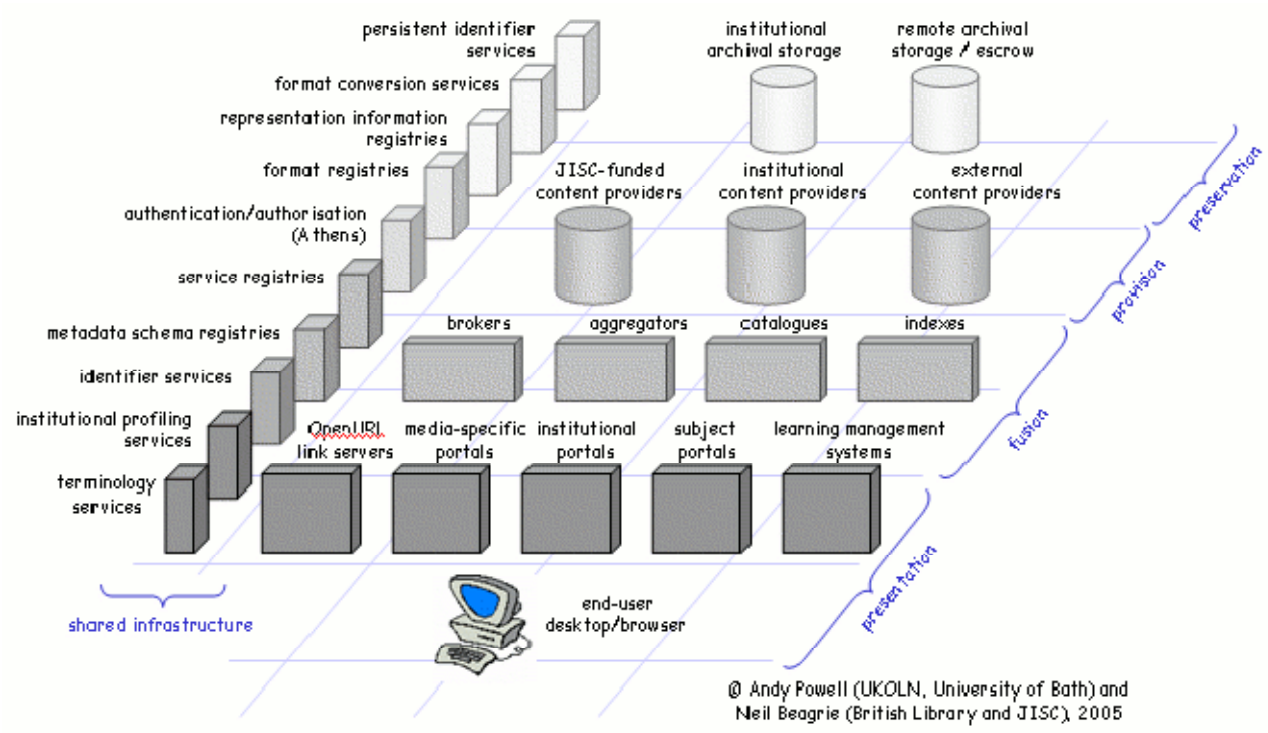
In the design and implementation of any form of aggregation or local hosting service, a number of technical infrastructure issues will need to be addressed and choices made.

7.2.2 Distributed, Centralised, or Hybrid (Both) Service Models

A hosting service or archive could be run from a single place, or distributed or be a hybrid combination of both (some functions centralised and shared). A distributed preservation model could take the form of a limited series of distributed mirror or peer sites for archival

storage and/or hosting, and some distributed but shared (centralised) functions (e.g. ingest?, preservation planning?, preservation action?, persistent identifier service?). The mirror or peer site models offer significant resilience against most causes of loss of data for preservation or communication for hosting respectively. Examples in the interviews include CLOCKSS/LOCKSS and the British Library/UK legal deposit libraries.

Another example of a model of this type (both distributed and some centralised services or shared functions) could be the JISC Information Environment Architecture developed for universities and research institutes in the UK.



JISC Information Environment Architecture

In undertaking this study, the Alliance has a goal to establish and operate an efficient infrastructure for the storage and long-term archival of digital documents, which can guarantee perpetual access to licensed commercial publications and retrodigitised library materials. The key component of such a storage system is a repository or coordinated network of repositories that store data in a reliable and accessible way and that allow this data to be made accessible to participating German scientific institutions in accordance with the applicable licence agreements. Standardised interfaces will allow the storage system (as a service-oriented architecture) to communicate with distributed applications (for example,

subject-specific portals, digital libraries, OpenURL resolvers, etc.) developed and operated independently by individual organisations or information service providers. Such a storage system will essentially form the backbone of the integrated information environment for the science and research community envisioned by the Alliance partners.

We suggest elements of the JISC Information Environment Architecture with some adaptation to its preservation and provision (hosting) layers might closely resemble the Alliance's overall ambitions for scholarly information provision and existing developments for authentication and similar services in Germany. We recommend development of a similar service orientated architecture model to illustrate how different service components can contribute to the Alliance's vision for an integrated information environment for the German science and research community.

Recommendation 7: **Develop a service orientated reference architecture model and corresponding exchange formats and communication protocols to illustrate how different service components can contribute to the Alliance's vision for an integrated information environment for the German science and research community.**

7.2.3 Challenges of Scope, Scale and Consistency

Academic journal publishing currently encompasses a very large number of titles and publishers – SWETS estimates there are approximately 52,730 commercially licensed academic journals worldwide: of these 24,369 academic journals are available from large STM publishers and approximately 28,361 academic journals are from small STM publishers (ranging from a few to one journal published).

These may be accessed using a variety of services and mechanisms, including publishers' own websites, aggregators' websites, and local hosting services. One approach is to replicate the paper distribution model electronically by downloading and locally hosting each and every subscribed journal issue. The LOCKSS system fulfils this functionality in part, though the local copy is for backup access, not routine viewing. It should be recognised that arranging this for an entire portfolio of subscriptions is a challenging task and there are currently very few existing examples.

One of the challenges is the wide difference in technical competence and innovation between the small group of very large publishers and the remainder produced by a long tail of over 700 publishers, many of whom are single title society publishers. This polarisation has been observed as being a developing trend (Rightscom 2008).

Another challenge is the fact that publishers don't all present their electronic offerings in the same way. For the same BL e-journal mapping study (Rightscom 2008), the authors examined the publisher outputs and metadata that one aggregator (Ingenta) had to deal with. They found that a number of different processes were involved "depending on the technical sophistication of the publisher and the level of service being purchased". In summary these were :

- Level A (PDF plus a header in XML or SGML, usually configured to the NLM DTD);
- Level B – variation 1 (manual extraction of metadata from either hard copy or PDF, generating an SGML header using Ingenta's own DTD);
- Level B - variation 2 (starting with hard copy or PDF, the metadata is re-keyed creating an XML header using Ingenta's DTD). How automatic these processes can be made depends on the quality of the file, especially the references;
- Level C (XML full text from which metadata is generated along with HTML full text).

In summary, any aggregation process is potentially complex and labour-intensive. This effort is not short-term, one-off, but ongoing as journal issues are published and processed and publishers merge or sell titles. It is also notable that the report found that a significant number of the smallest publishers did not answer a survey question about metadata file formats. It is assumed that they lacked the technical understanding required to answer the question. Some said they simply supplied a PDF to the hosting service for them to extract metadata on their behalf (see the Ingenta example).

7.2.4 Preservation Choices

An aggregation service will need to make decisions about what to load and preserve. There are two broad techniques. One is to preserve the source files which constitute a journal publication. The other is to capture the presentation files (referred to as "rendition archiving"). There are advantages and disadvantages to each and it should also be noted

that there can be significant variations in how each approach is implemented. Some services do both. The advantages of source file preservation is that it is very complete (and likely to include more content than appears in the journal), is received directly from the publisher and is frequently delivered or converted to a few normalised formats facilitating long-term preservation. The disadvantages are that it requires a large upfront investment, there is no assurance that the archive will actually be needed, and the presentation is very likely to differ from the original publication.

The advantages of harvesting presentation files (rendition archiving) are that it is possible to retain the look and feel of the publication and initial costs are likely to be lower. A disadvantage of this technique is that it may be more difficult to preserve the content over time. It should be noted that a strategy for the large scale migration of presentation files from one format to another is still untested.

As an example, Portico does some pre-processing, transforming the incoming material into a normalised format, but stores both source files and presentation files (creating the latter if not supplied). CLOCKSS lets the publisher choose whether to supply source or presentation files. LOCKSS allows subscribers to harvest the presentation files supplied by the publisher. The Portico approach requires most ingest effort, but may provide better future proofing. With CLOCKSS and LOCKSS, there is less effort required at ingest time, but any future migration effort and cost is deferred until the requirement arises.

7.2.5 Service Currency

Researchers want access to the latest publications in their field. If they find that papers and journal issues are appearing on a publisher's website before those in a local aggregation service, they will try to bypass the local service. Two important factors that influence the scale of any time difference are the publishers' performance in providing a copy to aggregators in a timely fashion, and the amount of pre-processing that the aggregator has to carry out before making an issue available. It follows that any successful service making available copies of publishers' e-journals will need to have a good working relationship with the publishers and have enough effort available to process and mount issues quickly and efficiently.

7.3. Managing Access

7.3.1 Authentication and access control

For non-open access publications, some form of access control is necessary in order to protect commercial products from unauthorised access and distribution. The degree of sophistication and strength of protection required needs to be proportionate to the real value of the material being protected.

Several different access control mechanisms have emerged. These essentially divide into two groups – IP authentication (location test) and some form of username and password based identity test. Further explanation of these mechanisms can be found in Appendix 3.

Because of the deficiencies of these systems, a new, standards-based, development has appeared over the last few years which separates the issue of authentication (proving the user's identity) from authorisation (their access rights). It places responsibility for authentication with the user's "home" institution, and authorisation with the service provider. The system, commonly referred to as Shibboleth, has been gaining traction in a number of countries over the last few years. See http://shibboleth.internet2.edu/ for more information on Shibboleth.

Currently publishers and service providers such as aggregators typically support more than one of these authentication mechanisms, and any Federated Strategy on Perpetual Access and Hosting will need to recognise this. However the trend towards a federated approach to authentication and authorisation as exemplified by Shibboleth should be noted. Shibboleth federations are now deployed or being developed in a number of countries, including the US, the UK (universities and schools), Switzerland, Australia, New Zealand, Finland, France, Denmark, Spain, Greece and China. In Germany a project (AAR) started in 2005 with the goal of building a Shibboleth infrastructure for German Higher Education and Research. This has now evolved into the DFN-AAI federation for German Higher Education and Research. It is anticipated that any integrated approach to e-journal access in Germany would make use of the DFN-AAI federation to control access.

7.3.2 Electronic Licence Management

As the number of digital resources that libraries have to manage grows, so does the complexity of ensuring compliance with the wide and growing variety of licensing conditions.

Two approaches to trying to tackle the problem of licence management and compliance are represented by EDItEUR's ONIX for Publication Licences and the Open Digital Rights Language Initiative (ODRL).

EDItEUR is an international group which aims to coordinate development of the standards infrastructure for electronic commerce in the book and serials industries. It claims 90 members from 17 countries, including "most European countries". ONIX for Publications Licenses (ONIX-PL) is part of a family of XML formats for the communication of licensing terms under the generic name ONIX for Licensing Terms. (website reference - http://www.editeur.org/onix_licensing.html)

It addresses the need for libraries to be able to:

- express licences in a machine-readable format,
- load them into electronic resource management systems,
- link them to digital resources, and
- communicate key usage terms to users

The Open Digital Rights Language (ODRL) Initiative is an international effort aimed at developing and promoting an open standard for rights expressions. It aims to provide flexible and interoperable mechanisms to "support transparent and innovative use of digital content in publishing, distributing and consuming of digital media across all sectors and communities".

The ODRL initiative has an international advisory board with representatives, covering industry, research and the academic sector, from the USA(4), the UK, Germany, Australia and Finland. ODRL is an XML-based standard Rights Expression Language using in DRM systems and open content management systems. It has created a profile that supports Creative Commons licences. (Website reference - http://odrl.net/)

7.3.3 Digital Rights Management

Digital Rights Management (DRM) is a class of technologies that allow rights owners to set and enforce terms under which people can access and use their intellectual property. The core concept of DRM is the digital licence. Instead of buying content, the consumer purchases a licence granting certain rights defining usage rules for the content. The content

itself is encrypted and can only be accessed or used once the consumer has been authorised (e.g. by paying an access fee).

DRM clearly has a potentially useful role in the world of video, audio and computer games products where commercial interests are at stake. However its usefulness in controlling access to academic research publications, where researchers are more interested in high visibility than content protection, is probably limited. Its use also raises potential problems for preservation services. If DRM encryption has been applied to a preserved document, future access will also require knowledge of the keys used. The supporting technology also needs to be preserved. Some believe the trend is away from the use of DRM.

7.4. Standards

7.4.1 Object Structures and Metadata

As noted earlier, there is diversity in precisely what content is stored and in what formats. Some current and emerging standards that are relevant include:

NLM-DTD – The National Library of Medicine Journal Archiving and Interchange Tag Suite. This was developed to provide a common format for exchanging data between publishers and archives.

PDF and PDF/A – Adobe's Portable Document Format, created in 1993, has become the ubiquitous encoding for e-published journals etc. Formerly a proprietary format, it was released as an open standard in 2008 and published by ISO. **PDF/A** is a file format for the long-term archiving of electronic documents.

TIFF – Tagged Image File Format is a file format for storing raster images (as opposed to vector images), including line art.

JPEG and **JPEG 2000** –JPEG (the acronym comes from the name of the group that issued the standard in 1992 – the Joint Photographic Experts Group) is a lossy compression method (ie information is lost in the process of making the file smaller) which uses a compression process known as quantisation to simplify images. JPEG2000 is an updated version of the standard.

MPEG-7 is a standard for describing the multimedia content data. MPEG–7 is not aimed at any particular application; rather, the elements that MPEG–7 standardises support as broad a range of applications as possible.

ONIX (ONline Information eXchange) is an international standard for representing book, serial and video product information in electronic form. Many on-line book traders such as Amazon and Barnes & Noble use this metadata standard to transfer information about their products.

Dublin Core Metadata Initiative (DCMI). The Dublin Core was developed to provide a simple and concise description of Web-based documents.

Metadata Encoding and Transmission Standard (METS). 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.

More information about these and other related standards can be found in Appendix 3.

7.5. Persistent Identifiers and Services

One challenge with electronically published materials is being able to find and reference items in a consistent manner. The exact website address (URL) of a particular document may change with time. The concept of the Persistent Identifier has been developed to provide a way of directing access attempts via a lookup process so that the reference remains constant, even if the location of the object changes.

Two important standards are **DOI** – the Digital Object Identifier, and **URN** – a Uniform Resource Name. More details on these standards can be found in Appendix 3.

8. Gap Analysis for a Federated Strategy on Perpetual Access and Hosting

8.1. Overview

The purpose of the gap analysis is to identify discrepancies between current and ideal states for perpetual access and hosting for electronic resources in Germany. The results will be used to formulate the Strategy and as evidence in a call for action and implementation plan to achieve it in coming years.

The gap analysis draws on previous sections outlining the current state internationally (section 4) and within Germany (section 5) and for technical infrastructure issues and standards (section 7). The ideal state is drawn from interviewees' views on the future position in Germany (section 6).

The Gap analysis is structured around 4 areas of analysis in the study:

- Content, Licensing, Perpetual Access, Preservation and Hosting;
- Technical Infrastructure and Standards;
- Business Models, Costs and Funding;
- Organisational Models and Policy.

Of these, the gaps in Content, Licensing, Perpetual Access, Preservation and Hosting can currently be identified and discussed in the most concrete and tangible way. Both an ideal state is suggested and a summary of the current state provided derived from previous sections. Based on a comparison of the two we provide a broad indication (Low, Medium, Large) of the gap between the two. This is followed by recommendations on how the gaps (particularly the most significant ones) might be addressed.

The ideal state for Technical Infrastructure, Business Models/Costs/Funding, and Organisational Models/Policy, is more dependent on future choices that are made and therefore the discussion and analysis cannot be made in the same way. These are therefore addressed in more general terms with recommendations for approaches in the Strategy provided as appropriate.

8.2. Licensing, Content, Perpetual Access, and Hosting Gap Analysis

8.2.1 Licensing and Content (e-journals)

Ideal state for e-journals: Develop DFG national licences, standardisation of license terms, and text of clauses for model licences; allow for continuing access at short notice and without publisher veto in licences if subscription ceases; ensure access to journals orphaned by transfers between publishers; allow for materials to be in a common non-proprietary platform for long-term access; standardise metadata output; supply of metadata for authorised applications.

Current state for e-journals: Licensing criteria exist for the DFG national licences, criteria include perpetual access and local hosting rights and work is underway to develop these into model licence clauses; only the DFG national licences and MPDL currently have the specific focus on local hosting clauses with publishers; the TRANSFER Code of Practice (UKSG 2008) has developed a set of guidelines for transfer between publishers but these are not yet incorporated into licences within Germany; there is considerable variety in metadata and formats across the publishing industry but some efforts by libraries and publishers to encourage standardisation around the NLM DTD/schema; MPDL is developing a set of technical requirements for publishers to accompany its licences and these are under discussion with Alliance partners.

Current gap level for perpetual access and hosting clauses in licences:

Medium for DFG national and MPDL licences.

Large for regional and institutional licences.

Recommendations to address current gaps for perpetual access and hosting clauses in e-journal licences:

Recommendation 8: **Develop model licence and clauses for DFG national licences and seek to share, maintain and extend these within Germany.**

Recommendation 9: Incorporate TRANSFER Code of Practice into model licences and agreements. Seek to influence future development and strengthening of the TRANSFER Code of Practice.

Recommendation 10: Develop and agree technical guidelines and requirements to accompany licence agreements. These should promote common standards e.g. use of the NLM DTD/schema.

8.2.2 Licensing and Content (retro-digitised materials)

Ideal state for content and rights of retro-digitised material: Retro-digitisation should employ and generate standardised metadata and file formats which support the preservation and hosting of the content; copyrights in works proposed for digitisation should be identifiable and a rights permissions process in place.

Current state for content and rights of retro-digitised material: The DFG have funded a comprehensive set of Digitisation Guidelines which include guidance on metadata and file format standards that are widely referenced and used; the control of the end-to-end processes is within key digitisers - they create material so can specify standards; work done to date to develop standards and infrastructure to handle retro-digitisation within Germany is very advanced - the two national competence centres in digitisation were established in 1997 and have been in operation for over a decade; there is widespread adoption of common standards such as TIFF or JPEG 2000 for file storage formats; since 2007 DNB has collaborated with the Publisher Association and the collecting society (VG Wort) on a model that aimed to facilitate the digitisation of out-of-print and orphan works whilst respecting the relevant copyright protections. This work lead to the development of the ARROW project (Accessible registries of rights information and orphan works towards Europeana). [note the current state is primarily for larger institutions in Germany. Smaller institutions are covered in another report to DFG (BSB/Armed Forces University forthcoming].

Gap level for content and rights of retro-digitised material:

Low to **Medium** for standards;

Large for rights identification and clearance.

Recommendations to address gaps for retro-digitisation content:

Recommendation 11: **Maintain and continue to develop DFG Digitisation Guidelines.**

Recommendation 12: **Facilitate rights identification and clearance for the digitisation of out-of-print and orphan works. If outcomes of pilot project are promising continue development of the ARROW registry.**

8.2.3 Preservation and Hosting for Retro-digitised Materials

Ideal state for preservation and hosting for retro-digitised materials: National Centres of expertise are in place with workflow management systems for mass digitisation; mass storage is available to institutions for preservation of digital master images; tools exist for preservation planning and preservation action (e.g. migration) when needed; hosting services to provide access to outputs of digitisation are sustainable.

Current state for preservation and hosting for retro-digitised materials: Two national competence centres for digitisation are established at the State and University Library at Göttingen and the Bavarian State Library and have been in operation since 1997. They are developing work flow systems for mass digitisation for example BSB is developing "ZEND" which covers all processes from preparation, scanning, indexing, storage and preservation up to access and reuse of the archived digital master files under certain circumstances. Partnerships with nearby super-computer centres at GWDG and the Leibniz Super-Computer Centre respectively give access to relevant expertise and technical infrastructure for mass storage at currently marginal or no additional cost. BSB is working with Google books (public private partnership) on mass digitisation of German books. The current amount preserved is 134 Terabytes and expected future growth is 100 Terabytes plus per annum as the output from the Google partnership starts to be included. The Göttingen Digitisation Centre currently has around 25 Tb of digitisation files and over the next 5 years

their storage requirements may grow to approximately150 TB depending on the archiving strategy and redundancy of objects stored. Some regional Verbünde and library centres are also developing mass storage for retro-digitised materials. For example at KOBV digitisation storage is currently 50 Tb and they anticipate 1-2 petabytes of storage will be required in the next 3 years as they are involved in some very large digitisation projects with Berlin State Library and other partner libraries. Mass digitisation is also a feature on a smaller scale in the research organisations. For example, The Max Planck Digital Library plans to increase the number of retro-digitised collections, which will result in an increase of about 0.3 Tb annually for scientific articles. Due to an ongoing digitisation project of photographs they expect an increase in storage space for digitised photographs of about 5 Tb annually over the next five years. National legal deposit law for German electronic publications covers all German publishers including open access publishers and retro-digitised material and thus provides an additional level of protection for retro-digitised material. Archival image storage formats and metadata are relatively stable so change is unlikely in the near-term. Currently there are relatively few digital preservation planning and action tools (which will be needed in the medium to long-term) or service models to sustain them. However a number of digital preservation research projects both within Germany and internationally are currently working in this area and may address this requirement as tools and support services emerge.

The issue of sustainability for outputs from retro-digitisation is a concern shared in all the major countries engaged in mass digitisation. A major initiative in Germany has been DigiZeitschriften. It was created in 1997 to undertake a large-scale digitisation of back issues of scholarly journals, following the example of JSTOR in the US. DigiZeitschriften e.V. now operates as a registered not-for-profit organisation. It supports itself through institutional subscriptions (192 in late 2008) sold to research libraries and institutes both in Germany (65%) and the rest of the world (35%) and through the contributions in kind of its partner libraries. Sustainability is still a challenge elsewhere. A Nestor sponsored study of digitisation in German Museums identified concerns over infrastructure and staffing in small institutions to maintain outputs and services for them over the long-term and similar conclusions are emerging from a forthcoming DFG funded study.

Gap level for preservation and hosting for retro-digitised materials:

Low for national centres of expertise as two well established centres are in place; **Medium** to potentially **Large** over time for workflow management systems for mass digitisation given exponential increases in scale expected.

Medium for mass storage is available to institutions for preservation of digital master images. Position may be more severe for some institutions and become more so over time e.g. smaller institutions.

Medium for tools exist for preservation planning and preservation action (e.g. migration) when needed. Several tools emerging from projects and sustainable procedures and services for maintaining them are under development or planned.

Medium for hosting services to provide access to outputs of digitisation are sustainable. It is possible to point to success stories but challenges remain.

Recommendations to address gaps for preservation and hosting for retro-digitised materials:

Recommendation 13: Maintain and continue to develop investment in applied R&D in areas such as workflow systems, sustainability, and cost/benefit studies, for retro-digitisation, archiving, and hosting.

Recommendation 14: Invest in development/purchase of digital preservation infrastructures and tools, and services to maintain them.

Recommendation 15: Consider the federated role that Verbünde and regional library services could play in support of retro-digitisation storage, particularly for small and medium-sized institutions.

8.2.4 Perpetual Access, Preservation and Hosting Services for e-journals

Ideal state for perpetual access, preservation and hosting for e-journals: Agreed procedures and trusted services exist to implement perpetual access, preservation and hosting clauses for all e-journal licences in Germany; services should provide for continuing access if subscriptions cease (perpetual access) at short notice with certainty (no veto); service coverage of issues within the journal should be complete and their content quality assured at ingest; service coverage should include small specialist publishers as well as large ones so all e-journal licences are covered; services within Germany are available to support local hosting when appropriate; service levels and technical standards are agreed for all services; supplementary materials and research data associated with articles have persistent links and are preserved and accessible; services should be trustworthy and their competence to provide, perpetual access, digital preservation or hosting certified.

Current state for perpetual access, preservation and hosting for e-journals: This was a major area of concern for all interviewees. As noted in section 4.8, interviewees have very low levels of confidence in arrangements for perpetual access for 10 years or more for e-journals. It is also seen as a matter of high importance for their institutions. We also noted that most of the commercial digital content licensed within Germany is provided via servers based elsewhere and therefore not preserved as part of electronic legal deposit in Germany (see preservation services table end of this section). Interviewees noted that the majority of their commercial digital publications (e-journals, e-books, and databases) are currently accessed via publishers' servers. Some are also accessed via aggregators' platforms. Currently a minority are hosted locally for access (see hosting table end of this section). When prompted for the reasons for these access choices, interviewees pointed to the importance of services associated with accessing the content (e.g. article first features offered by the publisher) or the absence of specific services (e.g. publisher unable to provide selective access rights to previously subscribed material on their server). Sometimes licence conditions and terms when a licence subscription was cancelled also required local hosting or local hosting had been found to be a more cost-effective option for past issues than an offer from the publisher. It is important to note that these decisions on hosting normally had

been applicable to a specific point in time and set of circumstances (e.g. services offered, not offered, and cost) and could therefore change over time. We note there is currently no service within Germany that is providing a perpetual access, digital preservation, or local hosting service for more than a small part of this licensed material. Some respondents noted that they a have a dark archive escrow copy as a temporary archiving option or are piloting local hosting for specific journals negotiated through the DFG national licensing scheme. So far German institutions are not members of other digital preservation services such as Portico or CLOCKSS (although this is under consideration). There is some participation in LOCKSS, with 7 institutions in Germany participating in various pilot activities and projects based on the software. Only two current preservation service options support perpetual access: Portico and local hosting (e.g. LOCKSS). Coverage of issues within Journal titles by services is variable. Not all preservation services require deposit of complete contents of journals and only one service (Portico) currently offers a mechanism for checking its archive content against a subscriber's requirements. Coverage of small specialist publishers by third party services is extremely limited. Coverage of the 52 publishers in our interview list is very variable but sometimes good. At present supplementary data associated with journal articles are rarely included in archive workflows by publishers. TIB in conjunction with other European partners is promoting the use of DOIs for persistent linking of research datasets to articles. Mechanisms for certification of German institutional repositories have been developed by DINI and proposed for trusted digital preservation services by Nestor.

Gap level for perpetual access, preservation and hosting for e-journals:

Large for agreed procedures and trusted services to implement perpetual access, preservation and hosting clauses for all e-journal licences in Germany;

Large for services that should provide for continuing access if subscriptions cease (perpetual access) at short notice with certainty (no veto);

Medium to **Large** for service coverage of issues within the journal should be complete and their content quality assured at ingest;

Large for service coverage should include small specialist publishers as well as large ones so more e-journal licences are covered;

Low to **Large** [depending on scale required] for services within Germany are available to support local hosting when appropriate;

Large for service levels and technical standards are agreed for all services;

Large for supplementary materials and research data associated with articles have persistent links and are preserved and accessible;

Medium for services should be trustworthy and their competence to provide, perpetual access, digital preservation or hosting certified.

Recommendations to address gaps for perpetual access, preservation and hosting for e-journals:

Recommendation 16: Develop funding, and a financial strategy and economic models to provide a framework for decision-making on support and prioritisation of hosting and perpetual access services.

Recommendation 17: Develop an archive content selection and development policy to scope required initial coverage and help prioritise content to be included as it develops.

Recommendation 4: Develop management information for licensing of e-journals and requirements for perpetual access and hosting. This should facilitate analysis of the need and size of demand for federated action.

Recommendation 18: Select or establish service provider(s) who can deliver implementation of your perpetual access rights and/or local hosting rights for e-journals.

Recommendation 19: Establish dialogue with publishers, service providers and other potential partners on technical and service level issues, coverage of journals, and approaches for small publishers.

Recommendation 20: Evaluate archiving practice and preservation and access services for supplementary materials to journal articles; and use of persistent identifier services for linking supplementary materials and research datasets associated with journal articles.

Recommendation 21: Develop and apply criteria and standards for hosting and perpetual access/preservation services and certification processes for them.

Publisher	DNB	CLOCKSS	KB	LOCKSS	PORTICO
American Anthropological Association				● ●	● ●
American Association for Advancement of Science					
American Chemical Society (ACS)					●
American Institute for Aeronautics and Astronautics (AIAA)					
American Institute of Physics (AIP)		●			●
American Mathematical Society (AMS)					● ●
American Meteorological Society					●
American Physical Society (APS)					●
American Psychological Association (APA)					● ●
American Society of Civil Engineers (ASCE)					●
American Society of Mechanical Engineers (ASME)					● ●
American Society of Metals (ASM)					
Association of Computer Machinery (ACM)					
Annual Reviews				● ●	●
Berkeley Electronic Press		●		● ●	● ●
BioMed Central			●	● ●	
Brill			●		
Brookings Institution Press					
Cambridge University Press					●
Cengage Learning/Gale					
Digizeitschriften	●				
Duke University Press				● ●	● ●
Elsevier		●	●		● ●
Emerald				● ●	
Hindawi Publishing Corporation				● ●	● ●
IEEE and IEE//IEL/IET Digital Library					●
Institute of Physics		●		● ●	●
IOS Press			●		
Johns Hopkins University Press				● ●	●
JSTOR					
Karger	●				
Kluwer Law International			●		
Mary Ann Liebert			●		● ●
MIT Press				● ●	● ●
Nature Publishing		●			● ●
Ohio State Uni Press				● ●	
Oxford University Press		●	●	● ●	● ●
Palgrave Macmillan					● ●
Project Muse				● ●	
Proquest					
Royal Society of Chemistry (RSC)		●		● ●	●
Royal Society of Medicine Press				● ●	● ●
SAGE Publications		●	●	● ●	● ●
Society for Industrial and Applied Mathematics (SIAM)					● ●
Seismological Society of America				● ●	●
SPIE Journals					● ●
Springer	●	●	●		● ●
Taylor & Francis/Informa		●	●	● ●	● ●
Thieme	●		●		● ●
University of California Press				● ●	● ●
University of Chicago Press				● ●	●
Wiley-Blackwell	●	●	●		● ●

Key: ● - Preserved ● - Post Cancellation Access

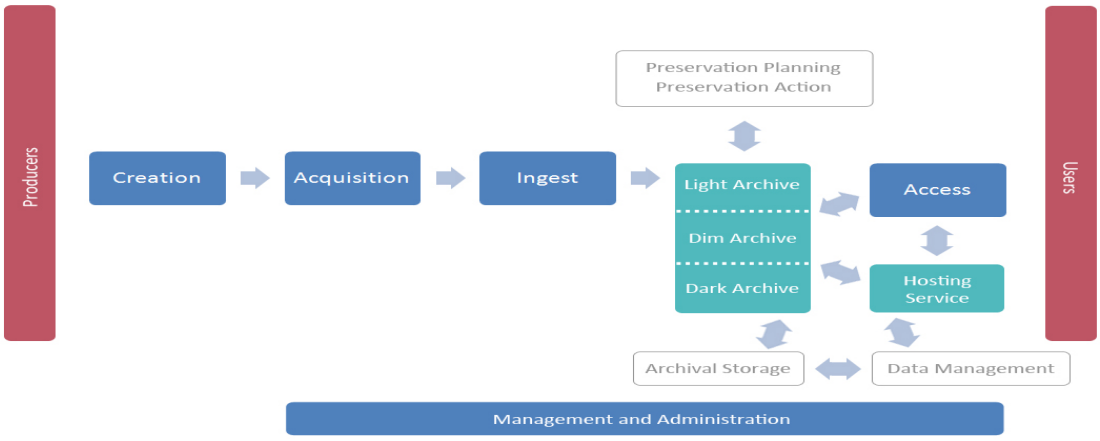
Publisher	Fraunhofer	MPG	Helmholtz	TiB	SUB	BSL	LSL	BC	HeBIS	FAK	NRWC	B-W Consortium	VZG	HBZ	ZiB/KOBV	BVB	Berlin St. Library	FiZ-Karlsruhe	Leibniz Centre	GWDG	Frankfurt UL
American Anthropological Association																					
American Association for Advancement of Science																					
American Chemical Society (ACS)																					
American Institute for Aeronautics and Astronautics (AIAA)																					
American Institute of Physics (AIP)																					
American Mathematical Society (AMS)																					
American Meteorological Society																					
American Physical Society (APS)				○														●			
American Psychological Association (APA)																					
American Society of Civil Engineers (ASCE)																					
American Society of Mechanical Engineers (ASME)																					
American Society of Metals (ASM)																					
Association of Computer Machinery (ACM)																					
Annual Reviews																					
Berkeley Electronic Press																					
BioMed Central																					
Brill																					
Brookings Institution Press																					
Cambridge University Press																					
Cengage Learning/Gale																					
Digizeitschriften																					
Duke University Press																					
Elsevier									○	○	○			●	●						●
Emerald																					
Hindawi Publishing Corporation																					
IEEE and IEE/IEL/IET Digital Library																					
Institute of Physics				□																	
IOS Press																					
Johns Hopkins University Press																					
JSTOR																					

	Fraunhofer	MPG	Gibbs-Helmholtz	TiB	SUB	BSL	LSL	BC	HeBIS	FAK	NRWC	B-W Consortium	VZG	HBZ	ZiB/KOBV	BVB	Berlin State	Fitz	Leibniz Centre	GWDG	Frankfurt UL
Karger																					
Kluwer Law International																					
Mary Ann Liebert																					
MIT Press																					
Nature Publishing				□																	
Ohio State Uni Press																					
Oxford University Press																					
Palgrave Macmillan																					
Project Muse									○												●
Proquest																					
Royal Society of Chemistry (RSC)				□																	
Royal Society of Medicine Press																					
SAGE Publications																					
Society for Industrial and Applied Mathematics (SIAM)																					
Seismological Society of America																					
SPIE Journals																					
Springer				□						○					●						
Taylor & Francis/Informa				□																	
Thieme																		●			
University of California Press																					
University of Chicago Press																					
Wiley-Blackwell																					

Key: ○ Licensee negotiator hosting locally (often via service provider) □Licensee negotiator proposing to hosting locally in future ● hosting service provider

e-journal Hosting in Germany by Licensees and their Service Providers

8.3. Technical Infrastructure and Standards Gap Analysis



Functions covered by the Technical Infrastructure and Standards Gap Analysis (shown in blue).

8.3.1 Introduction

At this stage of the Federated Strategy on Perpetual Access and Hosting, its preferred set of services, requirements and its technical implementation(s) are not determined.

However it is clear from requirements gathered to date that the service(s) will need to show to libraries and users, that it can provide a reliable, trustworthy, comprehensive, up-to-date service that is easy to navigate and use, and which can quickly acquire and load new issues as soon as they are published. At the same time it must have the co-operation and trust of publishers and show that it can protect their commercial interests and also be sufficiently technically supported or co-ordinated that the ingest process has minimal impact on publishers' workflows. As other parts of this report show, a number of different approaches and models have already been established, each with their own merits, but also with some potential drawbacks that deserve consideration.

The technical infrastructure supporting e-publishing is relatively immature and is certain to change and evolve. Any solution adopted will need to allow for regular updating and development to reflect these changes.

This section attempts to summarise some of the key technical and operational issues that will need detailed consideration when establishing a service(s).

8.3.2 Acquisition: Delivery and Storage Formats

For acquisition, agreement needs to be reached on appropriate standards. There appears to be convergence on using PDF and/or XML plus the NLM-DTD (http://dtd.nlm.nih.gov/publishing/) as a metadata standard for the delivery and storage of published material. [Note that the most recent release of the NLM-DTD (Version 3) is not fully backwards compatible with earlier versions though upgrading is said to be relatively easy]. A comprehensive service needs to account for a wide range of suppliers, including the large number of small publishers, many of whom might want to deliver their material in other formats and may not be able to provide any metadata. It will need to trade off a viable service against imposing too many restrictions on publishers.

Choices will need to be made on mechanisms for transmission of data between publisher and hosting service, including file formats, supported compression formats, authentication, and “push” or “pull” FTP procedures. A choice will also need to be made about whether or not to “normalise” material into a consistent internal format, or to accept and store whatever publishers provide.

For the main body of text the most common formats that will need to be processed are PDF and XML, together with HTML. But supplementary files are sometimes made available and a wide range of formats may be used for these, including .doc, .avi, .jpg, .mp3, .mp4 and .xls or .csv. Decisions will need to be made on how to handle these.

Recommendation 10: Develop and agree technical guidelines and requirements to accompany licence agreements. These should promote common standards e.g. use of the NLM DTD/schema.

Recommendation 20: Evaluate archiving practice and preservation and access services for supplementary materials to journal articles; and use of persistent identifier services for linking supplementary materials and research datasets associated with journal articles.

Recommendation 22: Decide on mechanisms for content acquisition and transfer for e-journals and whether to store material as acquired, or normalise to an internal format.

8.3.3 Access: Discovery and Navigation

A hosting service will need a front end with a searching and browsing interface. Such front end services may be provided decentralised at various sites even if the storage of the full documents (the back end) occurs completely or in large parts at a central site. Decisions will need to be made about how this works and what sort of front-end search engine to use. The contents also need to be made visible to search tools in common use. Google is the current favourite, even for academics, so visibility to this search engine would be a distinct advantage. Some aggregators have reached agreement with Google to enable the service to index material that is only accessible to subscribers – search engines are otherwise locked out at the "front door" of subscription services. At the same time, it should be recognised that Google is unlikely to be the dominant search tool for the indefinite future. The recent announcement of an agreement between Microsoft and Yahoo illustrates the fact that this is still a competitive environment. People are also getting used to search services that learn what you appear to be interested in and tailor results to suit. The world of Facebook and Twitter and the way these services work is likely to shape the expectations of a new generation of research students. All these issues and factors need to be taken into account in designing, developing and updating the interface to a hosting service.

Recommendation 23: Design a search and browse interface that meets current standards and expectations in terms of ease of use and presentation of results.

8.3.4 Access: Authentication

Access to subscribed material will need to be controlled. A Shibboleth style infrastructure (where authentication is separated from authorisation) is recommended, though alternatives, such as IP address authentication and/or username/password, will also need to be supported until Shibboleth becomes more widely adopted.

Recommendation 24: Develop an authentication and authorisation infrastructure based on SAML (such as Shibboleth) that meets international standards and supports single-sign-on to a wide range of resources. Continue to support legacy mechanisms such as IP and username/password.

8.3.5 Licence Management & Operational Issues

The task of handling multiple licences with differing conditions has led to several initiatives which should be considered for adoption. A UK-led initiative has led to a model licence for e-journals (http://www.nesli2.ac.uk/model.htm). This is used as a starting point for negotiation rather than a straightjacket, but helps bring some standardisation and order to licensing conditions. ONIX-PL (http://www.editeur.org/onix_licensing.html) provides for the electronic exchange of licensing terms using XML, while the ODRL initiative (http://odrl.net/) aims at developing and promoting open standards for rights expressions.

Publishers may sometimes wish to modify or withdraw a publication. A policy will be needed on how to handle these requests.

Recommendation 8: Develop model licence and clauses for DFG national licences and seek to share, maintain and extend these within Germany.

Recommendation 25: Implement mechanisms to maintain an archive of licence entitlements, administer rights management, and appropriate access control.

Recommendation 26: Establish a Technical Panel that can review and decide technical issues and develop appropriate guidelines, standards, and inter-operability.

8.4. Business Models, Costs and Funding Gap Analysis

As noted earlier in the study both digital preservation and electronic hosting and publishing are relatively new and immature areas. Business models, information on costs, and funding streams for them are still evolving and sometimes subject to rapid changes. A particular difficulty is that digital and electronic are often developing in parallel or in addition to print and other analogue media or a lengthy period of transition is needed between them: it is not simply a matter of one replacing the other at least in the immediate future. It will take time for the system to be brought into economic balance and for new processes, collaborations, or organisations to establish trust and viability. The Strategy and business models supporting it will need to recognise these factors.

In terms of business models identified in the study, institutions in Germany included in the interviews have a mix of “project” and “core” funding from a variety of sources and new services typically have a start-up phase with project or capital funding to establish themselves and then aim to develop operational services supported by core funding or external revenue streams.

We have noted that the DFG is limited by its constitution to funding projects of limited time duration but it is a potential source of funding for initiating activity in the start-up phase of the strategy implementation and for continuing its existing project funding in areas such as digitisation and R&D related to the strategy. Operational funding after that initiation phase is most likely to come from a mixture of state and federal sources with the potential for other income streams in some instances from subscriptions (e.g. as for DigiZeitschriften) or value-added services (e.g. as for document supply). Existing collaborations or extensions of them within Germany or with international partners could also contribute to developing economies of scale and risk-reduction.

There are a number of potential comparators in the study for costs of preservation and hosting services and information on this has been gathered on behalf of the Alliance. In deciding the Strategy there are a number of decisions that will need to be made about the content coverage to be included and the services and service levels required to be associated with them. These will directly influence the costs of implementing the Strategy and will affect the degree of variation expected from potential comparators.

Management information will need to be captured and methodologies for assessing costs and benefits will need to be put in place as part of the support needed for implementing the Strategy.

At this stage of developing the strategy, prior to defined options being decided, it is premature to identify “gaps” in this area, being uncertain of what will be chosen as the final direction of travel. However, the following issues came out in the interviews and are largely focused around funding concerns:

- Issues around “project” versus “long term” funding. Like in other countries funding for projects is easier to obtain than a commitment to long term funding. This does raise issues. For example, the problem of addressing the maintenance costs associated with the outputs of digitisation projects once the term of the “project” ceases. There needs to be recognition of the financial impact of the whole digitisation process.
- Recognise that existing budgets may have been fixed for up to the next 5 years and that any strategy implementation need to address at a minimum a bridging or transitional funding in the short to medium term with a view to fully developing a financial and business plan beyond this.
- As it is acknowledged that local hosting is both expensive and hard to do well, this may influence the preferred solution and the options adopted for the strategy.
- There is a gap in the cost data available from institutions to this study. The mixed model of most funding makes it difficult to determine costs for specific activity. Some cost data is available but it is limited. Data on funding is more accessible.
- There is no long term strategy to guide hosting/access arrangements or on which to build a sustainable funding model. Although nationally funded organisations see the archiving and preservation of material as part of the remit, currently not all content is provided for via these institutions.
- The distinction between a dark archive that a national library might hold and a hosting service for continuing access needs to made and both addressed within the strategy.

Our initial recommendations relating to business models, costs and funding are:

Recommendation 27: The implementation of the Strategy should address the need for sustainability and affordability and its phasing take into consideration the budgetary cycles and commitments of potential funders.

Recommendation 3: Develop a federated organisational and funding model to maintain and develop the national licensing initiative. Include perpetual access and hosting in its scope. Explore where savings achieved from licensing collaboratively might contribute to funds for shared perpetual access and hosting requirements.

Recommendation 4: Develop management information for licensing of e-journals and requirements for perpetual access and hosting. This should facilitate analysis of the need and size of demand for federated action.

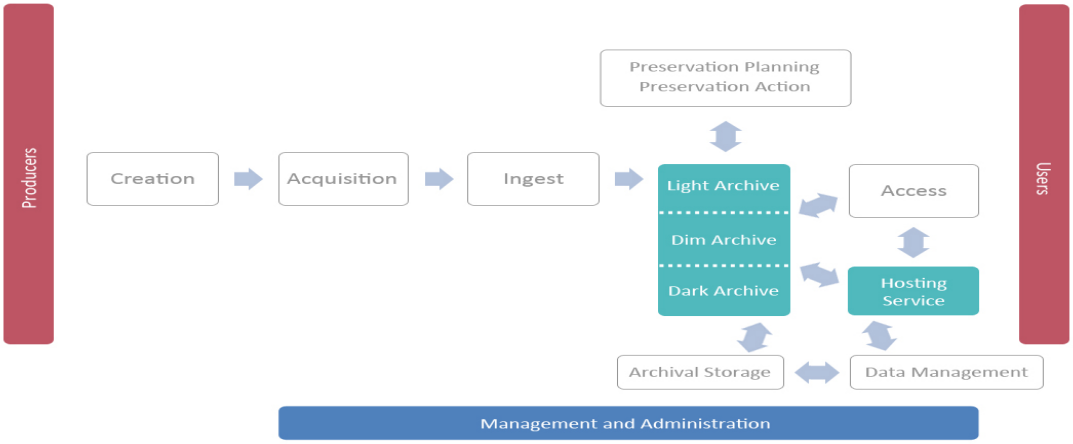
Recommendation 13: Maintain and continue to develop investment in applied R&D in areas such as workflow systems, sustainability, and cost/benefit studies, for retro-digitisation, archiving, and hosting.

Recommendation 16: Develop funding, and a financial strategy and economic models to provide a framework for decision-making on support and prioritisation of hosting and perpetual access services.

Recommendation 17: Develop an archive content selection and development policy to scope required initial coverage and help prioritise content to be included as it develops.

Recommendation 26: Establish a Technical Panel that can review and decide technical issues and develop appropriate guidelines, standards, and inter-operability.

8.5 Organisational Models and Policy Gap Analysis



Functions covered by the Organisational Models and Policy Gap Analysis (shown in blue)

At this point in the development of a Federated Strategy on Perpetual Access and Hosting, prior to defined options being decided, it is premature to identify "gaps" in this area, being uncertain of what will be chosen as the final direction of travel. However, based upon the information received from interviewees and from the other investigations undertaken as part of the study it is possible to make some key observations that potentially point to gaps that will need to be addressed no matter which options for the strategy are finally decided upon.

- **Establishing the organisational infrastructure**. Current organisations and associations will need to be either further developed, supplemented or new ones established in order to implement a strategy effectively. Interviewees had a number of suggestions including the following:
    - A "Strategic Oversight Committee" with members from both the central and de-centralised elements of the federal structure.
    - A governance body to represent all 16 states.
    - Establish a "DFN like" organisational model for a national hosting system, defining who should be in the starting group.

- A "Hosting Working Group" of the Alliance to coordinate and/or provide a hosting mediation functions.

From the international interviews, it is possible that the evolution over time of JISC Collections (from collaborative national working group for e-content, working group supported by small executive of staff embedded in some member organisations, to community owned not-for profit company with staff) and its organisation and policies may prove particularly informative.

- **New policies to cover the following aspects of implementation:**
  - Selection of content for archiving or hosting services;
  - Technical procedures and standards to enable taking full-text from different publisher archives and maintaining common archive(s) or hosting service(s).
  - The strategy will lead to national policies on "archiving" and "hosting".
- **Bridging interests and fostering co-operation.** Several of the interviewees noted the importance of the strategy recognising the federal structure of Germany and the need to build upon existing expertise and seek new synergies rather than creating something entirely new. The organisation established for the implementation of the strategy needs to be flexible and responsive to facilitate the success of this approach. This should include consideration of international co-operation.
- **Linking the Hosting Strategy to National Procurement**
  - The aim of which being to provide savings in procurement and hosting costs.

Note that the decisions made about business model and financial backing for the strategy, short through to long term, will impact upon the type of organisational model that is practical to adopt for the implementation of the strategy.

Our initial recommendations on organisational models and policy are:

Recommendation 28: Decisions made about business models and financial backing for the strategy, short through to long term, will impact upon the type of organisational model(s) that is practical to adopt for the implementation of the Strategy. The organisational model(s) should be in keeping with the type of archiving and hosting solution(s) adopted.

Recommendation 29: The organisation to implement the Strategy should aim to bridge interests, foster co-operation, and take into account the federal structure of Germany.

Recommendation 1: Maintain an international dimension to the Strategy, evaluate potential international partnerships and service providers, and maintain an oversight of emerging best practice and trends internationally.

Recommendation 2: Wide consultation and further consensus building amongst all stakeholders including publishers will be a critical element of further development and implementation of the Strategy given the complex organisational landscape in Germany.

Recommendation 3: Develop a federated organisational and funding model to maintain and develop the national licensing initiative. Include perpetual access and hosting in its scope. Explore where savings achieved from licensing collaboratively might contribute to funds for shared perpetual access and hosting requirements.

Recommendation 15: Consider the federated role that Verbünde and regional library services could play in support of retro-digitisation storage, particularly for small and medium-sized institutions.

Recommendation 17: Develop an archive content selection and development policy to scope required initial coverage and help prioritise content to be included as it develops.

Recommendation 19: Establish dialogue with publishers, service providers and other potential partners on technical and service level issues, coverage of journals, and approaches for small publishers.

Recommendation 21: Develop and apply criteria and standards for hosting and perpetual access/preservation services and certification processes for them.

Recommendation 26: Establish a Technical Panel that can review and decide technical issues and develop appropriate guidelines, standards, and inter-operability.

9. Use Cases

9.1. Introduction

The aim of these use cases is to explain the reasons why action is needed and the drivers for the strategy. As detailed elsewhere in this study, research is critical to the health of Germany's economy and some 230 million euros are invested each year in scholarly resources for researchers in German Universities alone. E-journals are a critical and heavily used part of these resources particularly for researchers in science, technology, and medical fields.

At the same time significant funds are also being invested in retro-digitisation of scholarly resources still in paper form such as books, journals or other materials such as photographs. The DFG's investment in retro-digitisation was 12 million euros in 2008 and has been between 25-50 million euros in total over the previous 5-10 years. These resources are also heavily used by researchers.

These are large public investments which need to be safeguarded and maintained and whose loss of access could severely impair research.

The use cases illustrate the different challenges libraries and research organisation face in maintaining access to electronic journals and retro-digitised materials and link these to potential solutions discussed in the study. The potential solutions themselves (and their potential strengths and weaknesses in relation to the Strategy) are discussed in section 10.

9.2. Use Cases for E-Journals

Over the last decade, e-journals have become increasingly popular and have begun to oust paper journals as a mechanism for delivering papers on academic research to the community. Although there are many obvious benefits that accrue from publishing and accessing academic journals electronically, there are costs and challenges which need to be addressed in maintaining access to them over the long-term. We need to be looking ahead if our growing collections of electronic research publications are still going to be accessible for researchers 5, 10, 50 or 100 years from now.

Electronic journals present new concerns and new challenges. High amongst the list of challenges are ensuring that this material can continue to be accessed for the indefinite future (perpetual access) and be preserved (digital preservation). A particular challenge is that, while libraries and librarians might be assumed to be the natural custodians of these materials, and to have the greatest interest in their long term preservation, the owners of the files are in the majority of cases the publishers, and publishers and their journal titles can merge, transfer or disappear over time. The only original copy of an issue of an e-journal now tends to be held by the publisher but the long term preservation of that copy is arguably of greater importance to the library and research communities. So we may need new procedures, collaborations and sometimes organisations to ensure safe custody of these objects is maintained through time, technologies, and organisational changes.

In order to illustrate the potential problems and the ability of different potential solutions to cope with them, five e-journal use cases have been identified and described in more detail below.

9.2.1 Use Case 1: Library Cancels Subscription to Journal

Few libraries have remained unaffected by the serials pricing crisis of the past several years and journal cancellations have been one outcome of this. Unfortunately the current economic environment is likely to lead to an increase in cancellations as library budgets are under pressure. A library may have invested significant funds over a number of years to purchase previous issues of the journal before finding that cancellation to the current journal issue is necessary. Before taking this step, the library will need to have developed its own risk assessment and a contingency plan for perpetual access to past subscriptions. This will include having negotiated at the outset where possible with publishers, post-cancellation access clauses in licences, any additional fees and the mechanism for access (e.g. via publisher's servers, trusted third party services, or copies for loading on local systems). If copies are provided for local hosting, the nature of those copies needs to be agreed. Provision for supplying a copy of the content for local hosting by the subscribing institution in the event of cancellation is a common provision of publishers' licences. However it may in practice be very difficult for the institution alone to put in place or afford the services needed for the access and preservation of that content. Institutions may therefore seek to achieve

economies of scale through co-operation on joint services or via a contractor. If post-cancellation access is provided via the publisher's server there will be long-term risks to the access from commercial fluctuations and changes in the publisher's business. This has led to the emergence of trusted third-party archiving services for e-journals that can provide a potential safety-net for perpetual access in these cases. Coverage of different journals and journal issues and participation of publishers in these services vary considerably and will need to be assessed. In all cases another major factor in selecting a solution will be the types and levels of service the library's users require to access the content not solely its preservation. Electronic access and preservation services may involve a significant level of investment and cost and lead even large libraries to seek collaborative solutions to post-cancellation access.

As the following table shows, only some of the potential solutions reviewed are set up to provide access where the trigger event is cancellation of a subscription.

Use Case 1: Library cancels subscription and needs access to past issues to which they subscribed		
Solution	**Triggered?**	**Comment**
Publisher's server	Yes	Provides a short-medium term solution but a poor perpetual access guarantee in the longer-term unless it can be combined with another service solution. Aggregator licences only guarantee access while a subscription is in force. There are no perpetual access rights or services.
National or regional Archive	Yes Immediate	A co-operative service is established by participating libraries and rights negotiated with publishers. Immediate access would be dependent on licence agreements and agreed authentication of entitlements for access.
Local Archive	Yes Immediate	An access and preservation service has been put in place by individual library and rights negotiated with publishers.
LOCKSS	Yes Immediate	LOCKSS assures members of access to participating publishers' content provided they have purchased or licensed that content.
CLOCKSS	No	CLOCKSS only makes titles openly accessible following a defined trigger event. Subscription cancellation is not such a trigger.
Portico	Yes Delayed	Providing the publisher has nominated Portico as a provider of post-cancellation access.
KB e-Depot	No	Except for onsite access.

9.2.2 Use Case 2: E-Journal or its past issues are no longer available from the publisher

This is a highly likely use case as publishers merge or change their business models, or as larger publishers review and adjust their portfolio of titles. Journal titles are also sometimes

traded between publishers, which may mean that access to past issues is no longer supported by the previous owner.

The UKSG TRANSFER initiative (http://www.uksg.org/Transfer/Code) has produced a Code of Practice aimed at easing the problems created when journal titles move between publishers. Of relevance are the following paragraphs contained in the latest version of the code (UKSG 2008):

> *It is very common for journal content to be included in one or more archiving services. The Receiving Publisher will not remove content that was previously deposited in an archive, or archives, even if the Receiving Publisher will not be continuing to deposit content in the archive, or archives. The Receiving Publisher is encouraged to continue the existing, or equivalent, archiving arrangements for a journal after the Effective Transfer Date.*
>
> *Customers that have been granted perpetual access rights to previously published content with the authority of the journal owner must have those rights honoured. Either the Transferring or the Receiving Publisher, or both, could fulfil perpetual access obligations.*

The real-life scenario of the publisher Sage no longer offering its publication Graft has provided an opportunity to demonstrate the success of two archiving solutions, Portico and CLOCKSS. In this case each is able to continue to offer access to the issues they hold, either as open access (CLOCKSS) or else as a service to members (Portico). While it cannot be guaranteed that the archive will include all back-issues of the title (as with Graft), participation in an archiving solution which covers at least some issues will significantly reduce the risk of disruptions to continuity of service.

Use Case 2: E-Journal or its past issues are no longer available from the publisher		
Solution	**Triggered?**	**Comment**
National or regional Archive	Yes Immediate	A co-operative service is established by participating libraries and rights negotiated with publishers. Immediate access would be dependent on licence agreements and agreed authentication of entitlements for access.
Local Archive	Yes Immediate	An access and preservation service has been put in place by individual library and rights negotiated with publishers.
LOCKSS	Yes	LOCKSS assures members of access to participating publishers' content

	Immediate	provided they have purchased or licensed that content.
CLOCKSS	Yes Delayed	The title would be made openly accessible to all.
Portico	Yes Delayed	The title would be opened up to all active participants, regardless of whether they previously subscribed to the content.
KB e-Depot	Yes Delayed	All archived titles will always be available onsite at the KB. This trigger would result in the title being made openly accessible, subject to publisher agreement.

9.2.3 Use Case 3: Publisher Ceases Operation

In this use case, the publisher is no longer in business and therefore unable to support a service providing access to their collection of previously published journal issues. Suitable strategies for coping with this event depend on risk management techniques. The probability of this trigger event for large publishers is arguably low. However, the impact for research institutions may be unacceptably high.

Use Case 3: Publisher has ceased operation and e-publication is no longer possible		
Solution	**Triggered?**	**Comment**
National or regional Archive	Yes Immediate	A co-operative service is established by participating libraries and rights negotiated with publishers. Immediate access would be dependent on licence agreements and agreed authentication of entitlements for access.
Local Archive	Yes Immediate	An access and preservation service has been put in place by individual library and rights negotiated with publishers.
LOCKSS	Yes Immediate	LOCKSS assures members of access to participating publishers' content provided they have purchased or licensed that content.
CLOCKSS	Yes Delayed	The title would be made openly accessible to all.
Portico	Yes Delayed	The title would be opened up to all active participants, regardless of whether they previously subscribed to the content.
KB e-Depot	Yes Delayed	All archived titles will always be available onsite at the KB. This trigger would result in the title being made openly accessible, if agreed with publisher.

9.2.4 Use Case 4: Catastrophic Failure of Publisher's Operations/Servers

In this use case, access to a publisher's e-journals suffers a major disruption. This is a somewhat unlikely, but not entirely implausible event. Examples can be thought of that might include fire, flood, explosion, lightning strike or terrorist action, which result in the total

destruction of a publisher's servers and, perhaps, damaged or destroyed backup materials. It should be added that major publishers are likely to take steps to protect themselves against such events by running mirror sites at various locations around the world, but smaller publishers might not have the resources to do this. Off-site or secure fire-safe storage of backup media is also common practice, but perhaps not universal.

Rebuilding a service following such an event might take many months. An agreement between a publisher and an archiving solution that this is an acceptable trigger to open up access to the archive would enable subscribers to continue to access back copies. For example Portico specifies the following condition for opening access: "Licensor has stopped publishing or providing access to the Publication for a period longer than ninety (90) days due to technical difficulties or any business interruption, bankruptcy, insolvency, receivership or business failure." Note, however, that if it is assumed that the publisher will be able to resume operations at a future date, such a service would only open up access to the material on the preservation service's servers until such time as the publisher was able to resume their own service. The normal legal protections against downloading bulk data for access on other servers would still have to apply.

Use Case 4: Catastrophic failure of publisher's operations/servers		
Solution	**Triggered?**	**Comment**
National or Regional Archive	Yes Immediate	A co-operative service is established by participating libraries and rights negotiated with publishers. Immediate access would be dependent on licence agreements and agreed authentication of entitlements for access.
Local Archive	Yes Immediate	An access and preservation service has been put in place by individual library and rights negotiated with publishers.
LOCKSS	Yes Immediate	LOCKSS assures members of access to participating publishers' content provided they have purchased or licensed that content.
CLOCKSS	Yes Delayed	Content moved to a hosting platform and made freely available
Portico	Yes Delayed	Portico subscribers get free access to content as long as publisher is unable to provide a service
KB e-Depot	Possibly Delayed	All archived titles will always be available onsite at the KB. Network access depends on agreements with publishers.

9.2.5 Use Case 5: Publisher Cannot Host Appropriately

In this use case, appropriate hosting cannot be achieved via a publisher's server. This may be due to a number of causes. Some small and medium-sized publishers do not have their own service infrastructure to provide access services to the content and this is the responsibility of subscribers or their contractors (this is more likely for databases than e-journals). In other cases a publisher may not be able to provide the services or service level guarantees required by the subscriber and the service has to been arranged via the local institution or its contractors. This is particularly important when institutions want to do more with the content than is possible via the publisher's platform. Publisher's access charges for access to past issues post-cancellation In Use Case 1 above, might also prove uneconomic and the subscriber with local hosting rights can arrange a more competitive service locally or via their contractor.

Note decisions in Use Case 5 are taken in light of specific circumstances at specific times. If circumstances change over time e.g. new services are available from the publisher which are appropriate, decisions can be reviewed.

Use Case 5: Publisher cannot host appropriately		
Solution	**An Option?**	**Comment**
National or regional Host	Possibly	Depends on required service/service level and whether it can be put in place.
Local Host	Possibly	Depends on required service/service level and whether it can be put in place.
LOCKSS	Possibly	Depends on required service/service level and whether it can be put in place via other systems linked to LOCKSS.
CLOCKSS	No	Not a defined trigger event.
Portico	No	Not a defined trigger event.
KB e-Depot	No	Not a defined trigger event.

9.3. Use Cases For Retro-digitisation

The challenges for long-term access and preservation of retro-digitised materials are somewhat different to those for e-journals as institutions have greater direct control over all stages from creation through to use and access to the material. In principle retro-digitised materials can also be re-digitised from originals so preservation may not be seen to be as pressing as for electronic "born-digital" originals. In practice however funding for digitisation

is limited and the quantity of non-digitised material is huge, so institutions and funders expect their investments in digitisation to be preserved and accessible.

We provide a use case for archival storage of outputs from retro-digitisation to illustrate the potential challenges and solutions involved for their preservation. A separate challenge exists for the access services: not strictly one of preservation but maintaining contemporary and up to date access services for users of these files. Business models for electronic publishing and hosting involved in this are strictly out of the scope for this study but we recognise their importance and the inter-relationship with that of preservation and have included a use case in the form of a case study for sustaining access because of this.

9.3.1 Use Case 6: Archival Storage and Preservation for Retro-Digitised Materials

Digital Preservation is a challenge for retro-digitised material as changes in technology gradually can make digital file formats obsolete and inaccessible, or because of the different infrastructure and staff expertise needed to maintain them compared to existing facilities for maintaining the analogue originals.

Typically the master files for digitisation (archive master images) are captured at relatively high resolution and are therefore very large. Hence small files in different formats are generated for access (access files). The archive master files and the access files are typically stored in different systems and have very different service requirements.

Archive master file formats such as TIFF are widely deployed standards, and practice guidelines for capturing images and accompanying metadata (e.g. DFG Digitisation Guidelines) well developed. If sufficient care is taken during a digitisation project, the archive master file and accompanying metadata should not be threatened in the short-medium term by file format obsolescence. A greater immediate threat may be the need for appropriate technical infrastructure and staff expertise to maintain and preserve the outputs once the project is complete. This may only be possible via a technical partner or collaboration, particularly given that the rapidly escalating scale of mass digitisation requires similar mass digitisation storage for preservation of the files.

Use Case 6: Archival Storage and Preservation for Retro-Digitised Materials		
Solution	**An option?**	**Comment**
National or regional Archive	Yes	Significant archival storage infrastructure available to two national digitisation centres/state libraries via technical partners. National Legal Deposit Law includes retro-digitised materials in Germany. National German Legal Deposit infrastructure may provide a safety-net of last resort for preservation? Infrastructure and services for archival storage available at some regional Verbünde and Library Service Centres to member institutions/customers.
Local Archive	Possibly	Depends on infrastructure available to the institution or its technical partner. Particularly challenging for small and medium-sized institutions.
Publisher's server	Not Applicable	For public/private partnerships. Applicable only for access files. Partner normally retains archival master files and provides archival storage for this.
LOCKSS	Not applicable.	
CLOCKSS	Not applicable.	
Portico	Possibly.	Portico provides a service for d-newspapers (digitized newspapers). It is evaluating the technology and costs associated with preservation of local library content and exploring the possibility of launching a new service.
KB e-Depot	Not applicable	The international e-depot is focussed on e-journals. However the National e-Depot includes very large Dutch retro-digitised holdings.

9.3.2 Use Case 7: Sustainability of Access Services for Retro-digitised Materials

As noted in section 4.6 another major concern raised by interviewees is around sustainability of access to the outputs of retro-digitisation projects. A recent study (Mahon et al 2009 p11) has defined sustainability as:

"...the ability to generate or gain access to the resources – financial or otherwise – needed to protect and increase the value of the content or service for those who use it. A sustainable project covers its operating costs through a combination of revenue sources and cost-management strategies and continues to enhance its value based on the needs of the user community."

It also notes:

"Furthermore, sustaining the value of the resource requires more than just 'keeping the lights on'. As new technologies develop and user expectations shift and grow, a resource risks fading slowly into irrelevance if it does not constantly grow and innovate in ways that continue to benefit its constituents. Not doing this, in the most extreme cases, can result in a resource becoming inaccessible. More often, though, a static resource will lose value

over time. Not only does this diminish the ability of the project to achieve its mission, but also the declining usefulness of the resource will make it even more difficult to generate the revenue needed to sustain a minimal level of activity."

That study also includes a case study on DigiZeitschriften, the German-language archive of scholarly journals. As of December 2008, this contained 3.5 million pages of content. It was created in 1997 following the decision of a group of library directors in Germany to work together, with funding from the German Research Foundation (DFG), to undertake a large-scale digitisation of back issues of scholarly journals, following the example of JSTOR in the US. It now operates as a registered not-for-profit organisation and supports itself through institutional subscriptions (192 in late 2008) sold to research libraries and institutes both in Germany (65%) and the rest of the world (35%) and through the contributions in kind of its partner libraries. The 9 founding partners has expanded to include 14 member libraries, all with Special Collection (SSG) status based around Germany. Lessons from DigiZeitschriften were seen as: partnerships among those with common goals can help to lower costs; subscriptions are a worthwhile option when the content is highly valuable to customers with the ability to pay for it; a good fit with a host institution can be a key aspect to sustainability; an active feedback loop with users helps a site stay current with evolving user expectations for online academic resources; continued growth and innovation requires committed leadership and dedicated staff.

Other business models to note encountered during our study for access to retro-digitised materials included open access funded directly by the institution and made available either through an institutional repository or a technical partner; and public-private partnerships (Google Books with BSB or British Library with publishers).

Use Case 7: Sustainability of Access Services for Retro-Digitised Materials		
Solution	**An option?**	**Comment**
National or regional Host	Yes	Hosting infrastructure available via FIZ and other regional library services as well as specific service collaborations such as DigiZeitschriften.
Local Host	Yes	Range of institutional repositories providing access to some local retro-digitised materials such as e-theses as well as pre-prints etc.
Publisher's server	Yes	For public/private partnerships. Partner normally retains archival master files and provides archival storage for this and may also provide a local hosting option.
LOCKSS	Not applicable.	
CLOCKSS	Not applicable.	
Portico	Not applicable.	
KB e-Depot	Not applicable	

10. Scenarios, Potential Solutions and Recommended Approaches

In the following section the study will outline potential scenarios in terms of different dimensions of risk and innovation and also describe potential solutions that are intended to address the use cases described in section 9. It concludes with our recommended approaches for the Alliance to consider for hosting, perpetual access and archiving for e-journals and retro-digitised materials.

10.1. RISK Profile Scenarios for you to consider

10.1.1 Introduction

There are many individual potential solutions or combinations of potential solutions to the use cases. In order to assess options further we have developed a number of strategy development scenarios. These strategy development scenarios are developed along different levels of risk and innovation and also take into account different starting points and constraints for the strategy's development. Scenarios are outlined in terms of a division of solutions between indigenous solutions within Germany, internationally, or potential hybrids of the two, list contributing solutions, and consider the levels of risk and innovation and content types included. Each scenario concludes with brief scenario characteristics including a visualisation of the assessment of risk and innovation. Finally, all scenarios are contrasted in a comparison table. Individual solutions and their potential strengths and weaknesses in relation to the Strategy are described in detail later in section 10.2.

Risk Levels

Levels of risk can change over different time horizons. The qualitative risk estimation for each scenario distinguishes the levels low, medium and high in short-term (1-5 years) and in long-term (6-10 years) and a brief explanation of the assigned levels of risk is provided in the description of each scenario. The qualitative risk estimation is a summation of the risk of failure or significant short-fall in relation to any of the different dimensions of the study. The risk estimations can be related to our gap analysis (section 8) and discussion of the international context (section 4) and the current position in Germany (section 5), including

the existing levels of expertise and experience and the levels of experience and expertise of content, technical, and organisational requirements, and funding models needed in each scenario. The principal risks that exist in different dimensions of the study can be summarised as follows:

Content, Licensing, Perpetual Access, Preservation and Hosting: loss of access to the content as the result of the lack of robust mechanisms to acquire and implement licensing rights (such as perpetual access) or to preserve the content (or its significant properties such as links); failure to provide required coverage of content and service.

Technical Infrastructure and Standards: inability to scale appropriately to very high-volumes; to ingest and potentially normalise complex content and metadata from heterogeneous sources; provision of national scale access services and access control; failure to identify appropriate service standards for preservation or access and ensure providers can meet and maintain them.

Business Models, Costs and Funding: inability to raise sufficient revenues or contain costs; potential “single points of failure” and critical dependencies that could be subject to sudden changes in the external environment over time.

Organisational and Policy Issues: level of complexity and effort needed to establish and maintain an organisation; inappropriate governance models with stakeholders unable to control or influence key policy areas such as future selection; failure to obtain buy-in from a core stakeholder group (e.g. publishers).

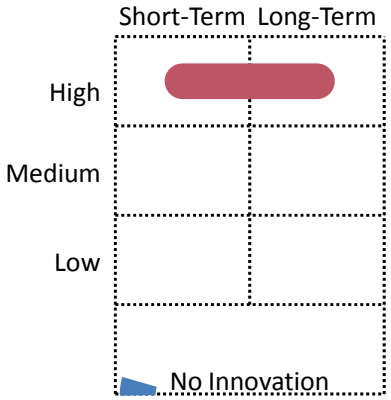
Levels of Innovation

Levels of innovation range from no innovation by remaining with the current position to ambitious strategies including radical innovation. It may refer to incremental and emergent innovation leveraging pre-existing organisations and areas of expertise (described as “conservative” in the innovation level) or radical and revolutionary changes in technology, business models or organisations (described as “ambitious” in the innovation level). While innovation is essential to add value, it typically also involves risk and the possibility of failure, particularly as the level of innovation increases towards radical or ambitious change. The innovation level therefore has a degree of correlation with levels of risk in the scenarios.

10.1.2 Do Nothing (Remain with Current Position)

This scenario envisages no forward momentum or action on developing this Strategy. Major investments in the purchase of commercial content and retro-digitisation would be left exposed to significant long-term risks. The major gaps identified in addressing perpetual access for e-journals would remain. Ad hoc reactive approaches would need to be taken in response to orphaning of journals or post-cancellation access of them; archiving and hosting of retro-digitisation materials would utilise existing infrastructure and standards but gaps may begin to widen even for the largest institutions as the creation of retro-digitised content accelerates. The position for small and medium sized institutions would not be improved. There would be no synergies or value-added to research in Germany through development of a collaborative federated network.

Scenario characteristics



Risk: This scenario was rated as being high risk both short-term and long-term.

Innovation: None

Contributing solutions:

- Publisher's servers
- Local hosting
- Local archiving

Content types:

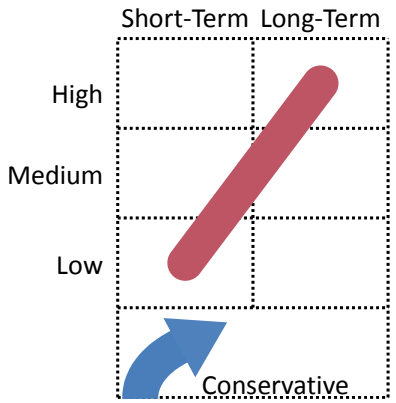
- Retro-digitised (partial)

10.1.3 Conservative National Scenario for Archiving and Hosting

A conservative Strategy development scenario can be outlined along a path that involves a low risk of failure during attempted development and deployment of a solution infrastructure within Germany. However, as the applied infrastructures have not been designed and developed to serve the purpose of this Strategy, it could also result in very slow and potentially inadequate development or very poor performance of the solution. The involved stakeholders and responsible service providers may also not have the mandate or funding to

act in such a scenario. In this conservative national scenario, the strategy development would seek to build on modest extensions and adaptations of the existing infrastructure in Germany such as the German Legal Deposit system as well as on the Digitisation Infrastructures at the National and State level (Kopal, GDZ, MDZ, BABS etc). It would furthermore build on and develop limited extensions and adaptations of the capacity of already existing Library Service Providers (Bibliotheks-Verbünde, FIZ). For such a scenario an appropriate stakeholder reflection as well as a feasibility study would be necessary in the first place to come up with an appropriate organisational and economic implementation plan. It remains doubtful in the long-term whether all requirements and recommendations could be fulfilled in an appropriate and economic way through such an approach. It is likely to be best suited to addressing the needs of retro-digitised and open –access materials that already have significant infrastructure that can be developed within a relatively conservative approach but to fall short for addressing the long-term needs the study has identified for international commercial e-journals.

Scenario characteristics



Risk: This scenario was rated as being low risk short-term and high risk long-term.

Innovation: Conservative

Contributing solutions:

- Publisher's servers
- National/Regional hosting
- National/Regional archiving

Content types:

- Retro-digitised
- Open access

10.1.4 Conservative International Scenario for e-journal archiving

Another rather conservative scenario (which would operate on the international level) would be to apply a follower strategy for e-journal archiving and perpetual access. It would require simply waiting for appropriate commercial or public infrastructures to develop internationally without any direction, involvement, and carrying of risk and responsibility locally. It would

involve monitoring and observing ongoing developments, evaluating their appropriateness for local interests and requirements and to sign up to relevant services and infrastructures that are emerging and provided by international commercial or public players. In such a scenario, the National stakeholders would become very much dependent on the availability, quality, and conditions of such commercial or public service providers and their offerings. The ability to influence, initiate, drive, or steer developments in accordance with local interests and requirements could be very limited unless some pro-active action is taken. It is therefore likely to fall short on the requirements for content selection and coverage and services identified in the study.

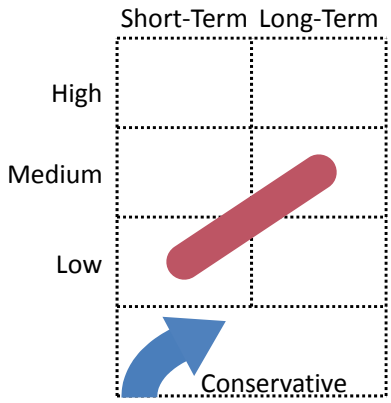
Current international service providers that could be involved in this scenario are Portico, CLOCKSS, and the KB of the Netherlands, who provide differing degrees of archiving service to complement hosting services via publishers' servers. Business models, coverage, and services for all of these are still evolving and represent a medium degree of risk. They have all passed beyond the initiation stage and have established experience and track record even if longer-term business and organisational models remain challenging. Of the service providers covered, currently only Portico provides post-cancellation access which is a requirement for this Strategy.

Using Portico as a partner would mean that a strategy would have to build on Portico's specific recommendation that the long term archiving function and perpetual access (post-cancellation) should be based on existing services and should be separated from the issue of day to day access (hosting) which would be in line with the recommendations of this study. Portico could then provide authenticated access to triggered material if and when required.

It would provide insurance to stakeholders on the National level that the e-journal and other content they have subscribed to will be preserved for the long-term. Portico would only provide access to the e-journals they have preserved after specified 'trigger events'. Portico would preserve publisher's source and presentation files and deliver content in renditions appropriate to the current technology. Portico would become a delivery mechanism in the event of a trigger event. In addition, if a publisher has designated Portico as such, it could also serve as a potential mechanism for post cancellation access.

It should be evaluated whether the coverage of Portico would be appropriate for the German demand). If reliance upon existing systems is possible, Portico believes that this may enable access and preservation needs to be addressed at substantially lower cost than a stand-alone national solution. Portico believes that it has the experience and mature technologies necessary to address the preservation aspects of the requirements referred to in the study.

Scenario characteristics



Risk: This scenario was rated as being low risk short-term and medium risk long-term.

Innovation: Conservative

Contributing solutions:

- Publisher's servers
- KB e-Depot, CLOCKSS, Portico

Content types:

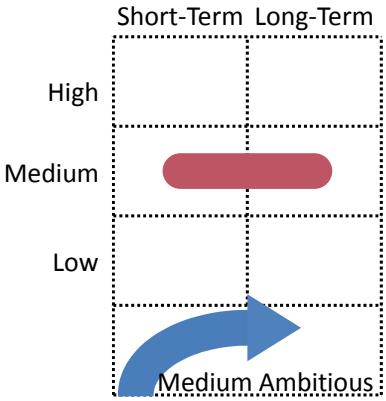
- E-journals

10.1.5 Medium Ambitious Scenario: Developing German infrastructure in partnership with international provision and partners.

A medium ambitious scenario would see German infrastructure developed in partnership with international provision and partners. For e-journals this might see a degree of local provision for hosting developed alongside use of publishers' servers which is well established. For archiving and perpetual access for e-journals it could involve development of new services or technical infrastructure with international partners. All international solutions we discuss for this are relatively new or emerging. We have noted that the landscape for both archiving and hosting is likely to see continuing significant change in the next few years. The Alliance is potentially in a position to pro-actively influence how existing solutions might evolve and develop and also to discuss potential partnerships which might generate new approaches. Other international organisations such as the Knowledge Exchange and partners within this such as the UK, Netherlands, and Denmark can also be

considered. Some of the German archiving infrastructure within this scenario may have dual purpose for archiving of retro-digitised materials as well as e-journals. The larger market for the service, spread of risk, and broader combination of expertise and experience potentially moderate risks in this scenario and therefore have been assessed as medium risk in the short and long term. The risks also exist that the potential partners may not have a common vision, budget horizon, or mandate to act in such a scenario.

Scenario characteristics



Risk: This scenario was rated as being medium risk both short-term and long-term.

Innovation: Medium ambitious through partnership of innovation and usage of existing capabilities

Contributing solutions:

- Publisher's servers
- KB e-Depot, LOCKSS, CLOCKSS, Portico
- Other international partners
- National/Regional archiving
- National/Regional hosting

Content types:

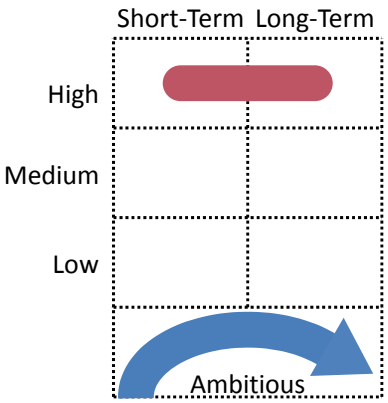
- E-journals
- Retro-digitised
- Open access

10.1.6 Ambitious Scenario: developing comprehensive, dedicated German Hosting and Archiving Options for both e-journals and retro-digitised materials

An ambitious scenario that operates solely on the national level an appropriate national organisation would have to be specified, negotiated, and finally set up that is determined and capable of specifying, implementing, deploying, and running a set of high-quality and potentially federated National and Regional Hosting and Archiving Services supported by a potentially hybrid/federated system with centralised/de-centralised back-end services and centralized/de-centralised front end services. Such an approach would have the advantage of providing a dedicated and tailor-made solution for the needs, requirements, and interests

of the different national stakeholders and could also take advantage of existing infrastructures and competences where appropriate. However, it would also require a very committed and competent project team that could develop, initiate, monitor and steer an implementation and deployment project as well as a very competent consortium of implementation partners and finally an appropriate business and financial model for the operation of the set-up and future long-term service provision phases. To date comprehensive approaches to hosting for e-journals having only been attempted at regional level elsewhere. Organisational and financial challenges and risk levels for a comprehensive hosting and archiving solution for e-journals in this scenario are likely to remain significant for some time. The level of changes from the current position in Germany in both hosting and archiving are large and are assessed as having a higher risk in the short and long term and an ambitious profile for innovation.

Scenario characteristics



Risk: This scenario was rated as being high risk both short-term and long-term.

Innovation: Radical

Contributing solutions:

- National/Regional archiving
- National/Regional hosting

Content types:

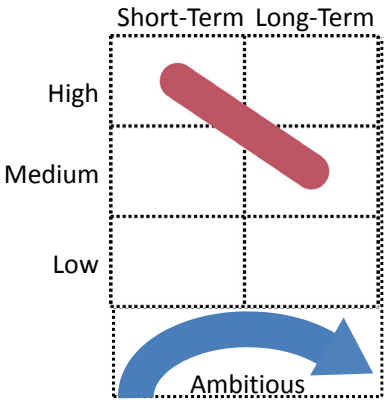
- E-journals
- Retro-digitised
- Open access

10.1.7 Ambitious Scenario: developing comprehensive, archiving options for other content types such as e-books, databases and supplementary materials

This scenario is similar to 10.1.6 outlined above but for different content types. E-books, databases and supplementary materials are excluded from the immediate development of the Strategy. The archiving and preservation of these materials is less developed and often significantly more challenging currently than that for e-journals and retro-digitised materials. These challenges do vary across and within these content types for example some eBooks are already included in existing archiving programmes. However, currently none are addressed comprehensively. The degree of variability in technical and organisational

requirements for them provides a higher level of risk in the short term for a comprehensive option. We envisage in this scenario they will need to be tackled as major research challenges at an international level with international partners and the outcomes of this research will then feed in to the proto-typing and development of services. In the longer-term the risk profile should be reduced to at least a medium level as a result.

Scenario characteristics



Risk: This scenario was rated as being high risk short-term and potentially medium risk long-term as further research and development addresses challenges.

Innovation: Radical

Contributing solutions: Hybrid German and international partnership options

Content types:

- E-books
- Databases
- Supplementary materials

	10.1.2 Do Nothing (Remain with Current Position)	10.1.3 Conservative National Scenario for Archiving and Hosting	10.1.4 Conservative International Scenario for e-journal archiving	10.1.5 Medium Ambitious Scenario: Developing German infrastructure in partnership with international provision and partners	10.1.6 Ambitious Scenario: developing comprehensive, dedicated German Hosting and Archiving Options for both e-journals and retro-digitised materials	10.1.7 Ambitious Scenario: developing comprehensive, archiving options for other content types such as e-books, databases and supplementary materials
Characteristics						
Risk **Short-Term / Long-Term**	High / High	Low / High	Low / Medium	Medium / Medium	High / High	High / Medium
Innovation	No innovation	Conservative	Conservative	Medium ambitious	Ambitious	Ambitious
Contributing Solutions						
Publisher's servers	•	•	•	•		
Local hosting/archives	•					
Regional hosting/archives		•		•	•	
National hosting/archives		•		•	•	
KB, LOCKSS, CLOCKSS, Portico			•	•		
Other international partners				•		•
Hybrid German and international partnership options						•
Content Types						
E-journals			•	•	•	
Retro-digitised	•	•		•	•	
Open access		•		•	•	
E-books						•
Databases						•
Supplementary materials						•

Comparison of strategy development scenarios

10.2. Potential Solutions

Potential solutions introduced in discussion of the use cases and risk scenarios are discussed in greater detail below with an assessment of their relative strengths and weaknesses for the Strategy. We have divided potential German solutions into the entries for National/Regional Hosting Services/Archives or Local Hosting Services/Archives. As described in section 5 The Current Position in Germany, some organisations currently operate at local/regional and national levels so these distinctions can represent gradations of the degree of federation and scope of services offered in these cases.

10.2.1 Using Publisher's Server

Content	Use Case cross references, and comments
e-journals	Use Case 1: Library cancels subscription to Journal (Yes short-medium term) Hosting Service. Publisher may provide post-cancellation access via its servers (hosting service). Only suitable for perpetual access over a short-medium timeframe and needs to be combined with another solution to address longer term risks.
Retro-digitisation	Use Case 7: Sustainability of Access Services for Retro-digitised Materials (Yes) Hosting Service. Public/Private publishing partnerships may help sustain access services. Digital preservation service remains with originator.
Others?	eBooks, Databases

As noted in previous sections access via publishers' servers to licensed e-journals is the most common means of providing access to that content currently in Germany. However as noted in Use Case 5 (Publisher cannot Host Appropriately), there are circumstances when the Publisher cannot host appropriately and therefore some degree of local hosting will always be required within Germany. As noted in Use Case 1 (Library cancels subscription to Journal), perpetual access via the publishers' servers cannot be guaranteed and therefore it

is desirable to combine it with another solution to provide a more robust long-term option for perpetual (post-cancellation) access and digital preservation services.

[note: aggregators are a specialised form of publisher but have no perpetual access or digital preservation services and would be excluded from the Strategy. However aggregators might be able to make a contribution in other ways. They can have lots of experience with ingest from many different publishers and perhaps could be possible contributors in a specific service role for ingest functions for e-journal Archives.]

Strengths

- perpetual access on the publisher's server may be included at no or moderate cost in the licence;
- Large-scale infrastructure at major publishers. Hosting service can have popular user features such as article first for new articles, not available elsewhere.

Weaknesses

- commercial publishing environment in constant flux – significant risk over any reasonable time frame to perpetual access via publisher;
- service orientated towards current access rather than digital preservation.

10.2.2 Using National/Regional Archive(s)

Content	Use Case cross references, and comments
e-journals	Use Case 1: Library cancels subscription to Journal (Yes); Use Case 2: E-journal or its past issues are no longer available from the publisher (Yes); Use Case 3: Publisher Ceases Operation (Yes); Use Case 4: Catastrophic Failure of Publisher's Operations/Servers (Yes); Use Case 5: Publisher Cannot Host Appropriately (Possibly). Dark Archive or Dim Archive.

Retro-digitisation	Use Case 6: Archival Storage and Preservation for Retro-Digitised Materials (Yes); :Sustainability of Access Services for Retro-digitised Materials (Yes). Dark Archive or Dim Archive.
Others?	eBooks, Databases

This solution would need to be developed and an option of federated development is assumed. There were a number of potential variants considered and we briefly discuss these further below:

- Standalone Local Evolution within Germany. This option would not address most key requirements raised during the study. It has not been elaborated further.
- Developing a highly-centralised Single Service in Germany. We believe this would not be sustainable given funding and political structures and distribution of relevant expertise. It has not been elaborated further.
- Developing a fully de-centralised Regional System in Germany. Probably not easily attainable in all 16 states and it would not provide the potential synergies and scale of wider collaboration. It has not been elaborated further.
- Developing a federated network of archives and services. A federated option could allow inter-operability between a defined network of archives and services and shared functions when economies of scale or other benefits can be leveraged. A number of technical implementations or parallels have been mentioned during the study for a federated option. The JISC Information Environment (see Technical Infrastructure and Standards section 7.2.2),The UK eLegal Deposit Network consisting of 4 mirrored archival storage nodes and further access nodes (see British Library section 4.3.2) or developing a private LOCKSS network (see LOCKSS section 4.4.2 and CLOCKSS section 4.4.1) are three examples.

It is possible for a federated network or components of it to solely focus on e-journal content, solely on retro-digitisation, or archiving both type of content. eBooks and Databases might

also be within scope although experience in digital preservation with these content types is currently more limited.

Strengths and weaknesses will depend on definition of the network and its required services but some general observations on potential strengths and weaknesses are provided below.

Strengths

- good fit to political landscape and funding streams in Germany;
- could build on existing digital preservation expertise and emerging infrastructure; potentially can address many archiving needs in the use cases (however actual degree of fitness as a solution could vary considerably).

Weaknesses

- relatively limited experience of archiving e-journal of international publishers with breadth, depth, and scale within Germany;
- existing infrastructure and skills may not always scale well; significant administrative effort needed to agree and set-up such a network.

10.2.3 Using National/Regional Hosting Service(s)

Content	Use Case cross references, and comments
e-journals	Use Case 5: Publisher Cannot Host Appropriately (Possibly). Hosting Service.
Retro-digitisation	Use Case 7: Sustainability of Access Services for Retro-digitised Materials (Yes). Hosting Service.
Others?	eBooks, Databases

Using national/regional hosting services share many similarities with using national/regional archiving services and a degree of federation is also assumed in this solution. In several cases this already exist regionally or nationally at least to some degree and could be

extended. There are already a number of hosting services within Germany particularly for retro-digitised materials on open access or as a subscription service. Hosting services may overlap with archive services at an organisational level although typically the infrastructure and staffing will be separate infrastructures within them. Organisations can also be solely focussed on hosting and access services.

Strengths

- good fit to political landscape and funding streams in Germany;
- could build on existing digital hosting expertise and infrastructure particularly for retro-digitised content and open access;
- potentially can address hosting needs in the use cases (however actual degree of fitness as a solution could vary considerably).

Weaknesses

- relatively limited experience of hosting e-journals of international publishers with breadth, depth, and scale within Germany;
- existing infrastructure and skills may not always scale well;
- significant administrative effort needed to agree and set-up such a network.

10.2.4 Using Local Archive(s) or Hosting Services

Content	Use Case cross references, and comments
e-journals	Use Case 1: Library cancels subscription to Journal (Yes); Use Case 2: E-journal or its past issues are no longer available from the publisher (Yes); Use Case 3: Publisher Ceases Operation (Yes); Use Case 4: Catastrophic Failure of Publisher's Operations/Servers (Yes); Use Case 5: Publisher Cannot Host Appropriately (Possibly).
Retro-digitisation	Use Case 6: Archival Storage and Preservation for Retro-Digitised Materials (Possibly); Use Case 7: Sustainability of Access Services for Retro-digitised Materials (Yes).

Others?	eBooks, Databases

The right to receive a copy of previously subscribed material for local hosting by the institution or its contractor post-cancellation is a relatively common provision in e-journal licences and the option of hosting locally is mandated in the DFG national licences for Germany. Although theoretically a subscribing institution can host such resources post-cancellation or on purchase and meet requirements of most of the e-journal use cases in practice this would be very challenging for them as the costs and complexity of doing so will be high. Effectively most examples of local hosting for e-journals are regional collaborations (e.g. OhioLINK or Scholars' Portal University of Toronto) or utilise a partnership with a major facility (e.g. TIB and FIZ Karlsruhe). Local hosting is more common for some retro-digitised material using local infrastructure although regional or national federation is normally encountered for hosting and sustaining materials at scale (e.g. DigiZeitschriften). Local digital preservation is rare outside of institutions with a preservation mandate as necessary skills and infrastructure are still relatively uncommon. Dark Archive escrows (archival storage only) however are widespread within larger information systems departments.

Strengths

- under direct control of the local institution;
- may be able to tailor access services and features specifically for local audience.

Weaknesses

- no economies of scale unless a partner can be used to provide this;
- may lack standardisation necessary for other re-uses within Germany.

10.2.5 Using CLOCKSS

Content	Use Case cross references, and comments
e-journals	Use Case 1 : Library cancels subscription to Journal (No); Use Case 2: E-Journal or its past issues are no longer available from the publisher

	(Yes Delayed); Use Case 3: Publisher Ceases Operation (Yes Delayed); Use Case 4: Catastrophic Failure of Publisher's Operations/Servers (Yes Delayed); Use Case 5: Publisher Cannot Host Appropriately (No). A Dark Archive: it could be considered as a possible solution to the problem of long term preservation, but it doesn't provide the post-cancellation access service identified as a requirement in this study.
Retro-digitisation	Not Applicable
Others?	

CLOCKSS (Controlled LOCKSS) is based on the LOCKSS software (see section below on LOCKSS) in which a limited number of libraries take on an archival role on behalf of a broader community. The CLOCKSS consortium and its scope are described in section 4.4.1.

Strengths

- it provides a community approach;
- low cost;
- it is likely to appeal to publishers with concerns about security of access;
- a German Node is under discussion for the CLOCKSS system.

Weaknesses

- post-cancellation access identified as a key requirement in the study is not supported;
- limited coverage of major publishers licenced within Germany;
- too early to be able to assess the long-term viability of the programme and its endowment funding.

10.2.6 Using LOCKSS

Content	Use Case cross references, and comments
e-journal	Use Case 1: Library cancels subscription to Journal (Yes); Use Case 2: E-journal or its past issues are no longer available from the publisher (Yes); Use Case 3: Publisher Ceases Operation (Yes); Use Case 4: Catastrophic Failure of Publisher's Operations/Servers (Yes); Use Case 5: Publisher Cannot Host Appropriately (Possibly). A Dim Archive: software.
Retro-digitisation	Not Applicable
Others?	

LOCKSS (Lots of Copies Keep Stuff Safe) enables participating libraries to collect, store, preserve and provide access to their own local copies of content to which they have subscribed. The LOCKSS application is open source software. LOCKSS and the LOCKSS Alliance are described in more detail in section 4.4.2.

Strengths

- it allows libraries to collect and exert control over the material they licence, as they have done when purchasing print journals;
- it covers a number of smaller and therefore probably more vulnerable publishers;
- it requires relatively modest investment in staff and equipment;
- it permits immediate access to the archive whenever there is a problem with communication with a publisher's server, even if very short term.

Weaknesses

- future software development will require an active LOCKSS developer community;

- library and collaborators needs to allocate resources to provide a service e.g. selection;
- it may be difficult to integrate with other institutional technical platforms;
- not all titles of a particular publisher, nor all issues of a particular title are necessarily included in the LOCKSS collection;
- lack of content from major publishers.

10.2.7 Using Portico

Content	Use Case cross references, and Type of Service
e-journal	Use Case 1: Library cancels subscription to Journal (No); Use Case 2: E-journal or its past issues are no longer available from the publisher (Yes Delayed); Use Case 3: Publisher Ceases Operation (Yes Delayed); Use Case 4: Catastrophic Failure of Publisher's Operations/Servers (Yes Delayed); Use Case 5: Publisher Cannot Host Appropriately (No). Dim Archive with post-cancellation access service.
Retro-digitisation	Use Case 6: Archival Storage and Preservation for Retro-Digitised Materials (possibly –new service under evaluation)
Others?	e-books, d-newspapers, possibly some full-text databases in future ("historic digitised collections"),

Portico is designed specifically as a third party archiving service for scholarly electronic literature, beginning with e-journals and e-books. If a publisher has designated Portico as such, it can serve as a potential mechanism for post cancellation access. Portico and its services are described in greater detail in section 4.4.6.

Strengths

- a full-service solution. It removes the onus of long-term archiving of the content;

- it can provide post cancellation access providing the publisher has nominated them as a potential mechanism for this (to date 88% of journals in Portico have this);
- good coverage of major publishers and publisher participation has grown at an impressive rate;
- a range of e-books and some retro-digitised material included;
- archiving approach is very thorough.

Weaknesses

- terms for post-cancellation need to be more transparent and as yet post-cancellation access service is largely untested by subscribers;
- governance model and lack of representation/vote may be a concern for German institutions;
- service is based in US and there is no mirror in Europe or elsewhere (although a tape copy of content is deposited at the KB).

10.2.8 Using KB e-Depot

Content	Use Case cross references, and comments
e-journal	Use Case 1: Library cancels subscription to Journal (No); Use Case 2: E-journal or its past issues are no longer available from the publisher (Yes Delayed); Use Case 3: Publisher Ceases Operation (Yes Delayed); Use Case 4: Catastrophic Failure of Publisher's Operations/Servers (Yes Delayed); Use Case 5: Publisher Cannot Host Appropriately (No). Dim Archive with post-cancellation access service.
Retro-digitisation	Not applicable (Netherlands only)
Others?	

The Koninklijke Bibliotheek (KB) is the national library of the Netherlands and operates the e-Depot which includes international journal publishers, its archive for the Dutch national deposit collection of electronic publications and other e-content (eg Dutch newspapers). The e-depot and its services are described in greater detail in section 4.3.3.

Strengths

- e-Depot aims to cover 20 major STM publishers and 12 major publishers on board;
- the KB has a strong reputation for leadership in digital preservation research and practice;
- close development link with Kopal system in Germany;
- service is currently underwritten by the Dutch government but may need additional income streams "In the coming years the KB intends to develop a sustainable business model for the e-Depot which will reflect both the public and private responsibility for our digital and cultural heritage."

Weaknesses

- no post-cancellation access service;
- no coverage for medium and small publishers;
- at the moment assured access is only available onsite at the KB.

10.2.9 Other International Partners and Hybrid German and International Partnership Options

All solutions discussed above are relatively new or emerging. We have noted that the landscape for both archiving and hosting is likely to see continuing significant change in the next few years. It needs to be recognised that the Alliance is in a position to influence how existing solutions might evolve and develop and also to discuss potential partnerships which might generate new approaches. The challenges we have described in the use cases are international in scope. We have recommended (recommendation 1) that an international dimension to the Strategy is maintained and various hybrids of solutions above could be considered as part of this. Other international organisations (in addition to those mentioned above) such as the Knowledge Exchange and partners within this such as the UK, Netherlands, and Denmark can also be considered within this option.

10.3. Recommended Approaches

The solutions outlined above often overlap and none will provide a unique or complete answer to the use cases we have described. Different levels of risk, effort and innovation apply for each solution and potential combinations of them, so we have outlined a number of risk scenarios for you to consider. Taking into consideration all previous sections of the study report including potential requirements, solutions and risk profiles these are our recommended approaches to the Alliance:

10.3.1 Hosting for e-journals

Access via publishers' servers (solution 10.2.1) is the most common and well-tested approach for hosting of commercial e-journal content licensed by institutions in Germany. On the basis of work completed for the study, we believe it currently provides the most effective main solution for hosting of this content in the majority of cases and is recommended to the Alliance. As noted in Use Case 5 (: Publisher Cannot Host Appropriately) however, some capacity to host e-journals within Germany will always be required in addition to this when publishers cannot host appropriately and will need to be provided for (solutions 10.2.3 and 10.2.4). To implement these approaches you will need to select or establish relevant technical guidelines and requirements (recommendation10), service standards and certification (recommendation 21), and select relevant service providers within Germany when required (recommendation 18). Requirements for, and the relative merits and costs of, different hosting solutions will change over time. Therefore we have also recommended that the implementation needs development of a financial strategy and economic models to provide a framework for appropriate decision making (recommendation 16). The JISC Collections (see section 4.3.1) has a strategy and models which may provide a useful comparator. Obtaining local hosting rights even if there is no immediate intent to exercise them is an important safeguard in the National and MPDL licences. We have recommended that model licences are developed and such rights extended wherever possible within Germany (recommendation 8).

10.3.2 Perpetual Access for e-journals.

Perpetual access for e-journals is much more immature in terms of development of appropriate services and successful business models. In the short-term many libraries can exercise post-cancellation access (and extension of access rights is also possible for National licences) via hosting on publishers' servers (solution 10.2.1). We have noted this is not sufficiently robust as a single perpetual access solution long-term. However given the immaturity of longer-term solutions this is the approach we would recommend for the immediate first horizon of the Strategy (perhaps years 1-3). As noted in Use Case 5 (Publisher Cannot Host Appropriately) however, some capacity to provide archiving and post-cancellation access for e-journals within Germany would be required in addition to this when publishers cannot host appropriately for post-cancellation access or other reasons even in the short-term and will need to be provided for (solutions 10.2.2 and 10.2.4).

In working towards longer-term solutions we would recommend two options are explored in parallel:

- Development of independent perpetual access capacity in Germany (solutions 10.2.2 and 10.2.4) with international partners (solution 10.2.9). As noted above, we believe if it is found necessary to develop a comprehensive German archiving solution for perpetual access to e-journals, this may be more sustainable if it is developed with other international partners. Some pre-existing working relationships e.g. with Denmark, the Netherlands, and UK in the Knowledge Exchange may provide a European basis for this. The larger market for the service, spread of risk, and broader combination of expertise and experience would make this our second preferred option for the long-term, in preference to developing a purely German solution which would have a higher risk profile. A number of technical or other international partnerships could also contribute to this option depending on the direction chosen (e.g. CLOCKSS/LOCKSS, KB, British Library, etc.).
- Participation in Portico (solution 10.2.7). Possibly one institution in Germany could pilot this service over the first horizon of the Strategy (there are no German library participants currently). A dialogue could be established to discuss potential

weaknesses from a German perspective noted in section 10.2.7 above and future directions that might address them.

We have recommended that an international dimension is maintained within the Strategy (recommendation 1) that will be essential for this. In addition, as with hosting of e-journals, to implement these approaches you will need to select or establish relevant technical guidelines and requirements (recommendation10), service standards and certification (recommendation 21), and select relevant service providers within Germany when required (recommendation 18). We have noted that many aspects of digital preservation planning and action are still an ongoing area of R&D and this will need ongoing work (recommendation 13) and investment in development or purchase of digital preservation infrastructures, tools and service as these move from R&D to services (recommendation 14).

We have also indentified within the study that there are significant differences in the e-journal market between the large and small publishers which may require different approaches.

The “core” licensed from large scientific publishers and demanded by a broad spectrum of scientific users represent some of the largest financial investments by organisations in Germany in commercial e-journals. The exact scope of a core list of publishers and titles will need to be determined. These resources are often licensed repeatedly and for long periods of time, so significant reductions in licensing costs can be achieved via national licensing. National licensing strategies and policies for their selection, delivery, ingest, management, hosting and access as well as preservation have to be determined.

Smaller scientific publishers are often offering more specialised scientific content which is furthermore quite often very discipline specific. As noted elsewhere in the study they represent a very long tail of often a single or handful of journals each with corresponding challenges of diversity and lack of economies of scale for archiving services. The set of users in demand of this information is also rather limited as compared to the major set of core titles of the large publishers. In summary this also means, that not all scientific information and content resources produced internationally can and needs to be electronically hosted and archived in a national scientific information environment in Germany. These scientific information and content resources will have to be treated

selectively. However many will be important for research and still need addressing within the Strategy. We have recommended that a dialogue is established with publishers, service providers, and other potential partners on developing approaches and coverage for small publishers (recommendation 19).

E-journals that are produced by means of aggregation also form a different type of scientific resources as their content is already available and only becomes available in a different form and selection via an aggregator. They have no perpetual access rights when accessed via an aggregator and therefore would be excluded from the Strategy.

10.3.3 Hosting for Retro-digitised materials

We have noted that retro-digitisation projects have an established history in Germany and infrastructure has been established at national, regional, or local level at different scales to support its hosting either via open access models (e.g. institutional repositories) or via subscription models (e.g. DigiZeitschriften). Our interviewees generally expressed a high-level of confidence in this infrastructure and existing solutions at national/regional level and local level (solutions 10.2.3 and 10.2.4). Areas of current concern mentioned the needs of small and medium sized institutions and we have recommended that the Strategy considers the federated role that the Verbünde and regional library services could play in support for them (recommendation 15). The other area of concern is around the sustainability of hosting. We have made a number of recommendations which may assist in addressing this including those which may help in containing short-term or long-term costs (recommendations 11, 12, and 13). Other solutions may have a niche contribution to the Strategy. In some cases private/public partnerships and hosting via a publisher's server may be appropriate (solution 10.2.1).

10.3.4 Archiving for Retro-digitised Materials

There are now a number of established national/regional archives for retro-digitised material based around the two national digitisation centres and state libraries of Bavaria and SUB Göttingen and the national legal deposit system Kopal maintained by DNB and SUB Göttingen. Other services such as the Verbünde and regional library services are also developing or have deployed archive storage services for digitisation archive master images.

We would recommend ongoing development based on these systems at national/regional level (solutions 10.2.2) to maximise economies of scale. Discussions on forming a "virtual national library", or exploration of archiving solutions for e-journals within Germany (see 9.3.2 above) might also contribute to this aim. In addition, as with archiving of e-journals, we have noted that many aspects of digital preservation planning and action are still an ongoing area of R&D and this will need ongoing work (recommendation 13) and investment in development or purchase of digital preservation infrastructures, tools and service as these move from R&D to services (recommendation 14). To implement these approaches you will also need to maintain relevant technical guidelines and requirements for digitisation (recommendation11). We have recommended that an international dimension is maintained within the Strategy (recommendation 1). Many of our international interviewees (see section 4) maintain archives for retro-digitisation and may contribute to evolving best practice and standards. It is also possible that Portico (solution 10.2.7) may develop services of interest in some areas such as d-newspapers or historic digitised collections.

10.3.5 Organisational Structure

We believe development of the Alliance and its working group structures provide a positive initial starting point for development of the Strategy. These working groups currently cover national licensing, open access, national hosting strategy, research data, virtual research environments, and legal frameworks. It will need to be either further developed, supplemented, or new ones established in order to fully implement the Strategy effectively.

We recognise that governance and funding underpinning a future organisational structure are likely to be key issues. We have recommended as an initial starting point wider consultation and consensus building amongst all stakeholders (recommendation 2). Successful examples of large-scale collaboration and co-ordination of funding at a federated level normally deliver such tangible levels of economic benefit that they overcome other competitive or organisational differences. Examples such as DFN in Germany or JISC in the UK provide clear economies of scale in shared networks that underpin collaboration and co-ordinated funding.

A key point of shared development to date for the Alliance has been work on national licensing. We have suggested that a federated organisational and funding model might be

developed around this (recommendation 3) as it provides similar compelling advantages that can underpin further development. Extending existing work to develop model licences and clauses and extend their use within Germany (recommendation 8) can contribute to this.

Decisions about business models and archiving and hosting solutions adopted will also influence governance and organisational structures (recommendation 28) for example governance issues will need to be a consideration with service providers. Advice may also need to be taken on the legal and financial implications (e.g. Value Added Tax – VAT) of different options as these discussions evolve.

Interviewees had a number of suggestions for a future organisational structure including the following:

- A "Strategic Oversight Committee" with members from both the central and de-centralised elements of the federal structure.
- A governance body to represent all 16 states.
- Establish a "DFN like" organisational model for a national hosting system, defining who should be in the starting group.
- A "Hosting Working Group" of the Alliance to coordinate and/or provide a hosting mediation functions.

A common element is the need to take into account the federal structure of Germany and foster co-operation (recommendation 29).

If the Alliance is taken as an organisational starting point, as it evolves it might potentially develop a main committee and sub-committees structure similar to that supporting the JISC in the UK. JISC accommodates a federal structure of devolved regions in the UK (England, Northern Ireland, Scotland, and Wales) and also supports community representation and ownership. The current JISC committee structure is as follows:

- JISC main committee
  - Sub-committees:
    - Infrastructure and Resources,
    - Learning and Teaching,

- Organisational Support,
- Support of Research.

These sub-committees are supported by fixed-term working groups (many focussing on similar issues to the Alliance working groups such as scholarly communication or VREs) and staff in an Executive and community-owned service companies http://www.jisc.ac.uk/aboutus/committees.aspx.

We believe it is possible that the evolution over time of JISC Collections in the UK from collaborative national working group for e-content, a working group supported by small executive of staff embedded in some member organisations, to community owned not-for profit service company with staff and its organisation and policies, may also prove particularly informative (see section 4.3.1) as the Alliance considers its future direction and organisational development path for work outlined in this study.

If a relevant organisational structure cannot be developed within Germany some recommended potential approaches in this study may not be implementable. Other recommended potential approaches however may require less new organisational effort or consensus. For example fully-out-sourced solutions such as Portico with independently evolved governance and funding models could still be implemented by individual organisations or smaller consortia (although probably less effectively in terms of the wider Strategy than could be achieved by a wider federation).

10.3.6 Financing

We have discussed current financial models in Germany in section 8.4. We have noted that the DFG is limited by its constitution to funding projects of limited time duration but it is a potential source of funding for initiating activity in the start-up phase of the strategy implementation and for continuing its existing project funding in areas such as digitisation and R&D related to the strategy. Operational funding after that initiation phase is most likely to come from a mixture of state and federal sources with the potential for other income streams in some instances from subscriptions (e.g. as for DigiZeitschriften) or value-added services (e.g. as for document supply). Existing collaborations or extensions of them within Germany or with international partners could also contribute to developing economies of scale and risk-reduction. There are a number of potential comparators in the study for costs

of preservation and hosting services and information on this has been gathered on behalf of the Alliance. In deciding the Strategy there are a number of decisions that will need to be made about the content coverage to be included and the services and service levels required to be associated with them. These will directly influence the costs of implementing the Strategy and will affect the degree of variation expected from potential comparators. The phasing and budget profile will also need to take into account the budget cycles and commitments of potential funders (recommendation 27).

Given the complex federal nature of activity in Germany, a useful funding model to consider may be a fixed “funding formula” which apportions funding support as a varying percentage across participating funding organisations. A funding formula for national government and federal states provides the budget for DFN. In the UK a funding formula also governs the relative funding contributions from regions to the budget for the JISC.

11. Future Development of the Strategy

11.1. Building blocks for the future

In considering the recommendations and approaches suggested in this study, there are a number which can be seen as initial building blocks towards the Strategy almost regardless of the final direction of travel. The consultation period required for developing the Strategy is likely to be significant as there is no single committee or funding source in Germany that can be engaged for support of the Strategy. In the interim, these building blocks can be pursued with relatively modest resources and extensions of existing work and contribute to the final Strategy adopted. We would recommend that they might be prioritised for the years 1 and 2 of the implementation. They are:

Content, Licensing, Perpetual Access, Preservation and Hosting

- Extensions to National, Regional, and Other Licences (recommendations 8-10)
- Develop Archive Content Selection Framework (recommendation 17)

Technical Infrastructure and Standards

- Establish panel for evaluation of Technical Options/R&D projects (recommendations 26)
- Short-term pragmatic options such as Dark Archive Escrows and extending access short-term on publishers' platforms (as a component of recommendation 18)

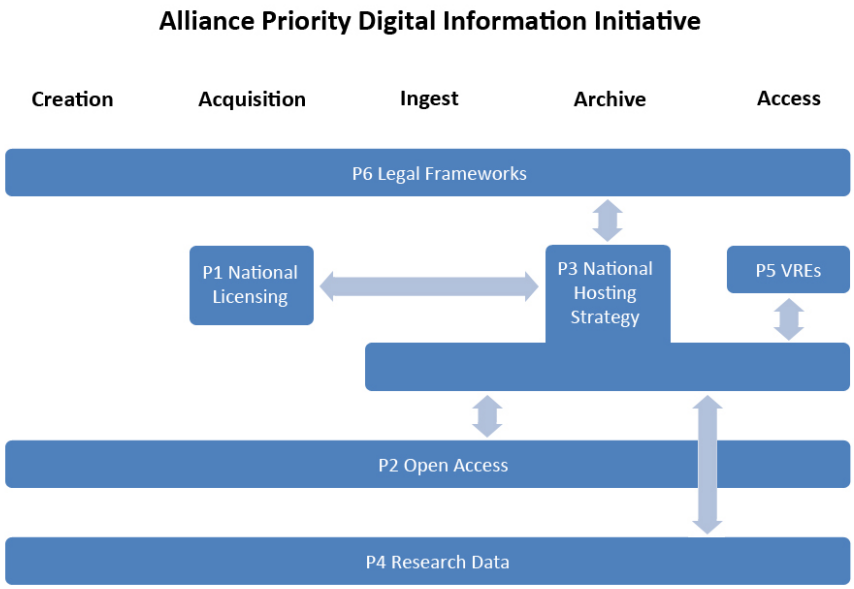
Business Models, Costs and Funding

- Develop Financial and Economic Models (recommendations 3 and 16)
- Certification Process and standards for Hosting Services and Archives respectively (recommendation 21)

Organisational and Policy Issues

- Consultation with all stakeholders (recommendation 2)
- Develop Governance and Participation (recommendation 3)
- Develop Advocacy and Standards (recommendation 7)
- Develop International Collaborations (recommendation 1)

11.2. Relationship to the Alliance Priority Information Initiative



Relationship of the "National Hosting Strategy" to the other Alliance Digital Information Priority Initiatives and Stages in Scholarly Information Workflow.

The recommendations of this study contribute to the establishment of a National Hosting Strategy for within the Alliance Priority Information Initiative. The recommendations build on the needs of the various types of Alliance member institutions and associations as well as on the needs of their respective scientific users of their information services. Furthermore, the recommendations establish a firm grounding for reflecting the strategy alternatives, reviewing available funding resources, taking a strategy decision, as well as specifying and developing a corresponding federated hosting, access support, and preservation infrastructure.

11.3. Recommendations to Other Alliance Working Groups

As the study results have also revealed requirements relevant for the work of the other Alliance working groups, some of the recommendations are targeted towards these working groups and therefore go beyond the scope of the responsibility of the working group that has been supervising this study. This holds especially true for areas related to licensing

templates as well as the general collaborative license demand analysis and areas such as supplementary research data.

11.4. Additional Future Content Types

Although asked to focus specifically on commercial e-journals and retro-digitisation materials we were also asked to consider how the Strategy might develop to cover other content types. To make sure that the implementation of an appropriate strategy covers all needed types of scientific information services and content, an implementation first of all has to thoroughly consider the different types of scientific information and content resources that are available. Within a requirements analysis for an infrastructure development phase these have to be clearly identified and appropriate policies as to how the different types of scientific information and content resources should be handled with respect to licensing, delivery, ingest, management, hosting, access, and especially preservation need to be defined. Furthermore, it seems reasonable to identify already existing competencies for handling the different content types in a systematic way.

We suggest the following approaches to content could be considered as part of the future development of the Strategy:

11.4.1 Databases

Databases cover a very broad field and the future strategy would need to distinguish between them. The main characteristic of databases versus journal collections is that the former contains a wide range of material of different types performing different functions for users. No one database is really like another and its structure and content are dependent on specific subject needs. The development of a schematic content map is recommended as a first stage. Existing standards for open database management systems and data interoperability should be investigated and effectively utilised. Future extensions to the national hosting strategy should include harmonisation and standardisation efforts in this area. A number of interviewees raised the issue of costs arising from this when considering local hosting. Databases illustrate the need to consider both the content and associated the features/services needed to access them in an electronic environment.

11.4.2 E-books

The label “e-book” is one of convenience covering many different types of resource, some of which would have value for perpetual access, whilst others were works of reference constantly updated where the latest version on the publisher’s server would be required by users. However, as these developments are still rather proprietary, a future extension to strategy needs to tackle this field systematically and make sure that existing open document formats are sufficiently utilised by the producers and publishers to avoid treating e-books as a special content type within the overall scientific information and content e-infrastructure. Many interviewees also noted the emerging and uncertain nature of the e-books market and business models for it and felt it would be too early to develop a strategy for them. The need to consider open access content such as e-theses that might be considered under e-book formats was also raised. The development of a schematic content map of scientific e-book materials and relative demands for their perpetual access and hosting can serve as an orientation within a first stage of preparation for their future inclusion in the Strategy.

11.4.3 Open Access Materials

As Open Access policies are becoming more and more successful in many scientific and research environments, these materials need to be considered as a special form of scientific information content resources by the strategy. Open access is being considered by a parallel working group off the Alliance and significant investment is being made in Germany in support of open access. Hybrid publication options for open access to journal articles will also means that a clear line may not exist between commercial content and open access content in specific future issues of journals. Despite the fact that these material are much more easily available (and can usually be pro-cured without any licensing cost) as compared to commercial scientific information and content resources, this does not mean that they are free of any intellectual property rights or permanently available by their own existence. This means that open access based materials need to be covered with almost the same care from a legal as well as from a technical and organisational point of view, as the material from commercially supported publishing channels. This means that for these materials appropriate agreements for the above outlined use cases need to be closed and managed within the infrastructure, too. As this is an emerging field and closely linked to key areas of

the Priority Initiative Digital Information and the Alliance working group on Open Access, it should also be covered by future extensions to the strategy.

11.4.4 Supplementary Material

As e-science methods are becoming more widely used, research data as well as many other forms of multi-media objects capturing relevant information and content are becoming more and more relevant parts of scientific information and publication processes. However, due to the dynamic and unpredictable nature of the underlying research processes and methodologies, their formats are quite often evolving very dynamically and are of a potentially complex nature. Supplementary materials are currently excluded from most archiving workflows for journal articles. As this is an emerging field and closely linked to key areas of the Priority Initiative Digital Information and the Alliance working group on Research Data, it should also be covered by future extensions to the strategy.

11.4.5 Conclusion

A conclusion that can be drawn from these comments is that future developments of the Strategy may need to be guided by a "content map" which provides a mechanism for distinguishing different types of databases and e-books and places commercial content into a context alongside open access materials. This would allow the strategy to focus on those materials where perpetual access or local hosting (e.g. to provide specialised services not available from a publisher) are seen as priorities. It would also allow work conducted for the Federated Strategy on Perpetual Access and Hosting to be placed in context of related work on Open Access.

Recommendation 30: The Strategy should segregate different content types and consider appropriate policies and implementations for them.

Recommendation 5: Future developments of the Strategy can be guided by a "content map" which provides a mechanism for distinguishing different types of databases and e-books and places commercial content into a context alongside open access materials.

Recommendation 20: Evaluate archiving practice and preservation and access services for supplementary materials to journal articles; and use of persistent identifier services for linking supplementary materials and research datasets associated with journal articles.

12. References

The Alliance of German Science Organisations, 2008, *Priority Initiative Digital Information*. www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/allianz_initiative_digital_information_en.pdf

Beagrie, N., Chruszcz, J. and Lavoie, B., 2008, *Keeping Research Data Safe: a cost model and guidance for UK Universities*, (Joint Information Systems Committee 2008). http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx

BSB/Armed Forces University, forthcoming, *From digitisation project to organised digital preservation: the development of business and organisational models for the long-term preservation of digital objects from DFG-sponsored digitisation projects.*

Consultative Committee for Space Data Systems (CCSDS), 2002, Recommendation for Space Data System Standards: Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1, Blue Book, January 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf

Consultative Committee for Space Data Systems (CCSDS), 2004, *Producer-Archive Interface Methodology Abstract Standard,* CCSDS 651.0-B-1 BLUE BOOK May 2004. http://public.ccsds.org/publications/archive/651x0b1.pdf

Davies, R. (ed), 2008, *The LIFE2 Final Project Report* http://eprints.ucl.ac.uk/11758/1/11758.pdf

Hilse, H.-W. and Kothe, J., 2006, Implementing Persistent Identifiers - Overview of concepts, guidelines and recommendations. NBN urn:nbn:de:gbv:7-isbn-90-6984-508-3-8.

JISC, Not Dated a, *JISC Guidelines for Publishers*, JISC and the Information Environment Fact file series.

www.jisc-collections.ac.uk/about_collections/publisher_information/coll_jiscfactfile.aspx

JISC, Not Dated b, *JISC Fact File: Managing Content Provision*.

www.jisc-collections.ac.uk/about_collections/publisher_information/coll_jiscfactfile/coll_factcards_mcprov.aspx

Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen (KOPAL), 2006, Universelles Objektformat - Ein Archiv- und Austauschformat für digitale Objekte. http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf

Lavoie, B., 2004, *The Open Archival Information System Reference Model: Introductory Guide*, Digital Preservation Coalition Technology Watch Report 04-01 January 2004. http://www.dpconline.org/docs/lavoie_OAIS.pdf

Maron, N.L. et al, 2009, *Sustaining Digital Resources: An On-the Ground View of Projects Today*. Ithaka Case Studies in Sustainability. http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_with_CaseStudies.pdf

OECD, 2008, *Science, Technology and Industry Outlook* (OECD Publications 2008).

Paskin, N., forthcoming, *Digital Object Identifier (DOI) System* in Encyclopedia of Library and Information Sciences, Third Edition. Taylor & Francis, DOI 10.1081/E-ELIS3-120044418. See http://www.doi.org/overview/DOI-ELIS-Paskin.pdf

Rightscom, 2008, *E-Journal Mapping Study*. Final Report to British Library Legal Deposit Advisory Panel 29 August 2008.

UK Serials Group (UKSG), 2008, *TRANSFER Code of Practice: Version 2.0*, September 2008. http://www.uksg.org/Transfer/Code

Weber, S., 2009, *National Licenses in Germany: Lessons learned from five years of central funding in a Federal Republic*. http://www.hebis.de/hebis-konsortium/praesentationen/Konferenzbericht_20-04-2009.pdf?PHPSESSID=2ce7b7d954fac3bd0973003f2d42d0dd

Witthaut, D. et al, 2005, Digitalisierung und Erhalt von Digitalisaten in deutschen Museen. Nestor Studies 2. http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20041223022

13. APPENDIX 1- GLOSSARY AND DEFINITION OF TERMS

ACCESS

There are varying degrees of access in the digital world. Full **subscriber access** refers to the ability of a subscribing organisation or a **consortium** of organisations and their patrons to make use of the entire range of electronic materials which has been licensed from a commercial publisher for the specific use of the registered members of that organisation during the licensing period. Most registered users will access electronic materials from a workstation within the organisation itself or by remote access using a designated librarian-supplied username and password or a via a specified authentication login such as Athens or Shibboleth. **Guest access** refers to the ability of users other than those belonging to a subscribing organisation to have partial access to the electronic materials of a publisher such as tables of contents, abstracts, etc.

Open Access (OA) literature is digital, online, free of charge to the user and free of most copyright and licensing restrictions. Traditionally most of this literature has been published solely in scholarly journals. What makes it possible is the internet and the consent of the author of the material or the copyright holder to make it available in this way. OA publishing removes price barriers (subscriptions, licensing fees, pay-per-view) and permission barriers which have previously restricted access solely to those with subscriptions to electronic journals and which have limited legitimate scholarly uses. The BBB (Budapest-Bethesda-Berlin) definition of OA embodies the above. OA is now practised in a variety of forms in different types of published works as commercial publishers have evolved new systems. Variations include Gold OA, Hybrid OA, Delay OA and Green OA.

ARCHIVE

An archive is a repository created to preserve material no longer in heavy current usage but nevertheless still needing to be accessed on occasions or for specific reasons. Three distinct terms to describe grades of digital archive for e-journals are in common use. A **dark archive** is an archive that cannot be accessed by any current users but may be accessible at future dates subject to the occurrence of specific pre-defined events ("**trigger event**"). Access to the data is either limited to a few set individuals or completely restricted to all. Typically

these dark archives can be divided into two main types: **type 1** - those only providing a form of escrow or “bit preservation” of content that is suitable as a short-medium term solution for guaranteeing access; and **type 2** - those providing the bit preservation of the content plus some degree of associated services for future access (this may include a “back-up” access service should a primary access service fail for any reason, and digital preservation planning and preservation action services such as file format migration which will keep the content accessible in the future). These provide for and can help guarantee long-term perpetual access. A **dim archive** provides bit preservation for the content plus digital preservation planning and actions for long-term perpetual access, and also limited current access (perhaps limited to on-site users or previous subscribers post-cancellation, etc); A crucial concept behind many archives is agreeing to restrict access over a period of time or until specific events occur in the future. At such points “dark archives” and “dim archives” may transform to a **light archive** that can be accessed by many authorised users. Access to the information is open to all members of the 'community' that has a need for the information. Access may be subject to access restrictions agreed upon by the publisher of the material and the archive. Current access under some circumstances would always be presumed and an access system would be maintained.

BACKFILE

That portion of an electronic journal title which is not included in the subscription to the current year of that journal and a number of designated years prior to the current year is known as the backfile. A specific cut-off date is used by many publishers to designate the difference between backfile and current content. Frequently the backfile can be purchased as a one-off but still under license and can contain all the digitised volumes from the start of the journal to the cut-off date. Cumulative backfiles for a group of electronic journal titles by the same publisher may be offered with a common cut-off date. Supplementary backfiles may be offered subsequently. Both may be subject to an ongoing access fee if accessed from a publisher's server.

CONSORTIUM

A group of individuals, libraries or organisations with common interests formed to undertake an enterprise or activity that would be beyond the capabilities of the individuals, libraries or organisations on their own. In the context of electronic materials it normally refers to a group (regional, national or international) that undertakes the negotiations for the purchasing and licensing of such materials on behalf of the constituent members of that group.

Continuing Access - See PERPETUAL ACCESS

Dark Archive - See ARCHIVE

DIGITAL PRESERVATION

Digital preservation is a term used to cover the series of managed activities necessary to ensure reliable access to digital materials for as long as necessary and beyond the limits of storage media failure or technological change.

Dim Archive - See ARCHIVE

HOSTING

Hosting refers to the service used for primary access to content. **Local hosting** of electronic materials is the holding of data files provided by a publisher under an agreement with the subscriber on a local server under the control of the subscriber themselves or a designated organisation, other than the publisher, working in partnership with the subscriber. It is the alternative to access/hosting via the publisher's own servers. Local hosting provides archival protection and rights to the resources under local control. Control of the local server is under the subscriber who must provide the labour to maintain it and also maintain access controls to the content. The service may encompass elements of **digital preservation** and an **archive** but this is usually not its primary function and it may be less comprehensive in these areas than dedicated services. In some cases institutions may co-operate to federate local hosting to regional or potentially national scale.

HOSTING SERVICE

Hosting service(s) provide the point(s) of full current access to authorised users and ongoing maintenance, updating, and security of that content (including routine back-ups and disaster recovery planning). These may provide for and help guarantee short-medium term horizons for access but are not focussed on digital preservation or organised for the long-term and would need to partner with or be combined with other services to achieve or guarantee longer-term perpetual access.

LICENCE

A licence is a legal document giving official permission to undertake an activity and is granted by one party (the licensor) to another party (the licensee) as an element of an agreement between these parties. In the digital world it is the authorisation for the licensee to provide and their users to use licensed electronic materials supplied by the licensor to the subscribing organisation. It is governed by contract law and may include specific elements determined by the laws of the issuing country. It is normally time limited to the subscription period, and needs to be renewed at given periods. A site license is a particular type of licensing agreement that permits access to and use of digital information at a specific geographical site or location agreed by both parties and under specific conditions and groups of users. Licences may contain clauses covering, a designated **archive**, **hosting**, **perpetual access**, **digital preservation** and **transfer**.

Light Archive - See ARCHIVE

Local Hosting See HOSTING

Open Access – SEE ACCESS

PERPETUAL ACCESS

Perpetual access refers to the right of the subscriber and their users to have ongoing permanent access to electronic materials which have already been leased and paid for by the subscriber from a publisher. It is a term used along with its synonyms, continuing access and post-cancellation access, in the information industry to describe the ability to retain access to electronic materials by the subscriber/licensee after the contractual licensing agreement with the publisher/licensor for those materials has ended, whatever the reason for the cessation. It may also cover as appropriate arrangements for **digital preservation** needed to guarantee some elements of perpetual access. The subscriber/licensee and the publisher/licensor, both of whom are party to the licensee, need to agree terms for the granting of perpetual access rights to the subscriber. It may refer to both the leasing of current content and the purchase of backfiles. For some publishers the payment of an ongoing annual access fee to cover remote hosting costs is required. An alternative to ongoing access via the publishers servers post-cancellation may be supplying a copy of the content to the licensee for local **hosting** or a third-party **archive** that also provides this service.

Post-Cancellation Access - See PERPETUAL ACCESS

TRANSFER

In the electronic publishing world 'transfer' refers to the change of publisher or owner which can occur for an electronic journal title and the consequent roles and responsibilities which must be adopted by the 'transferring publisher' and the 'receiving publisher' respectively, to ensure the continued access to content, both currently and in perpetuity, for the subscriber to that journal and their users. A TRANSFER Code of Practice has been created, now in Version 2 (UKSG2008), which seeks to set acceptable operational standards and publishers are encouraged to sign up to this code and become TRANSFER Compliant.

TRIGGER EVENT

This terminology is used when specific conditions relating to an electronic journal title and its continued delivery to subscribers are met. If the journal is no longer available to the users or subscribing organisation from the publisher or any other source for a variety of reasons then a trigger event is said to have occurred. They can set in motion access for users via an archive where the electronic journal may be digitally preserved. Examples of trigger events include four from the publisher's side and one from the subscriber's side - a publisher ceases operations, a publisher ceases to publish a title, a publisher no longer offers back issues or there is a catastrophic and sustained failure of a publishers delivery platform or the subscriber/licensee for economic or scholarly reasons cancels their subscription to the journal/s. 'Pain Reviews' triggered such an event in July 2009.

14. APPENDIX 2 - LIST OF INTERVIEWEES

Bavarian State Library: Werner Baur, Leibniz Supercomputing Centre Tobias Beinert, Digital Library Dr. Markus Brantl, Head of the Munich Digitisation Centre/Digital Library Dr. Klaus Ceynowa, Deputy General Director of the Bavarian State Library Matthias Groß, Bavarian Library Network Anna Kugler, Digital Library Dr. Hildegard Schaeffler, Head of Serials and E-Media Dr. Astrid Schoger, Digital Library/Head of long-term preservation unit.

Berlin State Library: Matthias Kaun, Director East Asian Collections , Acting Director Oriental Collections / with the help of Friederike Glaab, Sussane Maier (Collection Development Department) Dr. Holger Busse (IT-Department/Digital Library)

British Library: Richard Boulderstone, Director of e-Strategy

BMBF: Susanne Clobes, Head of Division Science System Policy Issues

B-WK: Dr. Antje Kellersohn, Head of Library, Hans-Adolf Ruppert, Head of IT Department, Dr. Michael Becht, Central Coordination of Consortium Baden-Württemberg

CLOCKSS/LOCKSS: Victoria Reich (Director, LOCKSS Program)

CSIRO: Cynthia Love, Library Services Manager; Paul Reekie, General Manager, CSIRO Publishing

DFG: Dr. Anne Lipp, Head of Scientific Libraries and Information Systems Division

DNB: Ute Schwens, Deputy Director, Reinhard Altenhöner, Head of IT

FAK: Michael Rieck (Acquisitions University of Potsdam + FAK), Helmut Voigt (Acquisitions Humboldt University + FAK), Bettina Golz (Acquisitions Berlin Technical University + FAK),

Uta Kaminsky (FAK office)

FIZ Karlsruhe: Sabine Brünger-Weilandt, President & CEO, Dr. Leni Helmes, Vice President Development and Applied Research, Dr. Karl-Heinz Weber, Senior Expert e-Science Co-operations

Frankfurt University Library: Sylvia Weber License Negotiator Frankfurt University Library/HeBIS Consortium/DF national licences, Berndt Dugall, Director Frankfurt University Library

Fraunhofer: Martina Lillmanntöns, Fraunhofer-Gesellschaft e.V. (FhG), Zentrale, Abt. C9, Informationsmanagement, Klaus Greschek, FhG, Birlinghovener Informationsdienste (BID) / Birlinghoven Library Competence Center Fraunhofer Informationsdienste (CC-FID)

Freiburg University: Prof. Gerhard Schneider, Chief information Officer and Director of the Computing Centre

GBV: Reiner Diedrichs, CEO

Göttingen State and University Library: Kristine Hillenkötter Head of Acquisitions Dept and Current Head of Lower Saxony Library Consortium, Arnulf Timm – Head of Acquisitions Dept, Frank Klaproth Head of Digital Library Dept ,Stefan Farrenkopf Deputy-Head of Digital Library Dept., Claudia Rölleke - Head of Periodicals and Digital Media.

GWDG: Burkhard Heise - 1st Head of Applications and Information Systems group, Oswald Haan - 2nd Head of Applications and Information Systems group, Anke Bruns - Care of library and web hosting, Thomas Lindeman - member of Applications and Information Systems group

HathiTrust: Jeremy York, Project Librarian, HathiTrust Digital Library

Hbz/GASCO: Hans Ollig, HBZ, Director, Dr. Silke Schomburg, HBZ, Vice/Deputy Director, Dr. Jochen Johannsen, HBZ, Manager Content, Kathrin Gitmans, HBZ, Manager ePublishing Systems, Martin Iordanidis, HBZ, ePublishing Systems, Anette Seiler, HBZ, Digitisation Projects, Werner Reinhardt, Library Director; Organizer of E-Journal-Consortia in North-Rhine Westphalia; Chair of GASCO

Helmholtz: Dr. Korinna Bauer, Executive Assistant to the President of the Helmholtz Association, Roland Bertelmann, Head of the “Library Wissenschaftspark Albert Einstein” at the German Research Centre for Geosciences, Potsdam, Dr. Hans-Jürgen Goebelbecker, Head of the Library and Central Media Department of the Forschungszentrum Karlsruhe

Humboldt University: Prof. Dr. Peter Schirmbacher, Director, Computer and Media Service and Professor at Department of Library and Information Science

JISC Collections: Lorraine Estelle, Chief Executive

KB: Hans Jansen, Director e-Strategy

LANL: Miriam Blake Director LANL Research Library

Max-Planck: Ralf Schimmer, Head of Department of Scientific Information Provision, Margit Paltzenberger, Information Management, Department of Scientific Information Provision, Prof Dr. Wolfgang Schön, MPG Vice-President

National Library of Scotland: David Dinham Head of Information Systems

National Library of New Zealand: Steve Knight, Associate Director National Digital Library & Programme Architect National Digital Heritage Archive Digital Innovation Services

OhioLink: Anita Cook, Director of Library Systems; Celeste Feather, Assistant Director, Licensing

Portico: Eileen Fenton (Executive Director), Amy Kirchhoff (Archive Services Product Manager)

Scholars' Portal , University of Toronto: Kathy Scardellato, Executive Director Ontario Council of University Libraries and Alan Darnell, Scholars Portal Manager

TIB: Dr. Irina Sens, Deputy Director and Markus Brammer, Head of TIB Licensing Team and Goportis Publisher Licensing Coordinator

University of Edinburgh (PECAN and PEPRS projects): Peter Burnhill Director of Edina.

ZIB/KOBV: Dr. Thorsten Koch, Zuse Institute Berlin, Head of the Dept. Scientific Information Systems and Head of the KOBV Zentrale Monika Kuberek, Zuse Institute Berlin, Deputy Head of the KOBV Zentrale

15. APPENDIX 3 SUPPLEMENT ON TECHNICAL STANDARDS

15.1. INTRODUCTION

Effective, reliable interaction between different services and systems are dependent on the adoption of internationally agreed standards. Storing material in widely adopted, standardised formats also increases the likelihood of successful future migration necessitated by the introduction of new systems and standards. This appendix describes and elaborates some of those most relevant to a perpetual access strategy.

15.2. REFERENCE MODELS

15.2.1 The Open Archival Information System (OAIS) Reference Model

An Open Archival Information System (OAIS) is an archive consisting of an organisation of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. The OAIS reference model is an ISO standard for Open Archival Information Systems defined by a recommendation of the Consultative Committee for Space Data Systems (CCSDS 2002).

A key point to note is that OAIS does not specify or even conceptually suggest any specific interfaces and protocols to support design and implementation of its functional entities responsible for bitstream encoding during ingest, transcoding during archival and preservation as well as adaptation and decoding during access.

Lavoie (Lavoie 2004) identified two areas of activity related to the OAIS model requiring deeper consideration: metadata requirements associated with the long-term preservation of digital materials; and attributes of a trusted digital repository. In addition to them, another key domain of development is the Producer-Archive Interface Standard (PAIS) which led to a new international standard (CCSDS 2004). PAIS was developed as a recommendation identifying, defining and providing structure to the relationships and interactions between a Producer and an Archive. It identifies four phases in the process of transferring information, suggests actions which should be carried out during each phase, and provides a general framework which facilitates the identification and/or development of standards and software tools to be used within the ingest process. This example illustrates the growing

acknowledgement of the importance of the pre-ingest collaboration between the producer and the archive.

15.3. OBJECT STRUCTURES, METADATA, AND ENCODING STANDARDS

Current and emerging standards that are relevant include:

NLM-DTD – The National Library of Medicine Journal Archiving and Interchange Tag Suite. This was developed to provide a common format for exchanging data between publishers and archives. The Suite provides a set of XML schema modules that define elements and attributes for describing the textual and graphical content of journal articles as well as some non-article material such as letters, editorials, and book and product reviews. These are the "header files" that accompany content and describe and define that content. In 2006 the British Library and the US Library of Congress announced their support for the migration of content to this standard. For more information see http://dtd.nlm.nih.gov/ .

PDF and PDF/A – Adobe's Portable Document Format, created in 1993, has become the ubiquitous encoding for e-published journals etc. Formerly a proprietary format, it was released as an open standard in 2008 and published by ISO. **PDF/A** is a file format for the long-term archiving of electronic documents. It is a subset of PDF, leaving out features not suited to long term archiving. No audio or video content is allowed, JavaScript and executables are prohibited, all fonts must be embedded and be legally embeddable for unlimited universal rendering, and use of standards based metadata is mandated. The embedding of fonts is a potential disadvantage as the repeated storage of identical information with every PDF file, increases filespace requirements. For more info see http://www.pdfa.org/ .

TIFF – Tagged Image File Format is a file format for storing raster images (as opposed to vector images), including line art. Originally created by a company called Aldus as a standard for the output of image scanners, it is now under the control of Adobe Systems. The current version 6.0 specification was released in 1992 and, although there have been some minor extensions; it has not had a major update since then. TIFF is a flexible, adaptable file format for handling images and data within a single file, by including the header tags (size, definition, image-data arrangement, applied image compression) defining

the image's geometry. Image data can be uncompressed, compressed using a lossless compression algorithm (LZW) or lossy (JPEG compression). It is still a popular format for storing high quality images, but without compression file sizes tend to be very large. The full specification can be found at http://partners.adobe.com/public/developer/tiff/index.html .

JPEG and **JPEG 2000** – The JPEG acronym comes from the name of the group that issued the standard in 1992, the Joint Photographic Experts Group. It is a lossy compression method which uses a compression process known as quantisation to simplify images. The degree of simplification is variable, the trade-off being between image quality and file size. In all cases the process is irreversible, with information being discarded.

JPEG2000 is an updated version of the standard using a wavelet-based image compression standard (as opposed to the original discrete cosine transform based JPEG standard). It has several advantages over the original standard including a more flexible codestream allowing images to be displayed with gradually increasing refinement. The standard allows for both lossy and lossless encoding. Since its publication it has only grown very slowly in popularity, partly because it requires more intensive processing to encode/decode. It is said to produce better quality compressed images than the original standard.

The Wellcome Library has recently (21/9/09) announced the adoption of JPEG2000 for its digitisation programme, specifically the "visually lossless" lossy compression variant to gain at least 75% storage savings in comparison to TIFF files.

There is a useful Library of Congress analysis of JPEG 2000 at http://www.digitalpreservation.gov/formats/fdd/fdd000143.shtml. The JPEG committee page is at http://www.jpeg.org/

PREMIS – PREservation Metadata: Implementation Strategies – was a working group sponsored in 2003 by OCLC and RLG whose goal was to define implementable, core preservation metadata, with guidelines/recommendations for management and use. In 2005 it released the “Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group”. It defined preservation metadata as “information that supports and documents the digital preservation process” including

- Provenance: Who has had custody/ownership of the digital object?
- Authenticity: Is the digital object what it purports to be?

- Preservation Activity: What has been done to preserve the digital object?
- Technical Environment: What is needed to render and use the digital object?
- Rights Management: What intellectual property rights must be observed?

More information at http://www.loc.gov/standards/premis/ .

MIX – The Metadata for Images in XML (MIX version 2.0) is an XML schema for a set of technical data elements. It is developed specifically to manage digital image collections by the Library of Congress' Network Development and MARC Standards Office, in partnership with the NISO Technical Metadata for Digital Still Images Standards Committee. MIX can be use to encode basic information on file types and sizes, to details of image capture and processing of image after capture. The concept of keeping record of subsequent processing of the original images should be especially analysed from the point of view of re-use of digital material; for the purposes of reuse, the storage of the new object with a record of the changes is essential. More information at http://www.loc.gov/standards/mix/ .

UOF - The Universal Object Format (UOF) (KOPAL 2006) was defined within the German project Kopal to describe a flexible way to construct packages as Submission Information Packages (SIP) or Dissemination Information Packages (DIP) in the terminology of the OAIS reference model (CCSDS 2002). A package compliant to UOF is a packed file (e.g. ZIP or tar) including all the content files of one Information Package and a METS file of a specific METS profile. The METS file in a UOF package must contain LMER metadata in the Administrative Metadata section and list references to all files of the package in the File Section. There are no restrictions on the kind of metadata schema used within the Descriptive Metadata section. The UOF is meant to bundle the files of one logical object together with all metadata needed to enable digital preservation strategies. Therefore, the LMER metadata should provide specific technical metadata and the history of changes like file format migrations.

Besides the already introduced standards, various metadata schemata have been developed for a variety of user environments and disciplines. Some of the most common and relevant ones are presented in a reduced, i.e. selected overview following below. More information about LMER at http://www.d-nb.de/eng/standards/lmer/lmer.htm .

MPEG-7 (ISO/IEC JTC1/SC29/WG11N6828) is a standard for describing the multimedia content data. MPEG–7 is not aimed at any particular application; rather, the elements that MPEG–7 standardises support as broad a range of applications as possible. MPEG–7 uses a Description Definition Language (DDL) based on XML. The Description Schemata (DSs) are metadata structures for describing and annotating audio-visual (AV) content. The DSs provide a standardised way of describing in XML the important concepts related to AV content description and content management in order to facilitate searching, indexing, filtering, and access. More information at http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.

MPEG-21 (Multimedia Framework (ISO/IEC JTC1/SC29/WG11/N5231) was developed to address the need for an overarching framework to ensure interoperability of digital multimedia objects. More information at http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

ONIX (ONline Information eXchange) is an international standard for representing book, and serial product information in electronic form. Many on-line book traders such as Amazon and Barnes & Noble use this metadata standard to transfer information about their products. More information at http://www.editeur.org .

Dublin Core Metadata Initiative (DCMI). The Dublin Core was developed to be simple and concise, and to describe Web-based documents. From its start in 1995 it evolved into a leading initiative for improving resource discovery on the Web. The description of the information resources is created using fifteen Dublin Core elements. They consist of strings or labels, which are paired to a value. The value may consist of free text or it may be taken from a standardised resource. The use of the Dublin Core is not bound to strict rules. All elements are optional, and may be repeated without any constraint. The description may reside in a separate file or it may be a part of the information resource itself. The DCMI is represented on the Web at http://dublincore.org/ .

Metadata Encoding and Transmission Standard (METS). 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. 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. More information at http://www.loc.gov/standards/mets/ .

Metadata Object Description Schema (MODS) is a descriptive metadata schema that is a derivative of MARC 21 and intended to either carry selected data from existing MARC 21 records or enable the creation of original resource description records. It includes a subset of MARC fields and uses language based tags rather than the numeric ones used in MARC 21 records. MODS is expressed using the XML schema language. Although the MODS standard can stand on its own, it may also complement other metadata formats. Because of its flexibility and use of XML, MODS may potentially be used as an extension schema to METS, a metadata set for harvesting, and for creating original resource metadata. More information at http://www.loc.gov/standards/mods/. The MARC formats and standards reside at http://www.loc.gov/marc/ .

One popular example of a technological solution which allows metadata to be embedded into a file is Adobe's Extensible Metadata Platform (XMP). Note that a metadata standard is not metadata per se, but a specific model and/or schema to represent metadata. Consequently, it is possible to have different instances and implementations of the same standard. To address the existence of multiple metadata standards, since 1994 ISO has started to develop ISO/IEC 11179, Information Technology – Metadata registries (MDR) (with various parts released in subsequent years). The XMP is described at http://www.adobe.com/products/xmp/ .

15.4. PERSISTENT IDENTIFIERS AND SERVICES

In the following, we will introduce the two most important Persistent Identifier standards – DOI (Paskin, N., forthcoming), and URN, as well as a comparison of them based on Factsheet DOI System and Internet Identifier Specifications Version 2.1.

DOI – The Digital Object Identifier (DOI) System is a managed system (service) for persistent identification of content on digital networks. It can be used to identify physical, digital, or abstract entities. The identifiers (DOI names) resolve to data specified by the registrant, and use an extensible metadata model to associate descriptive and other elements of data with the DOI name. The DOI system is implemented through a federation of registration agencies, under policies and common infrastructure provided by the

International DOI Foundation (IDF) which developed and controls the system. The DOI system has been developed and implemented in a range of publishing applications since 2000; by early 2009 over 40 million DOIs had been assigned. The DOI system provides identifiers which are persistent, unique, resolvable, and interoperable and so useful for management of content on digital networks in automated and controlled ways.

The management of content on digital networks requires identifiers to be persistent, unique, resolvable, and interoperable. Persistence can be considered to be interoperability with the future. There have been a number of efforts to address the need for such reliable identifiers, notable among them URN and URI specifications; however these do not of themselves provide an implemented managed scheme and registry for specific content sector applications. Such full schemes require more: a model for identifiers and their management; shared, standards-based, persistent identifier management infrastructure; support for adoption of persistent identifiers and services, and a plan for sustainable shared identifier infrastructure. The DOI system is represented at http://www.doi.org/ .

URN - A Uniform Resource Name (URN) is a Uniform Resource Identifier (URI) that uses the urn scheme, and does not imply availability of the identified resource. Both URNs (names) and URLs (locators) are URIs, and a particular URI may be a name and a locator at the same time.

The Functional Requirements for Uniform Resource Names are described in RFC 1737 (http://www.ietf.org/rfc/rfc1737.txt).The URNs are part of a larger Internet information architecture which is composed of URNs, Uniform Resource Characteristics (URCs), and Uniform Resource Locators (URLs). Each plays a specific role:

- URNs are used for identification,
- URCs for including meta-information.
- URLs for locating or finding resources.

Uniform Resource Names (URNs) are intended to serve as persistent, location-independent resource identifiers and are designed to make it easy to map other namespaces (that share the properties of URNs) into URN-space. Therefore, the URN syntax provides a means to encode character data in a form that can be sent in existing protocols, transcribed on most keyboards, etc.

A URN is like a person's name, while a URL is like their street address. The URN defines something's identity, while the URL provides a method for finding something: essentially, "what" vs. "where". Many namespaces have been registered for URN, of which the National Bibliographic Numbers (NBN) namespace, RFC 3188 (http://www.ietf.org/rfc/rfc3188.txt), is the most noteworthy approach aiming at persistent resource identification (Hilse and Kothe 2006).“This namespace has been designed to allow national libraries to integrate their identification concepts into a common URN namespace.... NBN is a namespace which is exclusively assigned to national libraries. The global registry for the URN:NBN namespace is the Library of Congress (http://www.loc.gov/). All National Libraries are responsible for sub-namespaces that are expressed by [country codes]....Several national libraries developed their own NBN-URN-based systems in the context of national and international research projects, and several implementations are already in practical use.... Some national libraries have established data exchange between each other to facilitate resolving of other national libraries' NBN-URNs. However, there is no central resolver for all possible NBN-URNs.” (Hilse and Kothe 2006).

15.5. COMPARING GENERIC IDENTIFIER STANDARDS

A DOI name differs from commonly used Internet pointers to material such as the URL, because it identifies an object as a first-class entity, not simply the place where the object is located. The DOI System also differs from standard identifier registries such as the International Standard Book Number (ISBN), International Standard Recording Code (ISRC), etc., because it can be associated with defined services and is immediately actionable on a network. The comparison of persistent identifier approaches is difficult because they are not all doing the same thing. Imprecisely referring to a set of schemes as 'identifiers' doesn't mean that they can be compared easily. For further discussion and comparison see the DOI factsheet at http://www.doi.org/factsheets/DOIIdentifierSpecs.htm

DOI Registration Agencies – Worldwide, seven registration agencies (RAs) provide services for the allocation of DOI name prefixes, the registration of DOI names and the declaration and maintenance of metadata and state data. Most RAs have chosen to provide value added services on top of the basic naming service. Their application profile and terms of business, however, can prevent registrants from arbitrarily choosing among the registration

agencies. For example, the Publications Office of the European Union (OPOCE) acts as publisher for European Union entities exclusively, the offer of German National Library of Science and Technology (TIB) does not target commercial registrants, CrossRef expects its members to register their complete portfolio, and Nielsen BookData requires registered objects to possess traditional book-trade identifiers like EAN or ISBN already. Also, the service charges vary and are calculated based on annual publication-related turnover, the amount of registered identifiers, and other factors.

Located in Germany, Marketing- und Verlagsservice des Buchhandels GmbH (MVB) and German National Library of Science and Technology (TIB) offer DOI registration services.

The **Marketing- und Verlagsservice des Buchhandels GmbH (MVB)** is the local partner of Multilingual European DOI Registration Agency (MEDRA). As a subsidiary of the Association of German Publishers (Gesamtverband Börsenverein des Deutschen Buchhandels), MBV offers to booksellers and publishers' publications and services supporting book marketing and market orientation. Besides DOI registration, ISBN registration and German Books in Print (Verzeichnis lieferbarer Bücher, VLB) are further core products. At last, MEDRA provides the technological platform for other DOI registration agencies.

The **German National Library of Science and Technology (Technische Informations-bibliothek, TIB)** is the German National Library for all areas of engineering as well as architecture, chemistry, information technology, mathematics and physics. The TIB's task is to comprehensively acquire and archive literature from around the world pertaining to engineering and the natural sciences. TIB became a non-commercial DOI registration agency for research data sets from the fields of technology/science and medicine. Additionally, the TIB also facilitates the registration of any kind of scientific contents resulting from publicly funded research in Europe."

URN Management – "The NBN namespace has no commercial background, but it is the sovereign territory of national libraries." (Hilse and Kothe 2006). DNB manages the National Bibliographic Numbers for Germany, namespace nbn:de.

The **German National Library (Deutsche Nationalbibliothek, DNB)** is the central archival library and national bibliographic centre for the Federal Republic of Germany. In general, its function is to collect all media works published in Germany, to permanently safeguard the

holdings, and to provide access to the general public. By law, any party entitled to publish or distribute media works residing in German is subject to mandatory legal deposit at DNB. Each net publication accepted gets identified by NBN. By default, the library assigns the identifier during ingest. Depositors can opt to create and assign URNs themselves.

“The German National Library provides the basic organisational and technical conditions to assign, manage and resolve URNs to all users of URNs in the namespace nbn:de. ... URNs will be referenced in national and international reference systems such as bibliographies, catalogues and search engines and can be transported across bibliographic exchange formats. “(http://www.persistent-identifier.de/?link=3352&lang=en).

DNB also operates the resolving infrastructure for Austria, namespace nbn:at, and Switzerland, namespace nbn:ch.