SEVENTH FRAMEWORK PROGRAMME

FP7-ICT-2009-6

BlogForever

Grant agreement no.: 269963

BlogForever: D3.2 Interoperability Prospects

Editor:	Hendrik Kalb, Paraskevi Lazaridou, Matthias Trier
Revision:	1
Dissemination Level:	Public
Author(s):	Paraskevi Lazaridou, Hendrik Kalb, Ed Pinsent, Yunhyong Kim, Vangelis Banos, Seamus Ross, Stella Kopidaki, Lea Berninger
Due date of deliverable:	30 April 2013
Actual submission date:	Public
Start date of project:	01 March 2011
Duration:	30 months
Lead Beneficiary name:	Technische Universität Berlin

Abstract: This report evaluates the interoperability prospects of the BlogForever platform. Therefore, existing interoperability models are reviewed, a Delphi study to identify crucial aspects for the interoperability of web archives and digital libraries is conducted, technical interoperability standards and protocols are reviewed regarding their relevance for BlogForever, a simple approach to consider interoperability in specific usage scenarios is proposed, and a tangible approach to develop a succession plan that would allow a reliable transfer of content from the current digital archive to other digital repositories is presented.

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)

The **BlogForever** Consortium consists of:

Aristotle University of Thessaloniki (AUTH)	Greece
European Organization for Nuclear Research (CERN)	Switzerland
University of Glasgow (UG)	UK
The University of Warwick (UW)	UK
University of London (UL)	UK
Technische Universitat Berlin (TUB)	Germany
Cyberwatcher	Norway
SRDC Yazilim Arastrirma ve Gelistrirme ve Danismanlik Ticaret Limited Sirketi (SRDC)	Turkey
Tero Ltd (Tero)	Greece
Mokono GMBH	Germany
Phaistos SA (Phaistos)	Greece
Altec Software Development S.A. (Altec)	Greece

History

Version	*Date*	*Modification reason*	*Modified by*
0.5	22/04/2013	Draft version	TUB, UL, AUTH, UG
0.6	23/04/2013	Extension of Section BlogForever succession plan, and updates in the introduction	UG, TUB
0.7	29/04/2013	Extensions and Adaptations according to the internal reviews	TUB, UL, UG, AUTH
1.0	30/04/2013	First version of the deliverable	TUB, UL, UG, AUTH
1.1	09/07/2013	Addressing comments of the BlogForever internal review process	TUB, UL, UG
1.2	01/08/2013	Including Delphi Study final report	TUB

Table of Contents

Executive Summary

The aim of this report is the evaluation of interoperability prospects of the BlogForever platform with 3rd parties. Thus, the future adoption and deployment of the BlogForever platform, e.g. by libraries, should be facilitated. While interoperability is a domain with a multitude of facets, with the following report, the BlogForever project addresses the following questions:

- What are current and future interoperability challenges in the web archiving domain?
- What are the relevant technical interoperability standards and protocols for BlogForever?
- How can a specific institution proceed to establish and maintain interoperability of their BlogForever implementation?
- How can sustainability for collections based on the BlogForever system be established?

To answer the questions the following research activities has been conducted and are described in detail in the report:

- Reviewing interoperability models in the existing literature,
- Conducting a Delphi study to identify crucial aspects for the interoperability of web archives and digital libraries,
- Examining interoperability standards and protocols regarding their relevance for the BlogForever platform,
- Proposing a simple approach to consider interoperability in specific usage scenarios, and
- Presenting an approach to develop a succession plan that would allow a reliable transfer of content from the current digital archive to other digital repositories.

The single steps represent also the structure of the report and the findings contribute in several ways:

Interoperability is still an ambiguous concept that can be interpreted either in a narrow sense that focuses on technical aspects, or a more comprehensive perspective that include non-technical aspects like organisational and legal constraints. Therefore, section 2 summarizes several conceptualizations that have been proposed about interoperability models and levels and provides an extensive overview of the current literature on interoperability.

A Delphi study, conducted within the framework and the needs of the project, about aspects of interoperability of web archives and digital libraries is presented in section 3. The study reveals remarkable insights regarding current problems, limitations, needs and challenges that are encountered in today's interoperations (or efforts to this direction) among systems of the web archiving and digital library communities. It contributes to the limited so far empirical research for interoperability, presenting the current barriers but as well suggestions for future approaches, and can be a useful study for the three involved communities: the web archiving, the digital library and the digital preservation community.

An extensive review of the standards that support, assist or establish interoperability is provided in section 4. Standardisation is probably the most essential aspect of interoperability since standards can be the bridge between two or more different environments. The review provides a useful guide about the most commonly used standards that address the needs for interoperation mainly in the domains of digital libraries and web archives since these are the most relevant systems for BlogForever to interoperate with.

Section 5 presents a 5-step approach to consider and configure the enabling of interoperability of the BlogForever system (or any digital library) with another potential information system. This approach offers a useful and concrete guideline for managers, not only to realize future interoperations with BlogForever, but also to use as a basis to build upon their own interoperability

methodology tailored to their own environments and needs. The description of the approach proposes several templates to assist the documentation of the steps. Furthermore, an example is given to facilitate the understanding of the application of the approach in the context of BlogForever.

Section 6 presents a tangible approach to develop a succession plan that would allow a reliable transfer of content from the current digital archive to other digital repositories in the case that a future BlogForever Archive is unable to continue for any reason. Concepts from the business model are employed for succession planning, and are adapted to the digital repository context, to suggest steps for achieving organisational interoperability. The succession plan development is also elaborated upon using the interoperability scenario development framework.

The results of this report inform on the one hand further development of the BlogForever platform, e.g. through revealing the relevance of the WARC standard for interoperability prospects. On the other hand, future deployment of the platform in real life scenarios is supported through the presented guidelines and approaches. Additionally, most of the findings are not just applicable for the BlogForever project but inform also the web archiving and digital preservation community in general.

1 Introduction

The following report aims at the evaluation of interoperability prospects of the BlogForever platform with 3rd parties. Modern computer networks have given wide access to a huge number of digital resources and systems. Naturally, one of the next steps in the process of leveraging all these resources is the information retrieval and management operations between one or more systems in order to assemble new information and/or provide new functionality. This task is not trivial given the wide range of computer systems, hardware, software, operating systems, applications, protocols and file formats. Therefore, interoperability is a major issue of growing importance, especially in the era of the World Wide Web.

While interoperability is a domain with a multitude of facets, the BlogForever project aims on answering the following questions:

- What are current and future interoperability challenges in the web archiving domain?
- What are the relevant technical interoperability standards and protocols for BlogForever?
- How can a specific institution proceed to establish and maintain interoperability of their BlogForever implementation?
- How can sustainability for collections based on the BlogForever system be established?

1.1 BlogForever: Background

The main aim of the BlogForever project is to develop robust digital preservation, management and dissemination facilities for blogs. The outcomes of the project are expected to benefit a number of stakeholders, in particular, libraries, information centres, museums, universities, research institutes, businesses, as well as blog authors and readers in general.

The investigation on interoperability prospects is part of Work Package Three (WP3) BlogForever Policies that has three main tasks[1]:

- *Task 3.1: Development of the Preservation Strategy.* This task focuses on weblog preservation and long-term accessibility. The outcome of this task was D3.1 Preservation Strategy Report.
- *Task 3.2: Assessment of Interoperability Prospects.* This task focuses on interoperability and compatibility prospects of the BlogForever platform. The outcome of this task is D3.2 Assessment of Interoperability Prospects Report.
- *Task 3.3: Development of the Digital Rights Management Policy.* This task focuses on the development of a Digital Rights Management Policy that will clearly define the access level and type of allowed use of all items stored in the BlogForever platform. The outcome of this task will be D3.3 Digital Rights Management Report.

The three tasks are loosely interconnected and will be running in parallel in order to support later steps of the project. The investigation of the interoperability prospects should inform the design and development of the BlogForever platform and the Preservation Strategies. The description of work describes furthermore as specific aims[2]:

- Investigation of interoperability and compatibility prospects regarding current efforts, which can either complement or success the BlogForever platform,

[1] For details see: Grant Agreement Annex I - Description of Work (DoW), page 10.

[2] For details see: Grant Agreement Annex I - Description of Work (DoW), page 10.

- Survey research and industrial complementary efforts on web archiving or digital preservation in general, and
- Outline the means for reliably transferring content from the current digital archive to other digital repositories should BlogForever platform discontinue the project for any reason.

Former deliverables of the BlogForever project informed the assessment of interoperability prospects. The Weblog Data Model[3] is a foundation of the BlogForever platform, and, therefore, provided important constraints for the interoperation of data. The considerations about Weblog Ontologies[4] delivered initial insights into interoperability prospects, mainly on the semantic level. The BlogForever Survey[5] provided first empirical information about the adoption of standards by blog platforms, and the Requirement Analysis[6] already specified five interoperability requirements for the BlogForever platform. Furthermore, this deliverable may influence further development of the BlogForever platform in work package 4.

1.2 Contribution of this report

The current report contributes to the current research landscape in the following directions:

- Interoperability is an area widely researched and discussed and has been conceptualized under different perspectives. Section 2 summarizes several conceptualizations that have been proposed about interoperability models and levels and provides an extensive overview of the present literature on interoperability.
- A Delphi study, conducted within the framework and the needs of the project, about aspects of interoperability of web archives and digital libraries is presented in Section 3. The survey reveals and shares remarkable insights regarding current problems, limitations, needs and challenges that are encountered in today's interoperations (or efforts to this direction) among systems of the web archiving and digital library communities. The survey is carried out among a small, purposively selected group of people with expertise on the topic, who shared their views and ideas, adding a valuable input to the research. This survey offers a unique contribution to the limited so far research field of interoperability, presenting the current barriers but as well suggestions for future approaches, and can be a useful study for the three involved communities: the web archiving, the digital library and the digital preservation community.
- An extensive overview of the standards that support, assist or establish interoperability is provided in section 4. Standardisation is probably the most essential aspect of interoperability since standards can be the bridge between two or more different environments. This section provides a useful guide about the most commonly used standards that address the needs for interoperation mainly in the domains of digital libraries and web archives since these are the most relevant systems for BlogForever to interoperate with.
- Section 5 presents a 5-step approach to consider and configure the enabling of interoperability of the BlogForever system (or any digital library) with another potential information system. This approach offers a useful and concrete guideline for managers, not only to realize future interoperations with BlogForever, but also to use as a basis to build upon their own interoperability methodology tailored to their own environments and needs. The description of the approach proposes several templates to assist the documentation of

[3] BlogForever Deliverable D2.2: Weblog Data Model

[4] BlogForever Deliverable D2.3: Weblog Ontologies

[5] BlogForever Deliverable D2.1: Survey Implementation Report

[6] BlogForever Deliverable D4.1: User Requirements and Platform Specifications Report

the steps. Furthermore, an example is given to facilitate the understanding of the application of the approach in the context of BlogForever.

- Section 6 presents a tangible approach to develop a succession plan that would allow a reliable transfer of content from the current digital archive to other digital repositories in the case that a future BlogForever Archive is unable to continue for any reason. Concepts from the business model are employed for succession planning, and are adapted to the digital repository context, to suggest steps for achieving organisational interoperability. The succession plan development is also elaborated upon using the interoperability scenario development framework.

1.3 Structure of the deliverable

The rest of this report is structured as follows. In Section 2, we review different models and perspectives of interoperability. Section 3 presents a Delphi study that has been conducted to examine current and future challenges for the interoperability of web archives and digital libraries. Section 4 surveys interoperability standards from a technical perspective. A simple approach to consider interoperability of a BlogForever archive in a specific usage context is presented in section 5. Section 6 comprises elaborations about the succession plan. Finally, a conclusion for this deliverable is drawn in section 7.

2 Models of Interoperability

In the following, an overview of different conceptualisations for interoperability is given. Thus, the understanding of the complexity should be improved and further considerations in this report facilitated.

According to IEEE, interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged (IEEE, 1990). Interoperability has numerous facets including uniform naming, metadata formats, document models, and access protocols (Lagoze & van de Sompel, 2001). Interoperability in a narrow sense describes how technical systems interoperate. Interoperability in a broader sense comprises also social, political, and organisational factors (Gottschalk, 2009). Compatibility is also a related term. A product is compatible with a standard but interoperable with other products that meet the same standard (or achieve interoperability through a broker).

The European Interoperability Framework for pan-European e-Government Services (Commission) defines three dimensions of interoperability (Anon., 2004):

- **Organisational interoperability** comprises necessary activities on an organisational level like defining business goals, modelling business processes and bringing about the collaboration of administrations that wish to exchange information and may have different internal structures and processes. Thus, requirements of the user community should be addressed by making services available, easily identifiable, accessible and user-oriented.
- **Semantic interoperability** is given if the precise meaning of exchanged information is understandable by the interoperating. Thus, the processing of received information in a meaningful manner, and the combination of received information with other information resources is enabled.
- **Technical interoperability** focuses on the technical issues of linking computer systems and services. Hence, key aspects like open interfaces, interconnection services, data integration and middleware, data presentation and exchange, accessibility and security services are considered.

DL.org adopted the dimensions to provide a framework for interoperability scenarios in digital libraries (see Figure 1). Thereby, interoperability is considered as the communication between a provider system that provides a specific resource and a consumer system that requests the resource to perform a specific task. The operation of the provider as well as the consumer depends on organisational, semantic, and technical aspects (Athanasopoulos et al., 2011). In former publications, the semantic level was also mentioned as content level (Arms et al., 2002).

Figure 1: Interoperability scenario (Athanasopoulos et al., 2011)

Further levels/approaches of interoperability specifically for digital libraries are (Arms et al., 2002):

1. **Federation**: A federation is a group of organisations that agree that their services conform to certain specifications or deploy common formal standards. This model requires some effort by each organization to implement and remains consistent to all agreements in order to provide some basic shared services. A typical example is the case of libraries that share

online catalogue records using Z39.50[7]. The cost of participation is relatively high and therefore, typical federations have few but dedicated members.

2. **Harvesting**: Harvesting uses the metadata about collections provided by digital libraries in a simple exchange format. Thus, additional services can be provided for information discovery and reference linking. The motivation of this approach lies in the difficulty to follow the former one, i.e. to create large federations. The harvesting approach is about forming looser groups, where participants agree to enable some basic services without adopting a complete set of agreement and without significant effort. Therefore, even if the provided services are less powerful than those of federations, organizations are more likely to join.
3. **Gathering**: this approach produces a base level of interoperability, which is possible for organizations that are not prepared or eager to cooperate with any of the former ways. The gathering approach is about gathering openly accessible information. Since this model does not incur a cost for the libraries, the services can embrace large numbers of digital libraries, although the quality might be poorer in comparison with cases that systems cooperate directly. For example, CiteSeer[8] is such a digital library built automatically by gathering publicly available information.

In the context of digital preservation, Digital Preservation Europe (DPE) is distinguishing six aspects of interoperability (Gradmann, 2007):

1. **Interoperating entities**, e.g. traditional cultural heritage institutions (libraries, museums, archives) offering digital services, digital repositories (institutional or not), E-Science and/or E-Learning platforms, or simply web services.
2. **Objects of interaction** are the entities that need to be processed, e.g. the full content of digital information objects (analogue/digitised or born digital) to mere representations of such objects.
3. **Functional perspective of interoperation**: Examples are the exchange and/or propagation of digital content, the aggregation of digital objects into a common content layer, enabling users and/or software application to interact with multiple Digital Libraries via unified interfaces (dynamic portals), the facilitation of operations across federated autonomous Digital Libraries, and the establishment of common service architecture and/or common service definitions.
4. **Linguistic interoperability (multilingualism)** can be either represented as multilingual user interfaces to Digital Libraries (relatively well known) or as dynamic multilingual techniques for exploring the Digital Library object space.
5. **Design and user perspectives** consider the different conceptions of distinct users or roles like the Digital Library manager, the content consuming end user, the technical administrator, the end user providing content as an author, the digital content aggregator, the 'meta user', or the policy maker.
6. **Technological standards** for different purposes, for example librarian metadata interoperability (Z39.50 / SRU+SRW), harvesting methods (based on OAI-PMH), web service based approaches (SOAP/UDDI), Java based API defined in JCR (JSR 170/283), or GRID based platforms such as iRods.

Furthermore, the DPE describes the following abstraction levels where the abstraction increases from technical/basic to semantic (Gradmann, 2007):

[7] See also section 4.5.3.

[8] http://citeseerx.ist.psu.edu

- Technical or basic level: common tools, interfaces and infrastructures providing uniformity for navigation and access.
- Syntactic level: Allowing the interchange of metadata and protocol elements.
- Functional or pragmatic level: Based on a common set of functional primitives or on a common set of service definitions.
- Semantic: Allowing to access similar classes of objects and services across multiple sites, with multilingualism of content as one specific aspect.

However, the order of a semantic level on top of a pragmatic level is inconsistent with the three parts of semiotics: (a) syntax (or structure), (b) semantic (or structure-based meaning), and (c) pragmatic (or context-based meaning) (Morris, 1938).

The Levels of Conceptual Interoperability Model (LCIM) has been developed for the interoperation of modelling and simulation applications. The LCIM distinguishes six levels of interoperability shown in (Tolk et al., 2007).

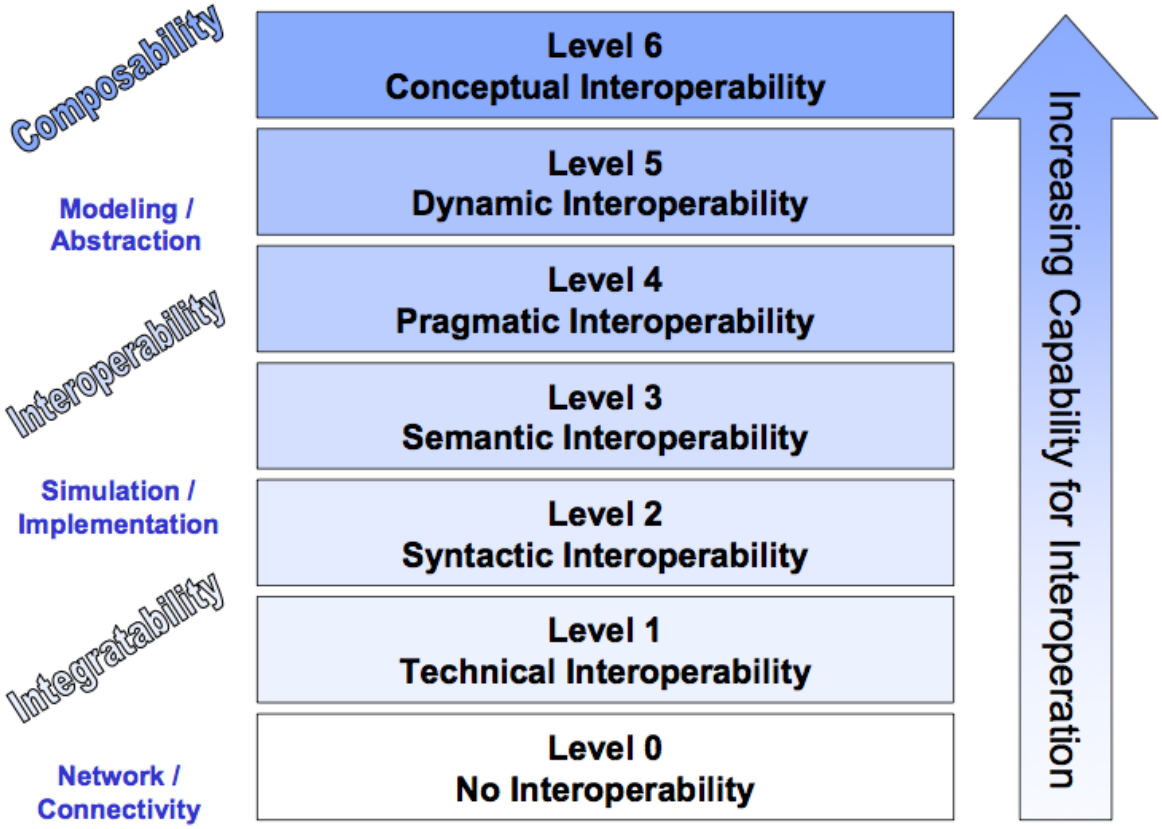

Figure 2: Levels of Conceptual Interoperability Model (Tolk et al., 2007)

The level of no interoperability consists of stand-alone systems. The technical level establishes a basic communication infrastructure through the definition of communication networks and protocols. On the syntactical level, common structures (e.g. a common data format) are defined to exchange information. The next level of semantic interoperability establishes the sharing of the meaning of the data. On the pragmatic level, the interoperating systems have to understand the context of each other data application through being aware of the methods and procedures that each other are employing. The fifth level of dynamic interoperability considers the changes in the state of the systems over time. The highest level is the conceptual interoperability. The interoperating systems align their conceptual model (e.g. assumptions and constraints of the meaningful abstraction of reality) on this level (Tolk et al., 2007; Tolk & Muguira, 2003).

Semantic Interoperability is important for handling a wide range of data context, mostly derived from heterogeneous information sources. The DELOS describe semantic interoperability on three levels (DELOS, 2005):

1. **Data Structures**, which are metadata, content data, collection management data, service description data.
2. **Categorical Data**, which are data that refer to concepts, e.g. classifications, typologies, and general subjects.
3. **Factual Data**, which are data that are related with specific records and are helpful to identify them, e.g. about people, items, places.

Pragmatic Interoperability is achieved when the interoperating systems are aware of the **methods** and **procedures** that each system is employing. Pragmatic Interoperability requires Syntactic and Semantic Interoperability. Pragmatic interoperability is related with the compatibility between the intended versus the actual effect of signs exchanged among systems (Camlon H. Asuncion, 2010). Although, pragmatic interoperability seems as a very important topic, a blurry picture covers its actual essence. Also, Pragmatic Interoperability is achieved if collaborating systems share not only the same understanding of the intended use of sings, but also the same context in which the signs exchange is applied. The last is referred as business intention exchange (Camlon H. Asuncion, 2010). The signs in that case may be business rules, organizational policies, organizational agreements and others.

The maturity of interoperability as the capability between two organisations to interoperate can be described on five levels (see Figure 3). On the level of computer interoperability, organisations are able to exchange information directly between their IT systems. Technical and semantic issues have to be solved in order to enable the IT systems to send, receive, and process message. The level of process interoperability can be reached if the processes in the interoperating organisations are linked to each other, for example a sub-process is fulfilled by the corresponding organisation. The next level of knowledge sharing demands knowledge exchange between the employees in inter-organisational relationships. Value interoperability requires that the interoperating organisations have the same configuration how they create value. The configurations can be distinguished in value chains, values shops, and value networks. The highest level of goal interoperability leads to a strategic alignment of the organisations. Thus, it is ensured that no conflicting goals occur between the organisations (Gottschalk, 2009).

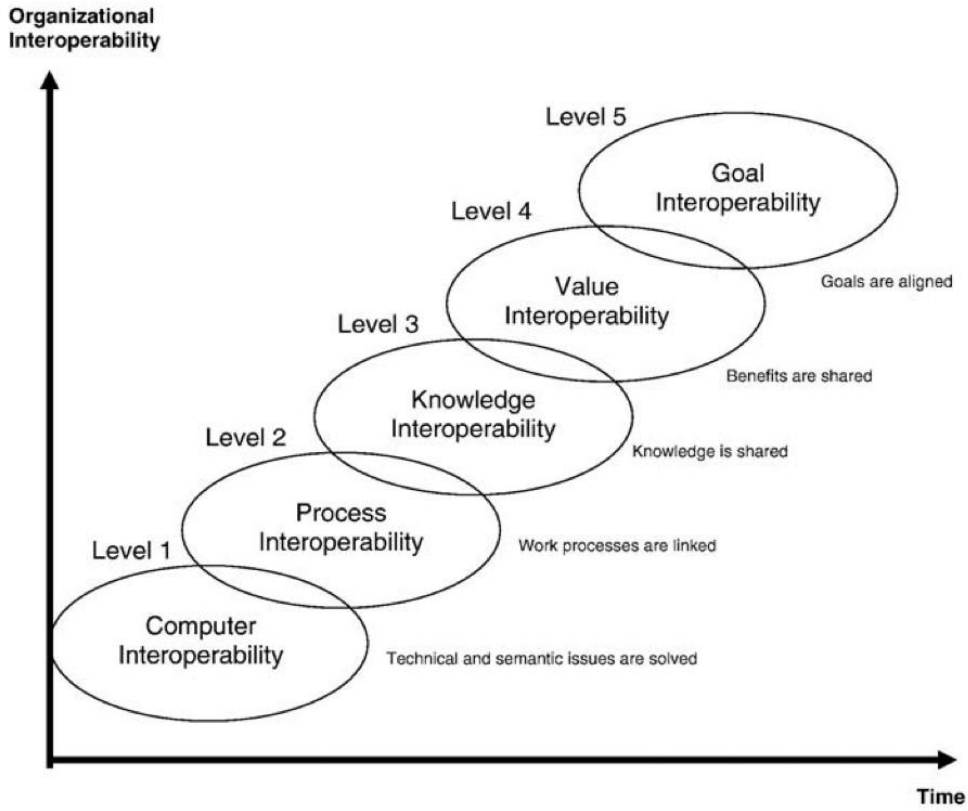

Figure 3: Maturity levels for interoperability in digital government (Gottschalk, 2009)

This section introduced different models that can be used to analyse and describe interoperability. While several levels of interoperability have been presented this way, additional levels like physical, empirical, social, political, legal, or international interoperability can also be found in the literature (Manso et al., 2009). However, the model of interoperability should be chosen according to the purpose of the interoperability considerations. In general, it can be stated that interoperability is much more than the interoperation of computer or IT systems but their interoperation is often the foundation for interoperation on higher levels.

The bottom line for the BlogForever project and this report should be that, on the one hand, interoperability has to be considered on an IT level because the main output of the project is a software platform but, on the other hand, the challenges on higher levels of interoperability (e.g. organisational aspects) should be reflected as well in order to facilitate further deployment of the BlogForever platform.

3 Current and future challenges of interoperability: a Survey

In this section, a Delphi study about aspects of interoperability of web archives and digital libraries is presented. The study has been conducted in order to provide a better understanding about the purposes, challenges, and possible solutions of interoperability for the BlogForever project as well as for the community of web archives and digital libraries in general. Just few surveys have inquired the subject yet (see section 3.1). Thus, there is a high risk that research and development to improve interoperability is based only on personal assumptions, beliefs, or experiences, but does not address the real needs of the community. Due to the fact that the subject is not well structured yet, a method for qualitative and explorative research has been chosen. The Delphi method (see section 3.2) provides a systematic way to survey the experience of a group of few experts in the field. The first round is presented in section 3.3, then section 3.4 presents the results of the second and final round, before we discuss and draw some conclusions and implications for the BlogForever project in sections 3.6 and 3.6 respectively.

3.1 Related Work

There have been several surveys in the domains of web archiving and digital libraries that address important issues. Most of these surveys examine several characteristics of the web archives like content and used technologies or other aspects like national legislations and policies about access and copyright.

In particular, the International Internet Preservation Consortium (IIPC)[9] conducted a survey on its members, which was more like a profile identification, examining the maturity of web archiving, the scope, the tools used for harvesting, curation and access, legal limitations by their countries and access restrictions etc. (Grotke, 2008). Similarly, the Internet Memory Foundation conducted a survey on European institutions aiming to characterise them and addressed several other aspects like the status of web archiving, legal aspects, access restrictions, policies and priorities regarding the scope and the types of archiving (Internet Memory Foundation, 2010). The 18th Conference of Directors of National Libraries in Asia and Oceania (CDNLAO) presented as well a report with the participants' answers about web archiving in this region. The questions were about cooperation, access and preservation policies, used tools and the legal framework (National Diet Library, 2010). Later, Gomes et al. (2011) presented an updated overview of the web archiving initiatives internationally, in which the addressed aspects were mainly the scope, content characteristics, file formats, technologies and the provided access (Gomes et al., 2011).

Regarding interoperability in particular, an important attempt in web archiving was made by Jacobsen (2007) who contacted national libraries examining several issues like scope of harvested resources, collecting and discovering policies, level of harvesting, access to archived content, level of cooperation with other web archives, how they solve ownership and technical issues and what kind of institutions would they engage in partnership with to solve such problems (Jacobsen, 2007). Also, in the sphere of digital libraries this time, the DL.org Policy Working Group made an experimental survey on policy interoperability of digital libraries to a selected sample of digital libraries, digital repositories and federated services. This survey addressed how their policies, strategies, frameworks and plans affect or are affected by interoperability (Innocenti et al., 2011).

BlogForever has also conducted surveys but specifically considering the Blogosphere. In particular, the first survey was carried out among bloggers in Deliverable D2.1[10] to enhance the understanding of blogs and the particular importance of specific entities for bloggers. Additionally, the technological foundations of the blogosphere have been examined (Banos et al., 2012). Later,

[9] http://netpreserve.org/

[10] BlogForever Deliverable D2.1: Survey Implementation Report

within the preparation of D4.1[11], the BlogForever team interviewed representatives of 8 stakeholders groups, including digital libraries and researchers. Both surveys provided outcome that was used to identify and implement potential requirements for the platform and design certain policies and approaches. However these surveys, even if they partially addressed web archiving activities, did not provide any particular input considering Interoperability.

Our survey aimed to gain insight into areas that have not been surveyed and derive from people who are highly involved and have some personal experience. Our aim was to examine a theoretical framework of interoperability in web archives and digital libraries with the help of people who have their own experiences on the topic. This survey can be more considered as a discussion about interoperability; the obstacles, the current limitations, the followed approaches, the forthcoming challenges, the ideas for improvement. Therefore, our contribution, not only to the research community but as well as to the involved communities, is the sharing of the valuable outcome of an enlightening virtual discussion from experts about interoperability.

3.2 Method

In this section, we outline the underlying method of our research. We aim on the identification of current and future main issues for the interoperability of web archives and digital libraries. We decided for an explorative, qualitative research in order to have the chance to identify also novel issues in this field. Our intention was not to extract statistical results from either the entirety of the web archives and digital libraries or from a representative sample of it, but to gain useful insights from a group of people that are highly involved and particularly interested in this topic and the future progress. Hence, we chose the Delphi method to survey a purposive sample of experts.

Delphi Method was developed from the need for a technique able to obtain the most reliable consensus of a group of experts (Okoli & Pawlowski, 2004). While it was initially conceived as a group decision technique aiming to obtain a consensus, now it is also used as a research method to obtain reliable opinions and valuable contributions from a group of experts in order to resolve a complex problem (Landeta, 2006). For example, several Delphi studies are ranking-type and aim to extract a consensus opinion on the importance of specific issues, but others emphasize differences of opinion in order to develop a set of alternative future scenarios (Okoli & Pawlowski, 2004).

A Delphi method undergoes two or more rounds. The first round is an exploration of the subject. The researchers design the initial questionnaire and select an appropriate group of experts who are qualified to answer the questions. In this round each individual panellist contributes additional information that he feels is important to the topic (Linstone & Turoff, 2002). The responses are then collected and analysed. Based on the analysed results, a second questionnaire (second round) is designed in which respondents are asked to revise their original responses and/or answer other questions based on group feedback from the first round (second round). The Delphi method is an iterative process and each subsequent questionnaire is developed based on the results of the previous questionnaire. The number of required rounds depends strongly on the purpose of the research. In general two or three iterations are suggested for most research but fewer could be also adequate to reveal sufficient information (Skumolski et al., 2007). However, the participants are usually given at least one opportunity to revise their original answers upon examination of the group response (Linstone & Turoff, 2002).

In general, it is highly important to ensure that in every round participants have the opportunity to refine, clarify or change their views in light of the progress of the group's work (Skumolski et al., 2007). Furthermore, anonymity is another important factor of the survey and should be also reserved in each round.

Aim of the Delphi study in this report is to inquire into interoperability of web archives and digital libraries. Therefore, a purposive sample of seven international experts from the web archiving and

[11] BlogForever Deliverable D4.1: User Requirements and Platform Specifications Report

digital library communities was created. While the research team knows the identity of the participants, the participants were anonymous among each other. Thus, a possible bias by reputation or hierarchy perceptions or an answering according to expected norms could be avoided.

The study consisted of two rounds. A purposive sample of seven international experts from the web archiving and digital library communities was created. While the research team knows the identity of the participants, the participants were anonymous to each other. Thus, a possible bias by reputation or hierarchy perceptions or an answering according to expected norms could be avoided.

Aim of the first round was a brainstorming about the purposes, obstacles, possible solutions to overcome limitations, and other future challenges. Therefore, a questionnaire has been created with four open questions (see A' Round Questions). The questions were created by two researchers and then reviewed by an archivist as domain expert. Based on the recommendations of the review, questions were adapted to improve the wording according to the participants' context. The final questionnaire was sent as word document and as online questionnaire to the participants at the beginning of February 2013. The participants had three weeks time to answer. Additionally, a reminder was sent in the middle of the three weeks to participants that had not responded yet. The final answers of the first round were analysed qualitatively by two researchers in parallel. Afterwards, results were compared and discrepancies in the interpretation were solved through discussion. The final results of the first round were documented (see section 3.3) and used to create the second round.

Aim of the second round was to verify identified results from the first round by all participants as well as to create further insights through evaluation regarding different aspects. Therefore, an online questionnaire with closed questions was created. The questions were created by two researchers according to the structure of the first round's results, reviewed afterwards by the archivist and, then, further improvements of the wording were made based on the review. Additionally, the questionnaire was tested with two individuals related to the archiving sector in order to test the understanding of the questionnaire as well as to confirm the time estimation for answering the questionnaire. Finally, the questionnaire (see B' Round: Questions & Result Tables) was sent to the participants at the beginning of April 2013.

3.3 A' Round: Analysis & Results

The first round comprised four open questions in which participants were free to develop their views as extensively as they might wish. The result was a brainstorming that provided several perspectives and multiple facets of the discussed topics. We analysed the responses qualitatively and organised the results in four main categories:

1. Purposes
2. Barriers
3. Solutions
4. Further & future challenges

Beyond our initial aspects for research, we identified some additional interesting topics for discussion. Particularly:

1. Interoperability perspectives
2. Additional Benefits
3. Usage Scenarios
4. Related projects that were highlighted

In the following, the analysis of the responses is explained in detail and related quotes from the participants are put as well to explain and support the statements.

3.3.1 Interoperability purposes

We collect under the term purposes the motivations and abstract use cases that require interoperability. The identified purposes can therefore be understood as answers to the question why a web archive or a digital library would consider interoperability with other systems. In particular, the identified purposes can be overlapping or complementary and should not be understood as disjoint classes.

Three dimensions to describe interoperability in terms of purposes and motivations have been identified. The first dimension describes the distinct uses for which interoperability is necessary. Thereby, uses that motivate interoperability can be differentiated between:

- Federated search,
- Federated access,
- Exchange, and
- Replication.

Federated search in the context of our research is the possibility to search from a single point or with a single query for data that are stored in several web archives or digital libraries. An example for federated search indicated by one of the participants was the following:

"For example, a collaborative of three of four cultural heritage institutions might digitize texts related to WWII and place them into a single collection. Each institution might house a copy of their own materials but create an aggregate index of all texts in the combined collection so that researchers may discover them and seek to access them from partner institutions as is feasible."

While federated search just requires that the desired data can be searched and found in different archives, **federated access** also enables the user to retrieve the data directly from a single point. This means that, for example, the data can be viewed or downloaded. We distinguish between federated search and federated access in order to emphasize the opportunity for the user to directly access through one interface the objects that are stored and managed in distributed locations. Therefore, a precondition of federated access is that the object has a digital form while federated search is also possible for non-digital, e.g. printed, objects. An example that indicated the desire for federate access was:

"One is to make it easier for people to access and use content despite the physical location of the content. For example a researcher can discover and bring together into one view content from many different repositories."

Exchange and replication are similar and describe different aims for the transfer of data between archives. **Exchange** of archived objects may be necessary to create or to complement specific collections like the collection of information about a specific topic or event. This is revealed in the aforementioned answer regarding a collection about WWWII and, as well, in the following response:

"collaborative constitution of web archives collections for the 2012 Olympic games in London".

Replication on the other hand aims on data redundancy in order to reduce the risk of data loss and improve reliability. While the specific reasons for replication were not further explained by the participants, the need for replication was mentioned in statements like the following:

"two preservation repositories could exchange content with each other so that they each have extra copies in geographically-distant data centers."

"The purpose of interoperability in the context of digital preservation is two-fold: exchange of information and distribution of replicas."

The second dimension derives from the differentiation in the scopes of the above uses (*"to transfer content between systems of any kind that store digital content"* and *"to make it easier for people to access and use content despite the physical location of the content".* Hence, it can also be

understood as a specialisation of the already described purposes. Particularly, the following refinements were made about interoperation across:

- National boundaries,
- Organisational boundaries, either among organisations of the same type (e.g. among several digital libraries), or among organisations of different type (e.g. between a national digital library and the national web archive).

The last dimension that we identified differentiates the motivations based on the objects in focus. Thus, interoperability may concern either primary objects entirely or only metadata. One participant gave us the following example:

"It may be exchange of collection if data are interoperable, or only collaborative referencing of collections if only metadata are interoperable".

3.3.2 Barriers to interoperability

The analysis of all the answers that mentioned difficulties, limitations or obstacles to interoperability led to the identification of the following more general categories:

- Standardization,
- Tools and implementation,
- Organisational obstacles,
- Legal problems, and
- Approach to handle interoperability.

While various standards already exist, the current state of **standardisation** and compliance seems to be unsatisfying. A lack of agreed standards has been reported. A lack of global unique identifier (URI) can be understood as a specific sub-problem. Similar to the lack of agreement, competition among the already existing standards has been reported.

However, even the agreements on standards do not lead to interoperability because problems occur when they are applied or implemented. One problem is the lack of **tools** that implement the existing standards. Next to this, the same standard can be applied or used differently in different contexts what in turn can hinder interoperability. More specifically, even if two archives apply the same schema (e.g. METS), the content can be modelled differently and thus impede interoperability:

"the differences in the metadata granularity in archives and libraries", "Technically we model content differently. Even when we use the same schemas (e.g. METS) we use them in different ways.".

While the barriers regarding standards are mainly of technical nature, barriers occur also from an organisational and legal perspective. **Organisational obstacles** concern the ability and willingness of an organisation to provide interoperability for its collections. Some organisations are not willing to commit in collaborations and partnerships or they are not willing to invest in standardizing processes:

"Too often organizations fear the process of becoming "dependent on another organization" when it is hard enough to operate alone")

Furthermore, organizations may feel unable to provide or invest into interoperability because of the expected effort as well as the lack of know-how and resources in the organisation:

"Large-scale collaborations can be time-consuming and require a lot of effort and communication, especially for mission-critical activities like preservation.".

Last, some organisations have actually no desire to provide any interoperability:

"In many cases, there is no desire for interoperability. Quite to the contrary, there are clear strategies aimed at not being interoperable in an attempt to lock in a user base, i.e. prevent users from seamlessly moving between information environments").

Legal barriers can hinder interoperability. Participants reported national regulations that limit or prevent any data exchange:

"exchange of data via ingest or export from other institutions outside of a "national" umbrella is strictly limited or forbidden. This is true today for many EU countries like Denmark, Sweden and Norway".

This particular point has also been raised in previous survey (Jacobsen, 2007) and was later addressed by the same author in detail (Jacobsen, 2008). Apart from this, the copyright holders define significantly the level of access and intellectual property laws hinder an open or public the access:

"We rely on the personal permit of copyright holders. National libraries can't or do not offer free access to the collections.").

Last, the **approach to establish or handle interoperability** seems to differ. For example, different perspectives of traditional librarians and web archivists were reported as a barrier to collaboration and interoperation between the two communities:

"there is sometimes a reluctance by the traditional library people to embrace web technology: harvesting and free text search versus a well controlled and high quality library catalog".

Furthermore, communities often define interoperability based on the specific systems they wish to interoperate and then define an approach to establish it, which is tailored to these systems:

"Often times, communities that are keen to achieve interoperability come at it from a perspective of determining which "systems" need to be interoperable" […] *This kind of system-to-system interoperability can effectively achieve desired interoperability levels among the targeted systems but leaves all other information environments unaffected and unable to benefit from the interoperability investment. ".*

3.3.3 Suggested solutions & improvements

Several suggestions to overcome current barriers and achieve better levels of interoperability have been proposed by the participants as possible solutions or improvements.

Clear Legislation and policies regarding the exchange of data/metadata: An essential change would be clarity in national legislations regarding the exchange of data/metadata because it seems to be a grey area in many countries that makes the institutions more reluctant to exchange information.

"Today many believe a precedence has been set for this through the efforts of the Linked Open Data community (LOD) in Libraries, Archives, and Museums around the globe but in fact it is still a gray area in many countries making national institutions hesitant to exchange information regarding their holdings. With clarity on this front, the global archival community could work more closely and in partnership on capturing and preserving representative samples of the Web.

Standardization: Regarding standards there seem to be a diversity of opinions. On the one hand, there is the belief that new, better, global and well-defined standards are needed, to handle interoperability limitations. For example, it should be very clear to institutions what is the minimum metadata information to be included in a single item:

"Defining a set of global standards and protocols for the exchange of this data will need to be ironed out including what minimal information must be contained in the core information package.".

On the other hand, there is the belief that there is not really need for new standards, but there should be a consensus on which standards to use and then conformity with them. Furthermore, an initiative that would somehow necessitate the use of specific current standards could be beneficial.

Implementation & other developments: even though the current standards seemed to be sufficient, the need for tools to implement them was also suggested:

"development of tools implementing current standards".

Further technical changes that are said to be supporting are the use of common APIs for search and retrieval and a central aggregation service that could bring all the information from several collections to the user. For example:

"we need to have common APIs for searching and retrieving content and metadata")

People's and communities' involvement: Communities and individual people are said to play also a part in this direction. The different communities should collaborate and be more involved in each other's activities so that their particular needs are also taken into account. For example, the web community could be more involved in the digital preservation community to ensure that web archiving needs are considered in the development of digital preservation standards:

"…it is necessary to be involved in the wider digital preservation community in order to ensure that web archiving needs are taken into account by main digital preservation standards (eg METS or PREMIS)").

Involved people are also said to be influential because sometimes their community may significantly influence their perspectives. As mentioned previously, web and library world seem to have different and even controversial priorities sometimes and therefore, people with broader knowledge should be involved in the interoperability efforts*:*

"Different cultures: web people versus librarians. There are few people who belong to both worlds.[…]the most pressing need is the right kind of people. People who talk both languages.").

Knowledge sharing is also another suggested important path. Sharing the experiences of various interoperability efforts, i.e. the successful stories, the failures and the practises that have been found to be best, would contribute to improve methods, avoid mistakes, and use resources more effectively. A consensus on the best practises and the sharing of them would contribute in more and more institutions joining and collaborating. This is not insignificant, since several institutions, especially libraries, don't have enough financial or personal resources to invest individually on such efforts. Therefore, an initiative or funded organization to provide support about technical and legal issues would be also beneficial*:*

"As a institution financed by the university, public fundings and by projects we can't afford the costs for the technical support we need for the preservation. This means, we need an institution that helps with technical support […] An EU-based organization that offers help for legal and technical questions").

Sharing knowledge should also include providing clear definitions and terminology about the digital preservation aspects.

Last, another recommendation suggests a different perspective, to consider **interoperability from the perspective of the web infrastructure** and implement it in terms of web and independently, creating information interoperability and diverge from system-based interoperability.

"…tackle interoperability not from a repository, digital library perspective but rather from the perspective of the web infrastructure. Assets in archives and digital libraries are web resources with URIs. If interoperability for such assets is required, define and implement it in terms of the web.".

3.3.4 Further challenges

Part of our research was to examine interoperability with a view to the future. Therefore the participants were asked about future challenges they consider. We include in this category either the forthcoming changes that will put additional difficulties to interoperations or the challenging goals that have to be considered in further steps. With respect to this, four challenges have been identified in relevance to the future. It should be noted in advance that not all of them are directly related to interoperability, but primarily related to web archiving issues. They are stated, nonetheless, on the one hand because the interoperability of web archives is significantly dependent on web archiving strategies, and, on the other hand, to support further web archiving discussions and developments.

- **Interoperability of the content.** While current efforts aim on the interoperability of the systems to enable search, access, and transfer of resources, future attempts will focus also on the interaction of content. The vision could be a seamless web of archived content.

  "The most immediate challenge I see is the need/desire to start looking at web archives and digital libraries not only as a collection of resources with URIs but also as big datasets. This means that, not only will it be important to be able to have interoperability expressed in terms of URIs, metadata but also in terms of content. Think mining web archives as done by the BL, mining book collections such as Hathi trust. There will be a need to determine what the cross-information-environment "primitives" are to allow such mining, i.e. a core set of access mechanisms to content (not resources) across archives."

- **New players** with different systems, needs, and tools are emerging in the field of web archiving.

  "However, new actors are emerging, eg research labs or private companies that may use specific tools and/or are not experienced with the necessity of respecting standards. [...] So there is a strong need: - to promote standards towards new actors in web archiving"

- The increasing efforts to archive as much of the web as possible combined with the immense growth of the web will lead to an **explosion of the amount of web data** to archive.

  "Furthermore, the volume of data has exploded to 500TBs to PBs of data per crawl of the Web."

- New and **complex media and web resources** (Web 2.0, Social Media, etc.) demand enhanced methods for web preservation.

  "The problem of preserving social networks. For example, Facebook is, for the moment, a very important communication tool in the literary field, but because of the legal obstacles it is impossible to archive Facebook-pages (it would be only possible, if it would be possible to cut all comments and posts from other authors than the rightholder)."

3.3.5 Additional insights

Since the initial questions were open, participants were totally free to develop their thoughts, as extensively they would like to and with absolute freedom. That led to identify more aspects than the initial core ones, which were as well worth to mention. They are described in this section.

3.3.5.1 Interoperability perspectives

The responses of the first round revealed another dimension of interoperability based on the perspective that it is considered. To this direction, two different perspectives can be distinguished:

- **System Interoperability** (used following as shorter term for system-to-system interoperability) is probably the most traditional and common perspective which communities tend to follow. It is the perspective of defining interoperability based on which

systems are desired to interoperate. This perspective might be quite successful but it is limited to the particular targeted systems:

"This kind of system-to-system interoperability can effectively achieve desired interoperability levels among the targeted systems but leaves all other information environments unaffected and unable to benefit from the interoperability investment."

- **Information Interoperability** is about putting the focus on the information itself and making the information interoperable with different systems. It is the perspective of considering interoperability not from the perspective of a digital library, repository or any other information environment but rather from the perspective of web infrastructure instead.

  "An approach that yields better return on investment is based on achieving the desired level of interoperability by specifying and implementing it in terms of the existing infrastructure (the Web and its fundamental building blocks): define the interoperability problem in terms of the web and its primitives and solve it using those primitives, web standards, widely embraced technologies. [...] Assets in archives and digital libraries are web resources with URIs. If interoperability for such assets is required, define and implement it in terms of the web."

3.3.5.2 Benefits through interoperability

Among the participants' views regarding interoperability, we identified also some benefits that arise from the institutions interoperation and the general attempts in this direction. We consider as benefits any advantage or opportunity for the institutions and the involved communities that occurs through the interoperation of the systems or through the research and other efforts towards this. We distinguish the benefits from purposes since the later are goals that we aim to achieve or problems that we try to overcome, while the benefits are the additional positive effects that arise through the process or the outcome. With respect to this, the following benefits were identified:

- Dissemination of the content of an institution's collections internationally. As stated from a representative of a digital archive which collaborates with a universal web archive organisation:

  "We are collaborating with X… thanks to the presentation of our project on the website of X we can (get) not only a larger, but international attention."

- Institutions and organizations are benefited in areas in which they are constrained to act individually in terms of budget and annual resources or because of lack of know-how

  "Creating interoperability requires more preparation and ongoing management but if executed well will result in benefits to an organization that could not be realized alone, especially in the domain of access or preservation, areas in which individual institutions are by nature constrained in terms of budget and resourcing on an annual basis"

  This point has been revealed as well in a previous survey (Jacobsen, 2007) were respondents indicated desire to engage in partnerships that could offer some technical assistance.

- Development of common tools to collect, exploit and preserve content

  "Example : all IIPC members use the ARC or WARC standard so IIPC funds projects to develop or enhance ARC or WARC files harvesting, managing or accessing tools."

- Longevity of collections since their content is described and encoded in common standards.

3.3.5.3 Example usage scenario

In the first round, participants were asked to describe the general purposes of interoperability preferably using detailed usage scenarios. Based on these we identified some of the purposes given above. However, we would like to also give the actual scenario descriptions as more detailed

examples where interoperability is needed. Following, we list the identified usage scenarios of interoperability between systems:

- **Interoperability between an institution's web and its traditional collections:**
  A library, which holds a traditional library system and a web archive, holds collections of reports of various institutions. One of them decides to stop printing and only publish them on the web instead. Therefore, the library user while searching for them in the usual catalogue will find those up to a specific time point. But it is very likely that additional reports will be also available in the web archive and thus the user should also get a list from those in the web archive.
- **Exchange of content between preservation repositories - Replication**:
  Two preservation repositories exchange digital content with each other so that they both ensure the existence of extra copies in geographically-distant data centres.
- **Collaborative constitution of collections:**
  Two cultural heritage institutions decide to create a collaborative constitution of their web archive collections for the 2012 Olympic games
- **Federated search & access of scientific resources:**
  A researcher can discover content from different repositories and bring together into one single view.
- **Replication & preservation of scholar's work:**
  A scholar submits his work to an institutional or disciplinary repository. The digital object is (automatically) copied to a preservation repository as well.

3.3.5.4 Highlighted related projects

Some of the participants referred to interesting related projects to emphasize remarkable current initiatives that could work as examples for future directions.

- ***Memento***[12] is a project for providing archived versions of web resources putting focus on time. It was mentioned from two participants as an example of important interoperability efforts through a Web-centric perspective.

  "The Memento project is a great example of a project focused on creating interoperable access services for archived web resources. They chose to emphasize date and time of capture and the protocol for requesting resources as they key layers of interoperability. They do not focus on the preservation layers or other considerations given that the primary objective is knowledge of the existence of a resource, then the number and location of captures based on specific date/time criteria."

  "Memento is a good example: it tackles a long standing cross-webarchive interoperability problem by introducing a variation on content negotiation, which is one of the primitive concepts of the web infrastructure. In doing so, it not only tackled the cross-webarchive problem but it also allowed the technique to be used for information collections other than web archives, e.g. content management systems that support versioning, version control systems, etc."
- **Hathi Trust**[13] is a digital library created by the partnership and collaboration of more than 60 institutions aiming to ensure that the cultural record is preserved and accessible long into the future. Hathi Trust was mentioned as an example of content interoperability (*"not only will it be important to be able to have interoperability expressed in terms of URIs, metadata*

[12] http://mementoweb.org/

[13] http://www.hathitrust.org/

but also in terms of content. Think mining web archives as done by the BL, mining book collections such as Hathi trust.").

- ***DuraCloud***[14] is an open source service that provides the possibility of storing copies of digital objects in several different cloud providers and can be a part of digital archiving and preservation to organisations. Therefore, it is a service for replication, which was one of the identified purposes of interoperability.
- **OAI-PMH** and **OAI-ORE** were mentioned as examples of the approaches of system-to-system and information interoperability correspondingly. Extensive reference of them is given in the technical standards section.

3.4 B' Round: Results

In the second round, each panellist received the group response, structured as closed type questions, and was asked to evaluate it. Therefore, participants had the chance, on the one hand, to revise or confirm their own original answers, and on the other hand to read and consider the other panellists' views. They were also given the option to add comments and, therefore, the chance to object, clarify, complete the existing statements or add a new one.

In the second round we decided to keep the focus mainly on the four core aspects: Purposes, Barriers, Suggested solutions and Further challenges related to interoperability in web archives and digital libraries. These sections were only based on the actual answers. Furthermore, motivated by some responses, we created an additional part regarding the perspectives of considering and realizing Interoperability, which is partly based on actual responses and partly extended with additional questions. Therefore the second round comprised of 5 sections.

As (Linstone, Turoff, & Helmer, 2002) mentioned, a Delphi study deals mostly with statements, arguments, comments, and discussion and, thus, in order to evaluate the ideas that were expressed by the group, rating scales must be established as the relative importance or feasibility, for example, of various policies and issues that came up from the group. Similarly we adopted Likert-type rating scales. In most of the questions an option for no judgment was also provided (specifically "I can't say / I don't know" or shortly mentioned here as N/A).

The questionnaire of the second round was sent to our 7 participants as a personalised online web form in the beginning of April 2013 and was completed at the beginning of May. The second round was completed by the 6 of 7 initial panellists.

The following presentation of the results below and the correspondent discussion are following the same structure as the first round, based on the core aspects, i.e. the purposes, the barriers, the suggestions and future challenges, plus an additional one that came up after the first round analysis. However, this section contains mainly graphical illustrations of the results in order to facilitate a faster reading and understanding. The detailed results, as well as the given questions, can be found in tables in B' Round: Questions & Result TablesB' Round: Questions & Result Tables

3.4.1 Purposes

The responses of participants on the identified purposes for interoperability, in web archives and digital libraries, are summarized in Table 14 and illustrated in Chart 1. The following remarks can be made:

- Each of the identified purposes was agreed by at least five of the six participants. Two times, a participant answered with N/A, and one participant disagreed on replication as a purpose. In summary, we assess therefore the purposes as verified.

[14] http://www.duracloud.org/

- The use cases of federated access and federated search received the strongest agreement and can therefore be considered as commonly more accepted uses cases that require interoperability.
- The agreement for an interoperability focus on metadata is stronger than for a focus on primary objects.
- The agreement that interoperability is used to overcome organisational boundaries is slightly stronger than for national ones.
- With a more rigour assessment of verification, replication could be considered a questionable point since even for the majority of participants is a considered as a motivation for interoperability; one participant disagreed on this but unfortunately didn't provide some additional comments on this. Therefore, this is a point that should be further examined.

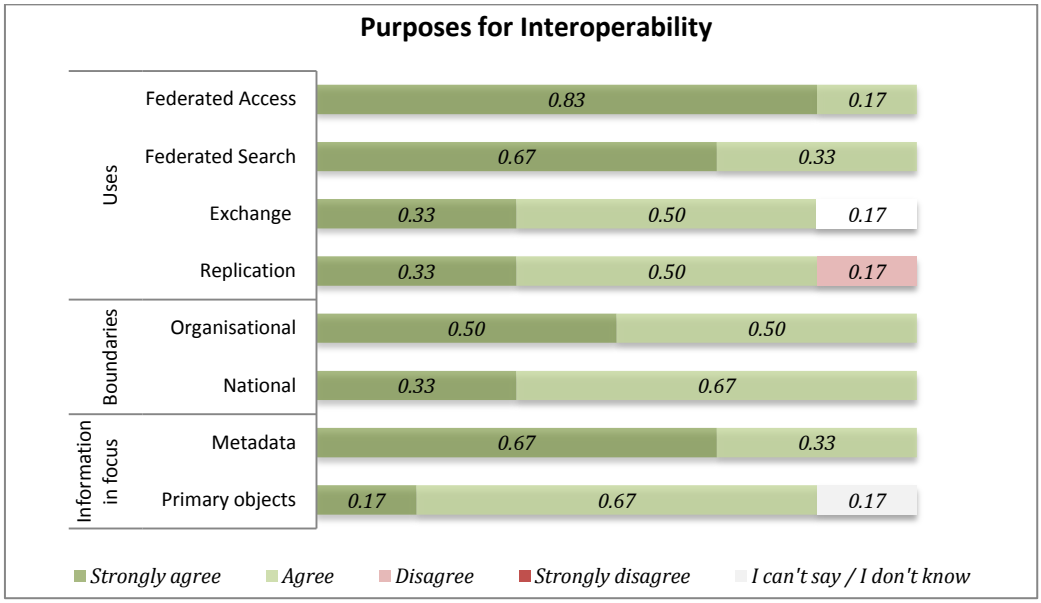

Chart 1: Response summary regarding the puproses for interoperability

It is worth mentioning that in this part we had two additional comments, both related to legal constraints. One of the participants wanted to clarify that most of the identified statements are prohibited in his/her country, and the answers were based on his/her personal consideration about how the situation should be. The other participant mentioned:

"Even though technical interoperability would allow some uses (eg federated access to web archives), legal constraints may prevent them."

However, by the end of the second round it drew our attention that the term 'purpose' could be ambiguous. While for most of the participants it was conceived as use cases we want to achieve with interoperability, for one participant purpose was more the upper reason with which such use scenarios can be achieved:

"Those are just examples of things one may want to achieve in an interoperable manner. That's not what I understand by the term Purpose. I see nothing re information interoperability, ie the ability to interpret information in a uniform manner across systems."

3.4.2 Barriers

The answers regarding barriers in the first round revealed several issues and, therefore, resulted in a long list of identified points that hinder or prevent the establishment of interoperability among systems. Participants were asked to evaluate them from the point of view of an organisation

individually and of the community as a whole separately. We adopted this distinction based on the assumption that there could be variations for some of them.

In this part one of the participants clarified that his/her answers are exclusively for web archives and one other did not evaluate the entire part regarding the community point of view (using N/A answer).

The detailed statements and responses are included in Table 15 (Appendix) while Chart 2 and Chart 3 following, present an illustration of the responses regarding an organisation individually and the community respectively. The barriers are ranked from most extreme barriers at the top to insignificant barriers at the bottom (based on the average assessment).

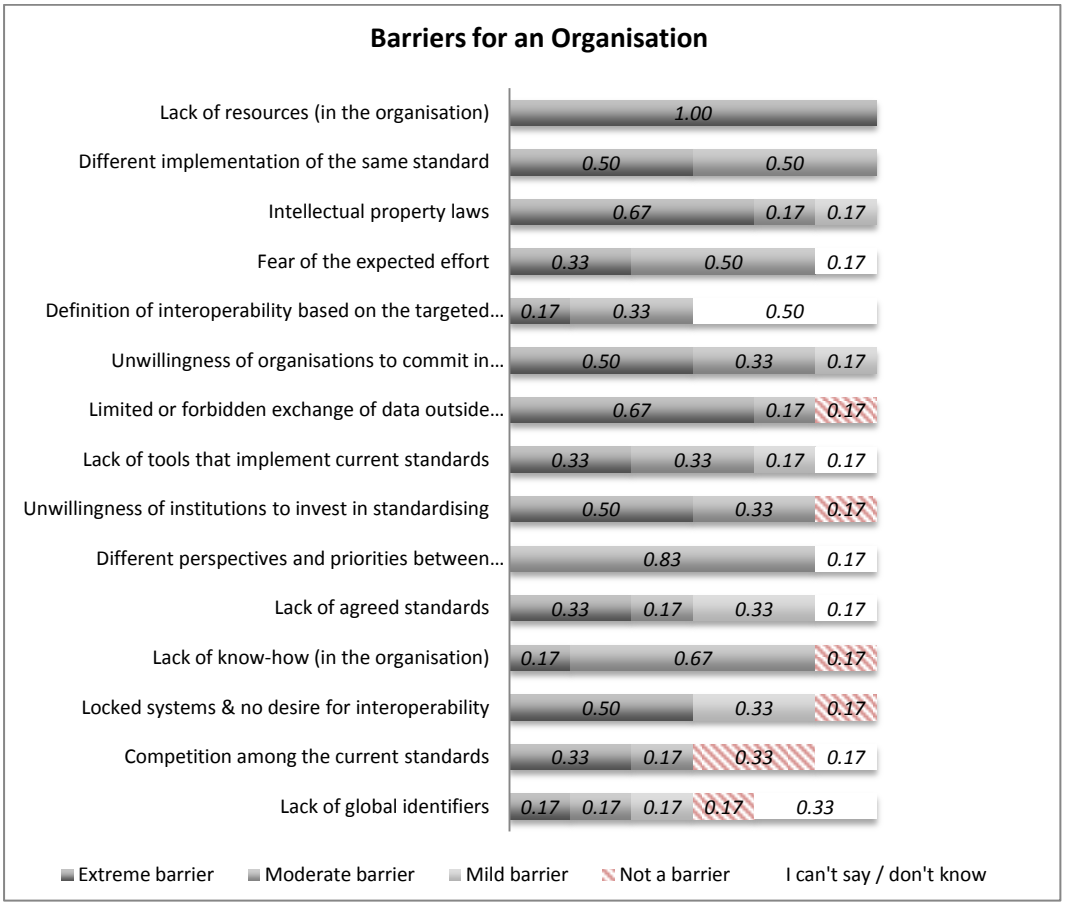

Chart 2: Response summary regarding the Barriers to interoperability for an organisation individually

Verification of an identified barrier had to be negated if it was assessed as "Not a barrier" for both cases of single organisation and entire community by at least one participant. Based on the results, all identified barriers were verified except the *competition among current standards* and the *unwillingness of institutions to invest in standardising*.

Furthermore, we evaluated the consistency of the group responses through analysing the standard deviation for the verified barriers. Thus, we can estimate the agreement among the participants for each barrier:

- The responses were most consistent for the barriers of *lack of resources (in the organisation)*, and *different perspectives and priorities between different communities*.
- The least agreement among the participants existed for a *lack of agreed standards*, and the barrier of *locked systems & no desire for interoperability*.
- In general, the responses for the community perspective were more consistent than for the view of a single organisation.

In addition, the impact of the barriers was in most cases higher for the community perspective than for the organisation's view. The strongest barriers from the view of a single organisation are the *lack of resources (in the organisation)*, *different implementations of the same standard*, and *intellectual property laws*. From the community perspective, the strongest barriers are *limited or forbidden exchange of data outside national borders*, *lack of resources (in the organisation)* and *intellectual property laws*.

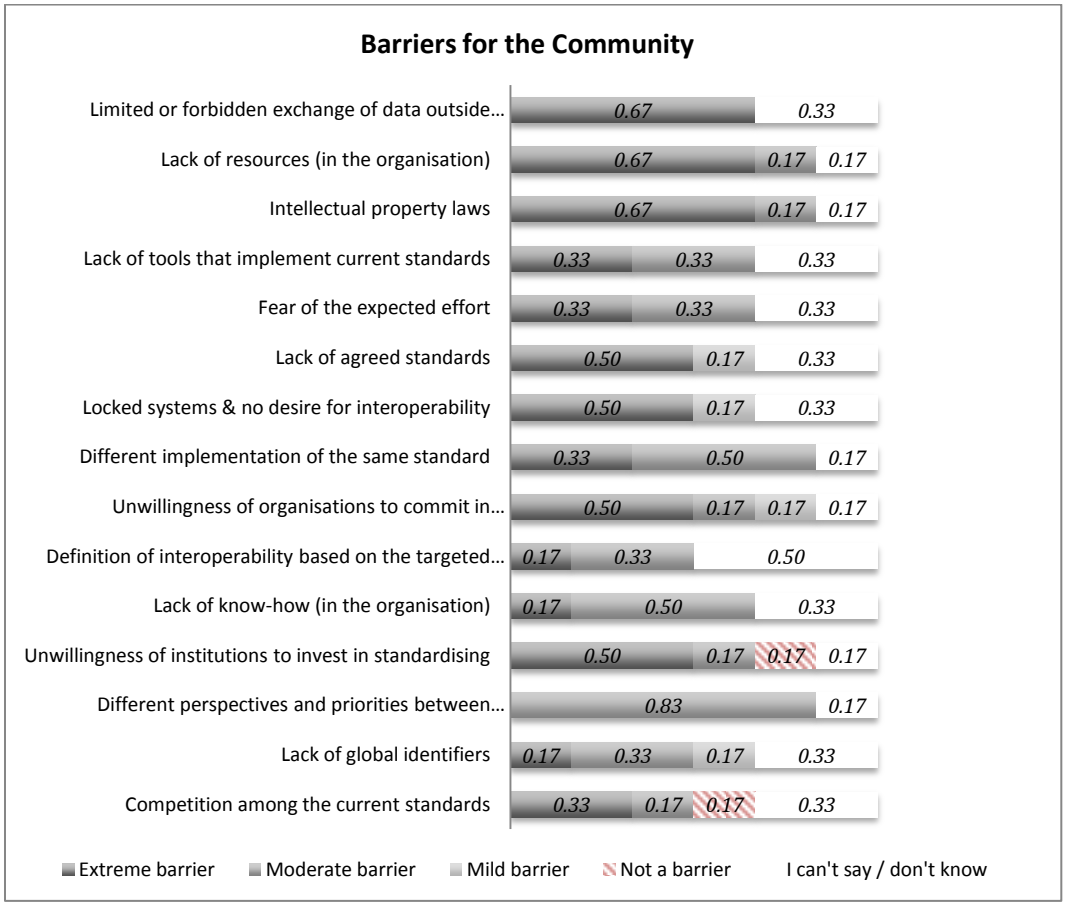

Chart 3: Response summary regarding the Barriers to interoperability for the community

Consensus, in the sense of agreement of participants not only to a point but also with exactly the same strength, was identified in the following points:

- The lack of resources is an extreme barrier for an individual organisation (and almost a high one also for community)
- The limited or forbidden exchange of data outside of national borders is an extreme barrier for the community
- Different perspectives and priorities between different communities is a moderate barrier

3.4.3 Suggested solutions

The initial question for changes that could improve the situation and help to overcome the current obstacles to the establishment of interoperability led to several suggestions. In the second round the participants were asked to evaluate these suggestions regarding *effectiveness* and *difficulty*; the effectiveness of each proposed solution/improvement and the difficulty to apply it.

The results regarding this section are included in Table 16 (Appendix). The evaluation of the suggested practices regarding their effectiveness is summarized in Chart 4 ranked from most effective to the least effective (on average).

One can notice that from the entire list of the 17 proposed ideas, the entirety of participants accepted 13 as at least somewhat solutions. Among them, some practices were accepted by the entirety of the panel as at least effective practices:

- Common APIs for search & retrieval
- Initiatives to necessitate the use of current standards
- Sharing of experiences, best practices & successful stories
- Clear definitions and terminology about digital preservation
- Involvement of people whose knowledge background is not confined to a specific community's aspect
- Better collaboration and stronger involvement of communities to each other's activities

The above reveal something quite interesting: changes people's attitudes and collaborations can significantly influence the status of the problem.

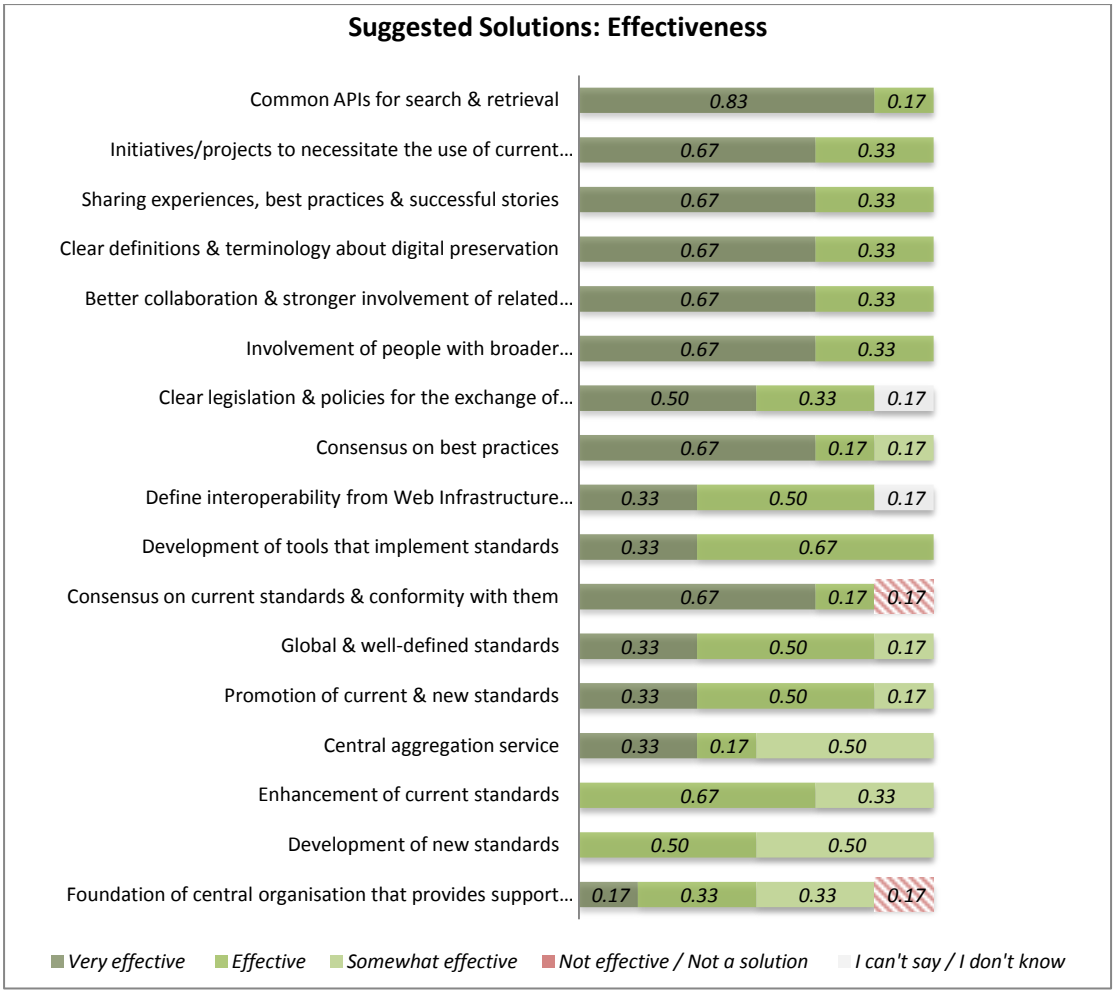

Chart 4: Response summary regarding the effectiveness of the suggested solutions/improvements

Regarding the difficulty to apply the proposed ideas, the responses are summarized in Chart 5, ranked from the easiest to the most difficult to realise (on average). We asked additionally for the evaluation of difficulty in order to reveal the practices that would be more effective and less hard to apply. This way we aimed to provide to the involved communities some efficient and relatively easy practices as directions that they could start with.

For verification of the identified solutions, we examine the results in Chart 4 and Chart 5 together. In Chart 4 two of the suggestions *Foundation of a central organisation that provides support for technical and legal issues* and *Consensus on current standards and conformity with them* were

assessed as "Not effective/ Not a solution". However, the assessment regarding difficulty included also the option "Not a solution". As we can see in Chart 5, no one of the statements was objected as a solution. In the particular case of the two participants, they gave a N/A answer in the second part. Therefore we can assume that more probably their first response can be interpreted as "not effective". Therefore, we consider all of the identified solution as generally verified.

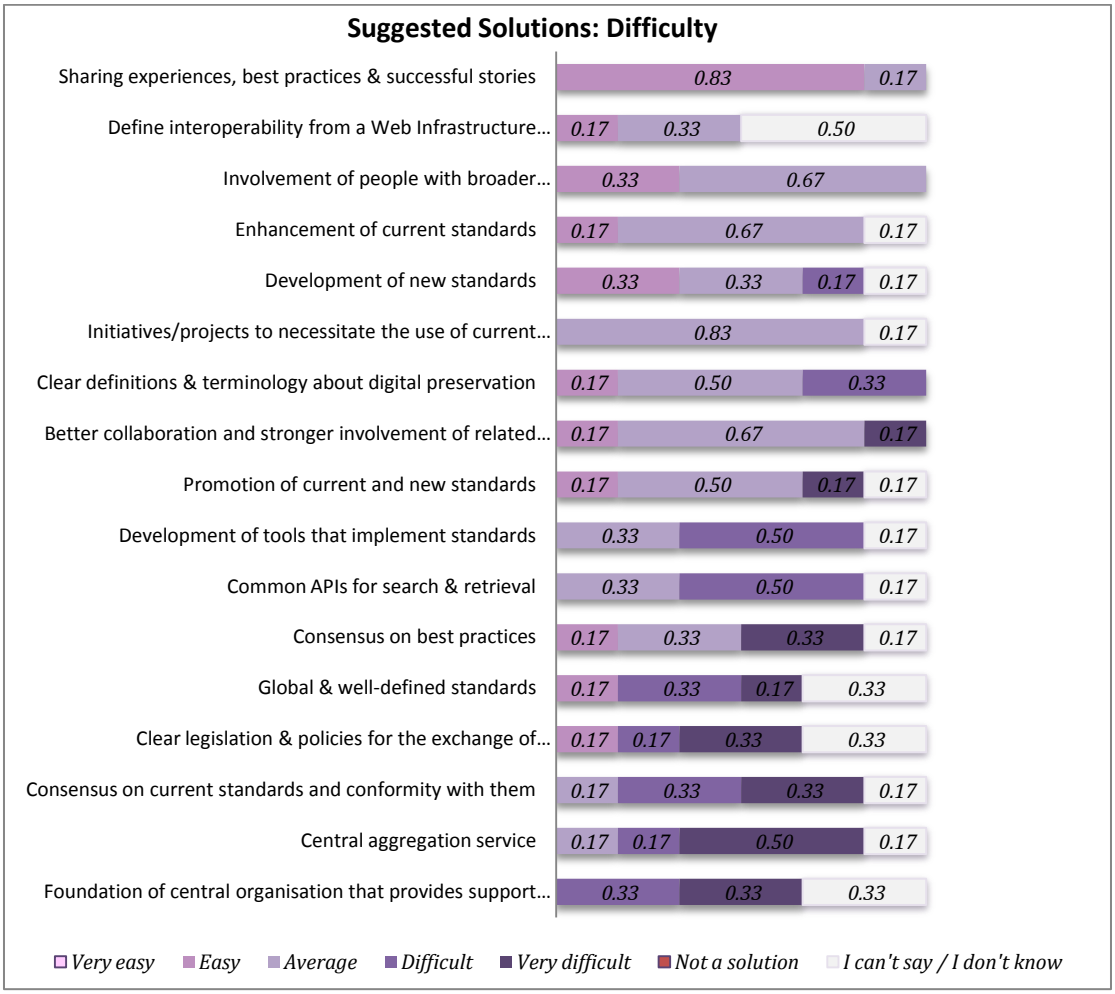

Chart 5: Response summary regarding the Difficulty of the suggested solutions/improvements

Chart 6 illustrates the proposed solutions projected both to the dimensions of effectiveness and difficulty. Each identified suggestion is plotted in this portfolio based its average effectiveness and difficulty as evaluated by the panel. Three clusters were identified:

1. **Highly recommended solutions** (area above the line from point *(very easy, somewhat effective)* to point *(very effective, difficult)*). These practices are considered as highly recommended because their estimated effectiveness is higher than the required effort to accomplish. The most promising practise is thereby sharing *of experiences, best practices & successful stories*.

2. **Recommended solutions** (area below the highly recommended solutions and above the line from point *(somewhat effective, average difficult)* to point *(very effective, very difficult)*). These practices can be assessed as efficient because their effectiveness still justifies their effort. It is notable that most of the solutions that are related to standards are located in this sector (A, C, D, E, F, and G).

3. **Inefficient solutions** (the bottom area). Solutions in this area can not be assessed as efficient because their effectiveness is much lower than the estimated effort to realise them.

With *a central aggregation service* and *a foundation of central organisation that provides support for technical & legal issues*, it is striking that the only two solutions that suggest a centralised service or institution are located in this sector.

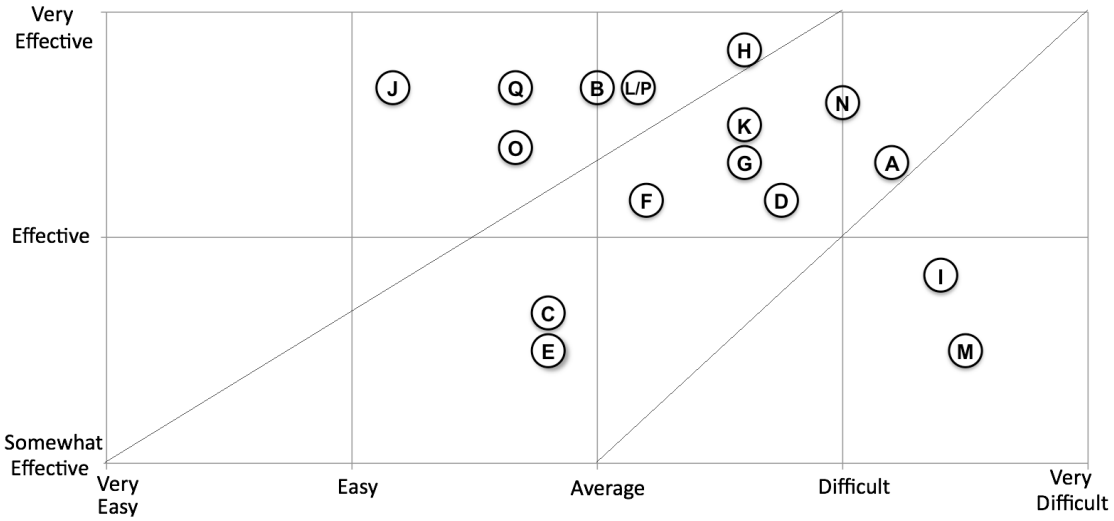

Chart 6: Portfolio of suggested solutions/improvements in both dimensions of effectiveness & difficulty

(A) Consensus on current standards & conformity with them

(B) Initiatives/projects to necessitate the use of current standards

(C) Enhancement of current standards

(D) Global & well-defined standards

(E) Development of new standards

(F) Promotion of current & new standards

(G) Development of tools that implement standards

(H) Common APIs for search & retrieval

(I) Central aggregation service

(J) Sharing experiences, best practices & successful stories

(K) Consensus on best practices (L) Clear definitions & terminology about digital preservation

(M) Foundation of central organisation that provides support for technical & legal issues

(N) Clear legislation & policies for the exchange of data/metadata

(O) Define interoperability from a Web infrastructure perspective instead of a system- to-system perspective

(P) Better collaboration & stronger involvement of related communities to each other's activities

(Q) Involvement of people with broader knowledge/ experience, not individually confined to community aspects

The highly recommended solutions area makes it more obvious that the changes should start from the involved people. If we could give a short summary of the three most promising practices (*J, Q* and *O*), people from the three involved communities, should start thinking more spherically and cooperatively, be open-minded and share the valuable knowledge. Furthermore, they should shift their thinking firstly, from their community's framework to more broadly beneficial aims, and, secondly, from the traditional system-based practices to the web infrastructure.

Last, an interesting addition to this section about legislation came as complementary comment from one participant in the second round:

> *Clear legislation can both promote and hinder interoperability.*

The participant explained further using as an example that the existence of clear legislation that, however, prohibits any exchange of data and metadata, would make interoperability a matter of little importance. Considering that legislation is one of the most significant factors that pose barriers to interoperability, this is certainly an interesting point for future research.

3.4.4 Challenges

For this part, in the second round, participants were asked to evaluate the priority for each of the challenges identified in the first round, as long as they accept them as challenges. All four challenges were indeed confirmed in the second round by the 6 participants. Therefore, the related to interoperability challenges that participants considered in the study, ranked by the highest priority are:

1. **The web resources become more and more complex**, new and complex media and web resources (Web 2.0, Social Media, etc.) demand enhanced methods for web preservation.
2. **Achieving interoperability of content**, consider digital libraries and web archives also as big datasets that should interoperate not only in terms of URIs and metadata but also in terms of content.
3. **Explosion of the volume of web data to archive**, as a result of the combination of the increasing efforts to archive as much of the web as possible and the immense growth of the web.
4. **New players are emerging in the field of web archiving** and, therefore, different systems, needs, and tools are emerging, the involved communities are increased and, as a result, interoperability may become more complex goal/affair.

Furthermore, it is worth mentioning the consensus of the panel that the increasing complexity of web is a high priority.

Therefore, the aforementioned 4 points are identified by this survey as issues of high importance that the involved communities have to consider and put in focus. The responses are summarized in Chart 7.

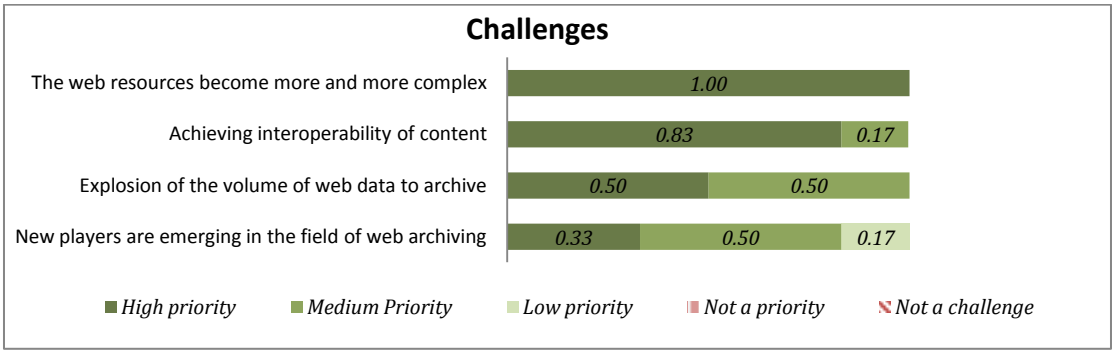

Chart 7: Response summary on fuTURE challenges

3.4.5 Perspectives on Information vs. System Interoperability

This part emerged in the second round motivated by some responses of the first round. It is partly based on actual answers, and partly extended with questions to examine it further.

Since the answers to this part, were quite conflicting and there was also a relatively high percentage of 'I don't know/ I can't say' answers we can't really provide a certain conclusion. Our aim with this addition is mainly to discuss and offer some topics for future exploration. The responses are illustrated in Chart 8 and Chart 9.

Few points that were clear, is that information interoperability is considered by all as a valid and reasonable way to establish interoperability, while in the case of system interoperability there was one disagreement. Furthermore, all participants agree on the fact that information interoperability is a broadly beneficial investment. However, one of them clarified to this point that it is beneficial in the case of more than two organisations, while when we speak for only two systems, a system-to-system is a better investment.

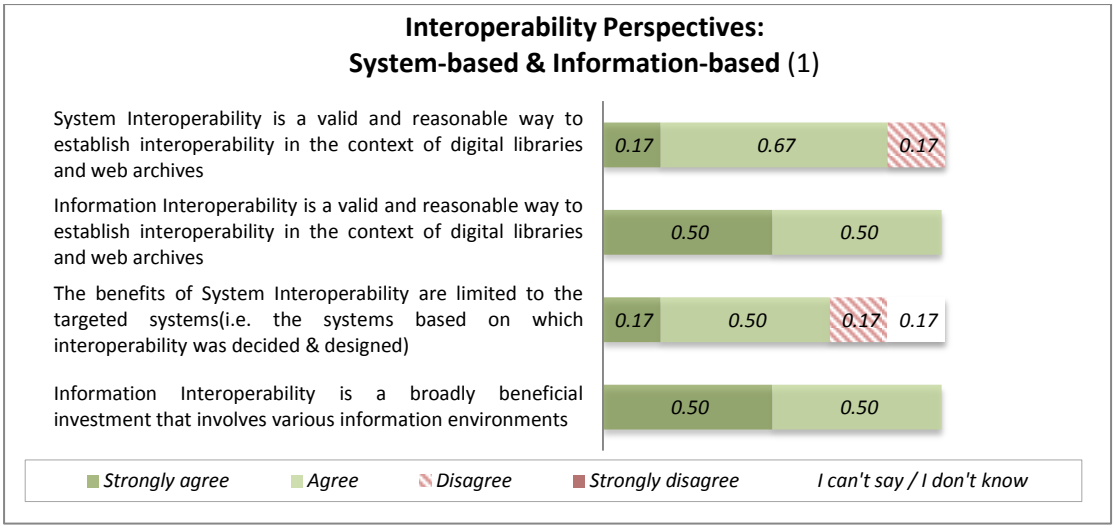

Chart 8: Response summary on interoperability perspectives (1)

Regarding the quality of interoperability that the two perspectives provide and the difficulty to realize them, there was no clear picture (Chart 9). Especially the quality seems to be a blur issue since half of the respondents answered "I can't say/don't know" and the rest gave different answers. The only part that seemed a bit clear is the sustainability of the two perspectives where information interoperability appears to be a more sustainable solution.

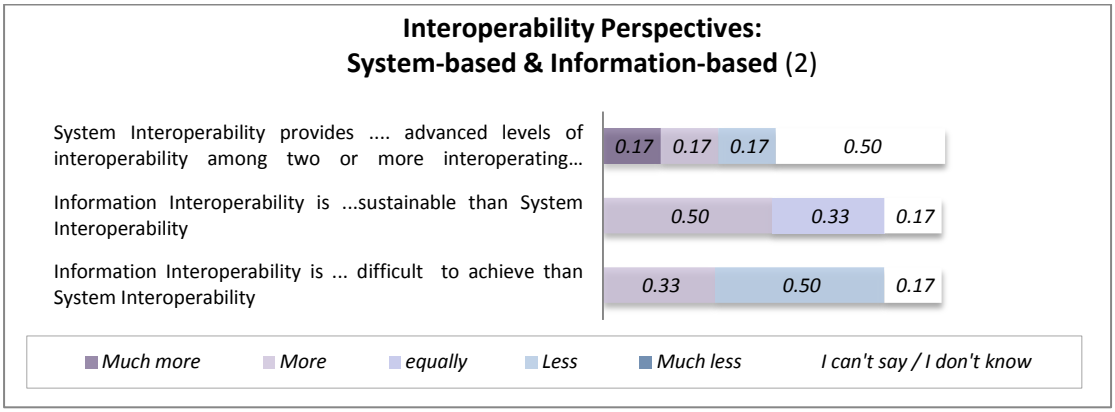

Chart 9: Response summary on interoperabiolity perspectives (2)

3.5 Discussion

In this section, we discuss the results of our Delphi study. As a first result of our study we identified several purposes or use cases that demand interoperability. The reasons for interoperation of web archives and digital libraries can be generalised into two aims. On the one hand, the user should be able to have access to collections or individual resources that are archived in one or more distinct repositories regardless of their location. This can be carried out by federated search, federated access, and through the exchange of objects in order to create a new collection. On the other hand, interoperation is required to establish the replication of objects into different locations, and, thus, reduce the risk of loss caused by possible threats. However, the identified purposes of interoperability were not as manifold as we expected. For example, interoperation that is necessary for sophisticated analysis on web archives, as well as, any interoperation demands for the ingest of new digital content into an web archive or digital library has not appeared in the participants' statements.

Additionally, we identified several benefits that are connected to interoperability. Thereby, the interdependence between collaboration and interoperability become apparent. For example, the

common agreement on specific standards for interoperation facilitates collaborative efforts for the development of tools as well as the knowledge exchange regarding common problems. This in turn facilitates higher levels of interoperability.

The identified barriers and the proposed solutions are connected by nature because a solution (or improvement) is conceived from the need to address one or more barriers. Therefore, the categories we identified are also similar for both. However, when we compare the identified barriers and solutions with the existing interoperability models (see section 2) two peculiarities have to be noticed. Firstly, perspectives that include also higher levels, e.g. the organisational level, seem to be more appropriate to consider interoperability for web archives and digital libraries comprehensively. A lot of problems on lower level can be addressed through further standardisation efforts while this is hardly possible on higher levels, e.g. the lack of knowledge or fears in the organisation. Secondly, a perspective or level that focus on legal issues is not mentioned explicitly in the reviewed models while it can be highly restrictive for interoperability attempts. Therefore, existing models for interoperability should be adapted in order to emphasise the importance of legal considerations, especially in the domain of web archives and digital libraries.

Another important finding is the identification of different ways to understand interoperability, and, thus, to establish the interoperation between different systems. Interoperability is most commonly considered as a task between two systems where both can take specific roles, for example a provider and a consumer of data (Athanasopoulos et al., 2011).Thus, the requirements are derived from the interoperation task and the systems characteristics, and the interoperability may be specifically adjusted to the corresponding systems even if the use of standards facilitates the same or similar interoperation with other systems. Contrary, the perspective of information interoperability abstracts from the specific systems, and aims on the provision of data as entities that support undetermined uses. Therefore, the entity must comprise or link all information that is necessary for processing in an undefined scenario.

In the second round of the study, almost all the results from the first round were verified and the evaluation allows further findings.

Federated search and federated access along with the exchange of the metadata seem to be more present as interoperability purposes than the replication and the exchange of primary objects.

The barriers that hinder or prevent interoperability are manifold. The most salient are the lack of resources to establish interoperability, different implementations of standards even if the same standard is used, intellectual property laws and limited or forbidden exchange of data outside national borders. They show that interoperability is dependent on organisational, legal, and technical aspects with little or no indication that one aspect may be more important than the other.

The evaluation of suggested solutions revealed that the most promising are these that comprise involvement or knowledge sharing of the community like sharing experiences, best practices & successful stories, involvement of people with broader knowledge & experience, clear definitions & terminology, and better collaboration and stronger involvement of related communities to each other's activities. On a lower level but still recommendable is the majority of solutions that are related to standards and tool development. However, the creation of centralised services or support institutions can be hard to recommend because the estimated impact does not legitimate the expected effort.

3.6 Conclusion

In this section, a Delphi study about aspects of interoperability of web archives and digital libraries has been presented. Some conclusions and implications for the BlogForever project should be drawn. The study has revealed the four main purposes of (a) federated search, (b) federated access, (c) exchange of data or content, and (d) replication. These purposes can be reused by the BlogForever project in order to reason and promote key features of the platform to target institutions or organisations. The results have further confirmed that technical issues are just a part

of the barriers that organisations have to overcome to establish necessary interoperability. While the BlogForever platform can only address directly the technical issues through the development of the platform, also other aspects could be considered in the project in order to facilitate a deployment of the platform, e.g. providing guidelines how to address organisational and legal issues. Furthermore, the results of the study revealed that general interoperability could not be taken for granted even if standards are implemented or supported due to degrees of freedom for the implementation of the standards. Therefore, statements about a general interoperability platform should be considered very carefully. The identified solutions to overcome barriers for interoperability can hardly be addressed by the BlogForever project. Most of the solutions require collaborative long-term effort of the web archiving and digital library community. However, the impact of decisions in the BlogForever project to support specific protocols and standards should not be underestimated. Given the positive scenario of a broad adoption of the BlogForever platform by various institutions, the support of specific standards can further boost their acceptance in the community.

4 Review of interoperability standards

The review in this section aims on the identification of existing standards that support interoperability, and their relevance for the BlogForever platform. The results of this work may help inform the development of the BlogForever platform directly.

4.1 Introduction

The scope of this review comprises

- Metadata and metadata standards that are currently used in the practice of interoperability,
- Protocols and standards associated with interoperability, and
- Technical aspects of interoperability.

In BlogForever, the aim of interoperability will be to share and exchange packets of metadata and content captured from blogs, and to assist with the digital preservation of blogs.

Our view on the use of interoperability standards is informed by the guidance from JISC Digital Media and their advice on management of digital collections[15]. Interoperability concerns resource discovery and sustainability, and can be understood as the ability of a collection to work alongside other collections. This can be done:

- Through shared resource discovery services, or
- By contributing metadata to other collections.

Interoperability from a technical perspective can be enabled:

- By the strict use of common standards, or
- By understanding how your 'non-standard' metadata can be mapped to or transformed to common standards.

In BlogForever, interoperability concerns the sharing of four entities:

1. Package of data,
2. Digital object,
3. Annotations, and
4. Collection of blogs.

The principal "entity" which we wish to share with others is a *package of data*, which represents the harvested content of a blog. This is closely connected with the *parsed metadata from blogs*, as captured by the spider and rendered by Invenio into a package of descriptive (mostly bibliographic) metadata.

The secondary shareable entity is *digital objects*, i.e. image files, audio files, moving image files, and attached documents (PDFs, text documents, spreadsheets, etc.) harvested from blogs. These objects have potential for reuse and repurposing in an interoperability scenario.

The third entity is *annotation*. Annotations are added to already archived objects by the repository users. Examples are tags, notes, structured comments, and links. Annotation Objects assist in the interpretation of the archived object, or give support or objections or more detailed explanations.

[15] From http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/metadata-standards-and-interoperability

The fourth potential entity is *collections of blogs*. When BlogForever, or a user of the platform, has aggregated sufficient collections of blogs, then further value can be obtained by the production of top-level cataloguing, indexing and tagging of such collections.

4.2 Methodology

This report focuses on standards used in the library and web archive domain. The list of standards is derived from the following sources:

- JISC Digital Media advice[16]
- Minitex (An Information and Resource Sharing Program of the Minnesota Office of Higher Education and the University of Minnesota Libraries)[17]
- DCC Briefing Paper on metadata standards[18]
- Library of Congress presentation on metadata standards[19]
- Lois Mai Chan's study of Metadata interoperability[20]

In all cases we concentrate on standards which are probably or potentially useful for interoperability. We have followed a simple structure to ensure coverage of standards for interoperability:

1. *What we need to know about an object*
2. *How it is encoded*
3. *How is it transmitted and accessed*

This is directly inspired by the Project Bamboo wiki on Candidate Collections Interoperability Standards[21], which proposed classes of standards as a way of identifying potentially relevant interoperability standards, protocols, application profiles, and best practices.

The standards and protocols are distinguished according to the categories of the JISC standards catalogue[22] (see Table 1).

Table 1: Standard categories and standards

Document Standards	Plain text docs	ASCII, Unicode
	Binary text docs	DOC, RTF, ODF, PDF
	Markup text docs	SGML, XML
Web Standards	Web format standards	HTML (including XHTML), CSS, DOM
	Web services	SOAP, UDDI, WSDL, REST

[16] http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/metadata-standards-and-interoperability

[17] http://www.minitex.umn.edu/Digitization/Standards/

[18] http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards

[19] http://www.loc.gov/standards/mods/presentations/intro-diglibstandards-ala07/intro-diglibstandards-ala07.ppt

[20] http://www.white-clouds.com/iclc/cliej/cl19chan.htm

[21] https://wiki.projectbamboo.org/display/BTECH/Candidate+Collections+Interoperability+Standards

[22] http://standards.jisc.ac.uk/catalogue/Home.phtml

	XML standards	XML, XML Schemas, XML Namespaces, XSLT
	Linking standards	XLink, XPointer
Image Standards	Vector Image standards	SVG, Flash, VML
	Raster image standards	BMP, GIF, JPEG, PNG, TIFF
	Image standards for web	SVG, Flash, GIF, JPEG, PNG
Metadata Standards	Resource Discovery Metadata Standards	Simple Dublin Core, Dublin Core, Encoding Bibliographic Citation Information in Dublin Core, OAI DC, IESR Metadata Schema
	Digitisation Metadata Standards	VRA, CDWA
	Other Metadata	METS, DIDL, DDI, OWL, RDF, RDFS, SKOS Core, PREMIS, IEEE LOM, UK LOM Core
Search Protocol standards		Z39.50, Bath Profile, SRW, SRU
Internet Transport Standards		HTTP
Metadata Harvesting Standards		OAI-PMH
Character Encoding Standards		ISAG, EAD DTD, TEI DTD, CIMI, MARC, MODS
Identifier Standards		URI, OpenURL, Z39.88-2004, DOI, PURL, Handle, ARK, INFO, COinS
Geographic Information Standards	GIS	GIS, GIS Metadata and Open Geospatial Consortium Specifications
Multimedia Standards	Multimedia Standards	SMIL, Flash
	Multimedia Containers	ASF, AVI (aka WMF), OGG (aka OGM), MPEG, RealMedia, RIFF, WAV
	Encoding standards	MPEG-1, MPEG-2, MPEG-4, AAC, AIFF, Flac, MP3, Ogg Vorbis, RA and WMA (audio), Dirac, H.263, MJPEG, Theora, WMV (video)

We will also discuss the web-archiving standard format WARC, standards for cataloguing, and other models where they may be relevant.

4.2.1 Structure of each report

We report on each standard following this structure:

- Name of standard

- Type of standard
- Main URL or URLs, i.e. the "home page" of the standard
- Purpose of the standard
- Who uses the standard?
- Further background information
- BlogForever deliverables which reference the standard
- Relevance and applicability for the BlogForever platform
- References, i.e. URLs for use cases, descriptions and interpretations of the standard, and/or its implementation

We have found it useful to bring out and restate relevant sections from the Description of Work (DoW) and published deliverables of the project, firstly to ensure that our findings about standards do not duplicate work already done, and secondly to ensure that any discussions or suggestions in this report do not depart from the intentions of the project.

We think it is particularly important to ensure that:

- Interoperability can support some of the aims of digital preservation
- The standards can be supported by the Invenio platform
- The standards are widely adopted and supported
- The standards are used by BlogForever's target audience / user base / potential partners

4.3 Metadata standards

Metadata is descriptive or contextual information which refers to or is associated with another object or resource. This usually takes the form of a structured set of elements which describe the information resource and assists in the identification, location and retrieval of it by users, while facilitating content and access management. A metadata standard will normally support a number of defined functions, and will specify elements which make these possible. Metadata standards may support many functions: Descriptive Metadata; Technical Metadata; Administrative Metadata; Structural Metadata; Preservation Metadata; and Rights Metadata (Higgins, 2007).

4.3.1 MARC 21 / MARC XML (MAchine Readable Cataloging)

Type of standard

- Character encoding standard

URLs

http://www.loc.gov/marc/bibliographic/

http://www.loc.gov/standards/marcxml/

Purpose of the standard

MARC is a metadata standard used to exchange bibliographic data in machine-readable form between integrated library systems. MARC is also a data structure standard used for describing bibliographic materials.

MARC was developed by the Library of Congress to facilitate the creation and dissemination of cataloging between libraries. There were created several versions of MARC with most predominant MARC21, developed in 1999 as an effort to harmonize the US format version, the Canadian one and UNIMARC.

MARCXML schema was developed in 2002 as an alternative record structure, allowing MARC records to be represented in XML. It is used to expose records via a web service or following the SRU or OAI-PMH standards.

Who uses the standard?

Primarily the library community uses the MARC format. It is used by digital libraries to encode and share information about books and other material they collect. MARC XML is intended for use by institutions already using MARC. One might use MARC XML to represent a complete MARC record in XML or to represent metadata for OAI harvesting.

In addition to the library community, library stock suppliers and the book trade also use MARC formats to varying degrees. Library stock suppliers can provide MARC format records for actual items supplied. Bibliographic data suppliers provide pre and post-publication records.

Further background information

MARC can be reused in an XML environment using MARCXML. MARCXML uses the MARC data element set in XML syntax.

The MARC standards define three aspects of a MARC record: the record structure, the field designations within each record, and the actual content of the record itself. MARC records are typically stored and transmitted as binary files (usually concatenated records in a single file).

Each field in a MARC record provides information about the corresponding item that the record describes. It uses a 3-digit code number to identify each field in the record. (e.g. 100 defines the primary author, 245 the title etc.).

MARC is a metadata transmission standard, not a content standard.

The MARC 21 formats except from Bibliographic Record Format, additionally includes: Authority Record Format, Holdings Record Format, Classification Record Format and Community Information Format.

BlogForever deliverables which refer to the standard

- Deliverable D3.1, Section 5.3.1

"MARC XML is an XML schema based on the fairly common MARC21 standard. MARC (MAchine-Readable Cataloging) is a data format and set of related standards."

- Deliverable D4.4, Section 2.1

"MARC is the standard format in the library world. It is well established and has been used since 1960s."

- Deliverable D4.5

The report contains details of how it is anticipated that MARC will be implemented in the platform. Of special interest to us is the implementation of export of content in XML formats, e.g. Repository Feature RF59, Export data using XML.

Relevance and applicability for the BlogForever platform

MARC allows interoperability with other XML schemes by taking advantage of free XML tools. It also allows for collaborative use of metadata for access (e.g. OAI).

Invenio is using the recently adopted standard MARC XML.

The Invenio system is using the MARC 21 standard to represent all the bibliographic data of blogs, and storing it in the database. Invenio is also capable of exporting the contents of a harvested blog to MARC XML. MARC is thus an essential standard for the functioning of the BF platform.

Invenio supports two important aspects of interoperability with its outputs:

1. Bibliographic metadata
2. Blog posts rendered in XML

The bibliographic metadata from BlogForever could potentially be used by any library which is committed to these exchange formats. The formats facilitate the transfer of bibliographic data between systems. Use of these standards reduces the duplication of effort in different libraries acquiring and cataloguing the same material. The potential for interoperability might include merging BF's bibliographic data from blogs with existing catalogues for collections of academic papers.

By allowing export of blog content into XML, BlogForever could theoretically exchange the entire contents of a collection with any repository capable of storing XML.

References

http://blog.lib.umn.edu/chapm157/metadata/024723.html

http://standards.jisc.ac.uk/

http://en.wikipedia.org/wiki/MARC_standards

4.3.2 METS (Metadata Encoding and Transmission Standard)

Type of standard

- Metadata standards: Other metadata

URLs

http://www.loc.gov/standards/mets/

Purpose of the standard

METS (Metadata Encoding and Transmission Standard) is intended to provide a standardized XML format for transmission of complex digital library objects between systems, so its value for interoperability is clear. METS was originally intended for digital libraries, which found the standard useful to express the hierarchical nature of a digital book or a library collection and model these in XML.

With its flexibility for expressing a wide range of metadata, METS can be used to model almost any digital object, as the BF project has demonstrated with its ideas about modelling of blogs.

Who uses the standard?

The METS community includes University Libraries, Archives, and Museums. The institutions, which have chosen to register their implementation, can be found on the METS Implementation Registry [23].

Further background information

The METS is a schema for encoding descriptive, administrative and structural metadata regarding objects within a digital library, expressed using the XML schema.

METS provides the means to convey the metadata necessary for both the management of digital objects within a repository and the exchange of such objects between repositories (or repositories and their users). It provides a mechanism for recording the relationships that exist between pieces of content and between the content and metadata that compose a digital library object.

A METS document comprises of the following sections:

- Header (metadata about the creation of the METS file like editor, agent, time of creation etc)
- Descriptive metadata (metadata that describe the preserved object)
- Administrative metadata (intellectual property rights, provenance, technical metadata regarding the content, information about the analogy source document)
- Behavioural metadata (executable behaviours with the content of the object encoded)
- File Section (a list of all content files that comprise the digital object that is described and their location)
- Structural Map (hierarchical structure for the digital object, links of the elements of the structure to content files and metadata that concern each element)
- Structural Link (hyperlinks between nodes in the hierarchy outlined in the Structural Map)

From the above, only the structural map is mandatory.

BlogForever deliverables which refer to the standard

- Deliverable D3.1, Section 5.5

"In the BlogForever project we have decided to use METS as the standard to keep all the metadata needed for the blogs archive. In this document we will describe a draft idea of how to use METS together with other formats identified in the previous sections."

Furthermore, see pages 114-115 of D3.1 for proposed implementation with specific reference to how METS could express different views of a blog or ways of modelling it differently

- Deliverable D4.4, Section 3.2.2

This report expressed the plan to encode the Information Packages in METS. It is envisaged that both the SIP and the AIP would be expressed in a METS wrapper. "The metadata for each component will be wrapped, encoded, and exposed using METS."

- Deliverable D4.5

The implementation of the platform describes the plan to transform the submitted METS package from the spider into MARC XML (pages 22-23).

Using RF59, the platform is also able to export the content as a METS file (page 44).

[23] http://www.loc.gov/standards/mets/mets-registry.html

Relevance and applicability for the BlogForever platform

BlogForever intends to use METS for the following outputs:

- Submission information packages in METS
- Archival information packages in METS
- Technical metadata for objects in METS - and see Section 4 of this report on object types
- A means of declaring and storing significant properties, as shown in the trial use of FITS (File Information Tool Set).

The National Library of Australia (NLA) experience has shown that:

METS can be used as a means of transmitting a representation of an object (physical or digital or partially digital) from one system to another. It can:

- Fully describe the object and its components.
- Encode the metadata needed to aid its preservation and future access.
- Represent the physical and/or logical structure of highly complex objects.
- Represent collections of objects, even where these objects are not stored in the same repository.
- Support a range of submission and dissemination scenarios.
- Deliver representations of an object appropriate to the scenario by using a protocol such as OpenURL to request the required parameters.

METS is also potentially useful for data exchange. A METS document may be expressed as a unit of storage or a transmission format.

In conclusion, METS potentially provides a means to exchange metadata and digital resources. It potentially could be used for transmission of a blog in METS to any institution platform that is also subscribing to the METS standard. Where MARC expresses the bibliographic metadata, METS declares metadata on the structure of the blog or blog posts, preserves the integrity of the archived blog, and also supports its component digital objects. METS is hence invaluable for a preservation strategy.

References

http://fedora-commons.org/download/2.2/userdocs/digitalobjects/rulesForMETS.html

http://ejournals.bc.edu/ojs/index.php/ital/article/view/1917

http://dlib.nyu.edu/metstools/

http://www.dlib.org/dlib/september08/dappert/09dappert.html

http://www.iwi-iuk.org/cashmere/htdocs/html/newsletter/data/mets.en.shtml

http://www.dlib.org/dlib/march08/pearce/03pearce.html

http://standards.jisc.ac.uk/

DL.org Digital Library Technology & Methodology Cookbook

Metadata Encoding and Transmission Standard: Primer and Reference Manual, 2010, Digital Library Federation

4.3.3 MODS (Metadata Object Description Schema)

Type of standard

- Character encoding standard

URLs

http://www.loc.gov/standards/mods/

Purpose of the standard

MODS is a bibliographic element set that can be used for a variety of purposes, and particularly for library applications. MODS is intended to complement other metadata formats. As an XML schema it is created to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. The standard takes a similar approach to resource description as MARC, with some rearranging, removing, and adding of data elements.

Furthermore, MODS offers more potential uses; It can be used as an extension schema to METS, to represent metadata for harvesting and for original resource description in XML syntax.

Who uses the standard?

It is primarily used by libraries and intended for use by library applications.

Further background information

MODS is an XML-based bibliographic schema, designed as a compromise between the complexity of the MARC format used by libraries and the extreme simplicity of Dublin Core metadata. That's because the element set is richer than Dublin Core but more simplified than the MARC format, and the schema is more end user oriented than MARC XML.

There is a complete list of elements, subelements and attributes[24]. The top-level elements in MODS are:

titleInfo	classification	language	part
note	genre	location	tableOfContents
name	relatedItem	physicalDescription	extension
Subject	originInfo	accessCondition	targetAudience
typeOfResource	identifier	abstract	recordInfo

The elements generally inherit the semantics of MARC and several of them have optional ID attribute to facilitate linking at the element level. Furthermore, MODS does not assume the use of any cataloguing code.

BlogForever deliverables which refer to the standard

- Deliverable D3.1, Section 5.3.3.
  "MODS…has potential for BlogForever, but it appears to be providing catalogue access at

[24] http://www.loc.gov/standards/mods/v3/mods-userguide-elements.html

a fairly limited level and thus may not offer enough richness of detail for describing blog content."

- Deliverable D4.4
  The BibConvert module is part of Invenio: "BibConvert allows conversions between various sequential and semi-structured formats, such as MODS (Metadata Object Schema)." (p 9)

Relevance and applicability for the BlogForever platform

WP3 concluded that MODS has some potential for the platform, but anticipated problems with the restrictions of the catalogue elements in MODS. Indeed MODS appears to function by the removal and rearrangement of selected data elements. However, it is frequently used as a descriptive metadata structure standard inside METS metadata wrappers for storage or exchange of digital objects. The capability of Invenio's BibConvert tool to support MODS should not be overlooked, meaning that MODS may have interoperability potential. Additionally it can be used to represent metadata for harvesting (OAI).

References

http://blog.lib.umn.edu/chapm157/metadata/024723.html

http://www.dlib.org/dlib/september08/dappert/09dappert.html

http://www2.archivists.org/standards/metadata-object-description-schema-mods

http://www.loc.gov/standards/mods/mods-overview.html

4.3.4 Dublin Core

Type of standard

- Metadata standards: Resource discovery metadata standard

URLs

http://dublincore.org/

Purpose of the standard

The Dublin Core set of metadata elements provide a small and fundamental group of text elements through which most resources can be described and catalogued. Using only 15 base text fields, a Dublin Core metadata record can describe physical resources such as books, digital materials such as video, sound, image, or text files, and composite media like Web pages. Additionally, it can be extended and combined with terms from other vocabularies for the definition of Application Profiles.

Who uses the standard?

Any institution using CONTENTdm (Online Computer Library Center's Digital Collection Management software) uses Dublin Core to describe their digital content. It is adopted be several European and international projects, including national libraries or vast web databases like

Musicbrainz[25]. It is also employed to describe several resources under the principles of Linked Open Data.

Further background information

The 15 base properties that comprise the Dublin Core element set are optional and repeatable, and in detail are: contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title and type.

Dublin Core can be expressed using:

- The DC-Text format
- DC-HTML, using HTML/XHTML meta and link elements
- DC-DS-XML, XML (DC-DS-XML)
- DC-RDF, using the Resource Description Framework (RDF)

BlogForever deliverables which refer to the standard

- Deliverable D3.1
  Section 5.3.2 mentions the standard. In Section 5.4.3, Dublin Core and Qualified Dublin Core are discussed as possible means of declaring Rights metadata.
- Deliverable D4.4
  The BibConvert module is part of Invenio. "BibConvert allows conversions between various sequential and semi-structured formats, such as Dublin Core)." (p 9)
- Deliverable D4.5
  The Invenio platform supports the export of data into Dublin Core. (See RF 8).
- Deliverable D2.3
  The standard is one of the core parts of the particular deliverable since it is an essential metadata schema. Section 3.3 describes the important vocabularies for exposing semantics and contains a more extensive description of the standard. Section 3.4 describes in detail how each one of the Dublin Core properties can be matched to the BlogForever data model.

Relevance and applicability for the BlogForever platform

Technically, the platform has the capability to support this standard. The Invenio platform already supports the export of data into Dublin Core (See RF 8). Beyond this, WP3 has not made a decision about whether Dublin Core ought to be used, how it could be used, or which metadata elements from a parsed blog should be expressed in Dublin Core. There may remain a mapping exercise to be undertaken here.

As to interoperability, the page http://dublincore.org/documents/interoperability-levels/ indicates that there are many design choices involved in designing applications for different types of interoperability. At Level 1, applications use data components with shared natural-language definitions. At Level 2, data is based on the formal-semantic model of the W3C Resource Description Framework. At Level 3, data is structured as Description Sets (records). At Level 4, data content is subject to a shared set of constraints (described in a Description Set Profile).

References

[25] http://musicbrainz.org/

http://dublincore.org/documents/dces/

DL.org Digital Library Technology & Methodology Cookbook

4.3.5 PREMIS (PREservation Metadata: Implementation Strategies)

Type of standard

- Metadata standard: Other metadata

URLs

http://www.loc.gov/standards/premis/

http://www.loc.gov/standards/premis/schemas.html

Purpose of the standard

PREMIS is a practical resource for implementing preservation metadata in digital archiving systems. Preservation metadata is information that supports digital preservation processes. A key component of archival systems, metadata helps to ensure that digital materials remain usable over the long term.

Who uses the standard?

See the PREMIS Implementors Group[26]. PREMIS is potentially useful for cultural heritage institutions, businesses, and government agencies with collections of digital materials.

Further background information

PREMIS is expressed using a data dictionary and an XML schema.

The Data Dictionary

The PREMIS Data Dictionary defines a core set of semantic units that repositories should know in order to perform their preservation functions.

Despite the fact that preservation functions may vary from one repository to another, they generally include actions to ensure that digital objects remain viable and renderable, that digital objects in the repository are not inadvertently altered, and that legitimate changes to objects are documented.

The Data Dictionary is not intended to define all possible preservation metadata elements, but only those that most repositories will need to know most of the time. Therefore, it excludes several types of metadata like format-specific metadata, implementation-specific metadata, business rules, information about rights and permissions that do not directly affect preservation functions. PREMIS also excludes descriptive metadata and therefore other independent standards can be used for this purpose (like MARC21, MODS and Dublin Core).

However, with extension containers, which are designed to give place to non-PREMIS metadata to be recorded, it can be extended to include metadata that is out of scope or not included in the Data Dictionary.

In general, PREMIS defines a subset of all the metadata needed by an organisation running a preservation repository. It defines only the metadata which are commonly needed to perform preservation functions on all materials.

[26] http://www.loc.gov/standards/premis/pig.html

The primary uses of PREMIS are for repository design, repository evaluation, and exchange of archived information packages among preservation repositories.

It should be noted that PREMIS Data Dictionary defines semantic units, not metadata elements. Therefore, it does not define how metadata should be represented in a system, but what the system needs to know and should be able to export to other systems.

The PREMIS data model defines 5 kinds of entities:

- *Intellectual Entities*, a set of content that is considered a single intellectual unit for purposes of management and description
- *Objects*, what is usually stored and managed in the preservation repository
- *Agents*, actors that have roles in events and in rights statements
- *Events*, the entity that aggregates information about actions that affect objects in the repository
- *Rights*, the entity that aggregates information about rights and permissions that are directly relevant to preserving objects in the repository

PREMIS in XML

The PREMIS Maintenance Activity provides an XML schema that corresponds directly to the Data Dictionary to provide a straightforward description of objects, events, agents and rights.

When PREMIS is used for exchange, it is expected (but not required) to be represented in XML. In practice, most of the preservation systems already use XML formats to import and export data.

PREMIS and METS

It is possible for PREMIS to be used inside of METS, but this cannot be done entirely straightforward. First, METS breaks up information into different sections according to whether it is technical, rights, or provenance metadata while the PREMIS schema has sections for objects, rights, events and agents. There is indeed some correspondence between the two structures but it isn't flawless. Secondly, PREMIS and METS have some overlap and if the two are used together, it has to be decided whether to record these overlapping elements in PREMIS sections, METS sections, or both.

Since such variations in the use by every preservation repository mean variation in how the data is represented, and, consequently, impede interoperability, there are several efforts in process to help define best practices for using PREMIS and METS together.

BlogForever deliverables which refer to the standard

- Deliverable D3.1
  Section 5.4.2 records the decision to recommend PREMIS as best standard for this purpose. It is also suggested in 5.4.3 as a means for expressing Rights Metadata. PREMIS is specifically recommended within the preservation workflow, section 6.1.4. Elements of PREMIS are also encoded in the draft METS profile, Appendix A, as an example.
- Deliverable D4.4
  See Section 3.2.3, the repository workflow. This expressed the intention to use PREMIS for provenance and preservation metadata, and notes its potential to express rights metadata.

Relevance and applicability for the BlogForever platform

PREMIS is part of the preservation strategy and is potentially very useful for encoding Rights Metadata, depending on results of deliverable D3.3. However, note that rights management in

PREMIS limited to permissions regarding actions taken within a repository. This means it won't work for resolving copyright issues.

If used with METS, note there are some conflicts between METS and PREMIS, as noted in Rebecca S. Gunether's DLib article [27].

For interoperability purposes, PREMIS is regarded as a standard for exchanging information packages between repositories. If BF's Information Packages are correctly rendered in PREMIS, this qualifies as interoperability.

References

http://www.loc.gov/standards/premis/understanding-premis.pdf

http://www.oclc.org/research/activities/premis-rlg.html

http://listserv.loc.gov/listarch/pig.html

http://www.loc.gov/standards/premis/pig.html

http://www.loc.gov/premis/v2/premis-2-0.pdf

4.4 Digital object standards

Digital object standards are any set of technical data elements required to manage particular types of object collections. For example the standard NISO Metadata for Images in XML (NISO MIX) defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files.

4.4.1 TextMD, MIX, AES57, VideoMD, DocumentMD

Type of standard

- Technical metadata standards

URLs

http://www.loc.gov/standards/textMD/

http://www.loc.gov/standards/mix/

http://www.aes.org/standards/schemas/aes57-2011-08-27.xsd

http://www.loc.gov/standards/amdvmd/index.html

http://www.fcla.edu/dls/md/docmd.xsd

Purpose of the standards

These are all XML Schemas designed for expressing technical metadata for certain types of digital object. For the project we looked at the following object types within WP3:

- Textual objects
- Images

[27] http://www.dlib.org/dlib/july08/guenther/07guenther.html

- Audio
- Moving images
- Documents

They can all be expressed within a METS schema.

These standards are in turn endorsed or backed up by other standards. For example MIX is based on the Technical Metadata for Digital Still Images Standard, NISO Z39.87. The MIX schema offers a way to implement NISO selectively.

Who uses the standards?

Digital libraries.

Further background information

All of these standards can be expressed in XML Schemas.

BlogForever deliverables which refer to the standard

- Deliverable D3.1
  Section 5.4.1 on Technical Metadata goes into detail about each of these standards and why they have been recommended for the support of common digital object types in BlogForever.
- Deliverable D4.4
  Describes how the spider will gather MIX metadata for images. In the 3.2.3 workflow section, the standards are explicitly named as "metadata related to renderability" (p 35).

Relevance and applicability for the BlogForever platform

If the project implements these object standards, and Invenio is able to support them with appropriate metadata schemas, such action increases the chances of preservation of common digital object types, thus assisting long-term support for blogs between institutions.

4.5 Protocol standards

A protocol is simply an agreed format for transmitting data between two devices. Protocols are used for many purposes (e.g. networks, communication, the internet); a protocol is a "set of rules or conventions formulated to control the exchange of data between two entities desiring a connection." (Kumar, 2009). The BlogForever project's interest is in specialist protocol standards that allow particular types of data transmission and exchange that are potentially useful for interoperability.

4.5.1 OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)

Type of standard

- Metadata harvesting standard

URLs

http://www.openarchives.org/pmh/

Purpose of the standard

A low-barrier mechanism for repository interoperability. It provides an application-independent interoperability framework based on metadata harvesting. It makes descriptive metadata about resources harvestable.

"In general, the Open Archives Initiative's (OAI) preferred method of re-use of repository data is harvesting. This differs from web crawling in that harvesting gathers data in structured XML formats - i.e. retaining separate fields for authors, titles, dates, and so forth - whereas web crawlers deal with everything as one big text. Structured data not only provides opportunities for richer search services, but also facilitates data analysis and data mining." (JISC Repositories Support project).

The above description is a very close match to what BlogForever will be providing, and why BlogForever is different to many conventional approaches to web crawling.

Who uses the standard?

The OAI Protocol has become widely adopted by many digital libraries, institutional repositories, and digital archives. See for example the list of Registered Data Providers at OpenArchives [28].

About 75% of repositories worldwide (~85% in the UK) provide an interface that uses the standard Open Access protocol OAI-PMH. Such repositories are designated 'OAI-compliant'.

Commercial search engines have started using OAI-PMH to acquire more resources.

Google is using OAI-PMH to harvest information from the National Library of Australia Digital Object Repository. NASA's Mercury: Metadata Search System uses OAI-PMH to index thousands of metadata records from Global Change Master Directory (GCMD) every day. [29]

Further background information

OAI-PMH is a protocol that provides an application-independent framework for metadata transfer. It was designed to offer easy implementation (based on widely accepted standards such as HTTP, XML and Dublin Core) and high efficiency.

According to the OAI–PMH framework, there are two actors: a data provider and a service provider. A data provider uses OAI–PMH to expose metadata about repository content to the service provider(s) and maintains one or more repositories. Similarly, the service provider uses OAI–PMH to harvest metadata from the data provider(s). In the context of OAI–PMH, the term harvesting refers to the act of collecting metadata from different repositories and the possible storing of all metadata in a central database. In OAI-PMH, the metadata is distributed and replicated in many different places and, potentially, provides a highly redundant and fault-tolerant system.

OAI-PMH uses XML over HTTP and XML Schemas to define record formats. Any kind of metadata is possible to be exchanged using OAI-PMH as long as it encoded in XML and defined with an XML Schema. OAI-PMH mandates the OAI_DC schema as a minimum standard for interoperability.

OAI-PMH documentation also describes the use of XML schema for other formats, and provides additional XML schemas for:

- rcf1807 (for RFC 1807 format metadata)

[28] http://www.openarchives.org/Register/BrowseSites

[29] http://en.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting

- marc21 (recommended for MARC21 metadata, provided by the Library of Congress)
- oai_marc (for MARC format metadata)

While OAI-PMH is intended as a machine-to-machine interface, it returns results as XML, which can also be displayed on web browsers for human consumption. EPrints.org provide a useful XML stylesheet for rendering OAI-PMH output that is used by many repositories that run EPrints software.

It should be emphasized that OAI-PMH is a protocol for the exchange of metadata only and does not provide mechanisms to expose and harvest full content.

There are 6 services requests or "verbs" defined in OAI-PMH:

- *GetRecord*, to retrieve an individual metadata record from a repository
- *Identify*, to retrieve information about a repository
- *ListIdentifiers*, an abbreviated form of ListRecords, retrieving only headers
- *ListMetadataFormats*, to retrieve the metadata formats available from a repository
- *ListRecords*, to harvest records from a repository
- *ListSets*, o retrieve the set structure of a repository

BlogForever deliverables which refer to the standard

- Description Of Work:
  OAI-PMH is mentioned as one of the essential required standards to achieve compatibility with a large set of libraries and other information services: "OAI-PMH is also an important standard which rapidly gains adoption in e-prints servers and digital repositories. BlogForever's digital repository must become part of the Open Archives Initiative and be capable of publishing OAI metadata in a variety of schemas".
- Deliverables D3.1 and D4.1
  Refer to Invenio requirements (IR):
  IR3 - Export data using OAI-PMH protocol and Dublin Core schema
  IR4 - Expose parts of the archive via OAI-PMH based on specified criteria
- Deliverable D4.4
  Multiple references in the standard:
  Section 2 confirms Invenio has this capability built in: "The development strategy used to implement Invenio ensures that it is flexible in every layer. Being based on open standards such as MARC and Open Archives Initiative metadata harvesting protocol (OAI-PMH), its interoperability with other digital libraries is guaranteed."
  Furthermore in 2.3.1: "The OAI Harvest [module] represents the OAI-PMH compatible harvester. It allows the repository to gather metadata from other OAI-compliant repositories and is also in charge of OAI-PMH repository management."
  The standard is also part of DIP assembly since repository features RF7 describes how to export data using the OAI-PMH protocol (p 34).
  OAI-PMH is also one of many data export options from Invenio (Section 4.1.4): "OAI-PMH: The Open Archives Initiative metadata harvesting protocol (OAI-PMH) can be used in Invenio to import and also export data."
- Deliverable D4.5
  Section 3.1 confirms the RF7 repository feature (p 15).

Relevance and applicability for the BlogForever platform

As a virtual global registry

"What gives the OAI-PMH process its power is the way that individual institutional repositories can each have their own particular collection policies and administrative systems, but to be linked into one large, a virtual, global repository through the use of the OAI-PMH. This allows individual institutions or subject communities to build their own individual repositories for their own purposes, but for users to be able to search just one service to gain access to all of the content of all of the repositories." (JISC Repositories Support project).

The above description would seem to match BlogForever's interoperability goals

Implementation

There are two classes of participants in the OAI-PMH framework: *Data Providers* administer systems that support the OAI-PMH as a means of exposing metadata; and *Service Providers* use metadata harvested via the OAI-PMH as a basis for building value-added services.

We can have confidence that such participants will be catered for in Invenio. For example, RF38 declares that "Users can communicate within the archive sharing and exchanging resources", and this is achieved using the components WebSession, WebMessage, and WebComment. This meets the requirement UI30, the Creation of a community of providers and recipients within the archive platform.

Partners

In terms of potential partners for interoperability, a number of software systems support the OAI-PMH, including Fedora, GNU EPrints from the University of Southampton, Open Journal Systems from the Public Knowledge Project, Desire2Learn, DSpace from MIT, HyperJournal from the University of Pisa, Primo, DigiTool, Rosetta and MetaLib from Ex Libris, DOOR from the eLab in Lugano, Switzerland, panFMP from the PANGAEA (data library), SimpleDL from Roaring Development, and jOAI. [30]

References

http://www.oaforum.org/tutorial/english/page5.htm

http://www.openarchives.org/pmh/

http://www.rsp.ac.uk/grow/registration/harvesting/

Assessing the Design of Web Interoperability Protocols (Jorgina Paihama, Kyle Williams, and Hussein Suleman, 2012)

4.5.2 OAI-ORE (Open Archives Initiative Object Reuse & Exchange)

Type of standard

- Standards for transmission and access

URLs

http://www.openarchives.org/ore/

Purpose of the standard

Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called

[30] http://en.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting

compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The intent of the effort is to develop standards that generalize across all web-based information including the increasing popular social networks of "Web 2.0".

Who uses the standard?

Digital libraries and digital repositories.

Further background information

The ORE Data Model builds on the following foundation technologies and architectures.

- The architecture of the World Wide Web
- Semantic Web concepts including RDF and the RDF Vocabulary Description Language (RDFS)
- Cool URIs[31] and Linked Data

In particular, OAI-ORE is based entirely on the architecture of the Web and encourages use of recent developments in the areas of the Semantic Web, Linked Data and Cool URIs. It is highly influenced, however, by the RDF model, which uses the idea of triples to describe things.

OAI-ORE focuses on objects and the relationships between these objects and introduces the concept of Aggregations and Aggregated Resources. An Aggregation is simply a set of Aggregated Resources, all of which are represented by URIs. Resource Maps represent the highest level of the OAI-ORE model. A Resource Map has a URI and is used to describe a single Aggregation.

The OAI-ORE protocol adds new functionality while it is a completely separate standard that 'neither extends nor replaces' OAI-PMH. The basic idea of OAI-PMH is "a mechanism for harvesting records containing metadata" from one repository for reuse elsewhere. However, as the uses of repositories and the types of content expand, more comprehensive methods for sharing content and more capability in terms of what is harvested and how it is reused are required. This requirement is what OAI-ORE is believed to solve.

BlogForever deliverables which refer to the standard

No mention made in Deliverables. The standard was shortly mentioned in the Description of Work regarding web content preservation.

Relevance and applicability for the BlogForever platform

This standard is predicated on the idea of managing data content, and on web resources as *objects*, which is not the trend of the BlogForever project to date – at least not for interoperability purposes, where the thinking is heading in the direction of metadata exchange rather than object exchange.

One advantage of digital objects over fixed physical objects is the flexibility of 'binding' them into publications or other useful aggregated intellectual entities while retaining the ability to reuse them independently in other contexts.

OAI-ORE may be useful as part of a digital preservation strategy for BlogForever. There is some evidence it has been used as an exchange mechanism for moving repository contents from one system to another, in the projects ChemistryFM, ADM-OER and HumBox.

[31] http://www.w3.org/TR/2008/NOTE-cooluris-20080331/

References

http://journal.code4lib.org/articles/1062

http://www.ariadne.ac.uk/issue57/rumsey-osteen#What_Does_OAI-ORE_Mean_for_Digital_Preservation

http://journal.code4lib.org/articles/1062

4.5.3 Z39.50

Type of standard

- Standard for transmission and access
- Interchange protocol

URLs

http://www.loc.gov/z3950/agency/

Purpose of the standard

The Z39.50 protocol is an application layer protocol. The objective of this standard is to facilitate communication between a client and a server for applications where clients search and retrieve information from server databases. It is a protocol that specifies data structures and interchange rules that allow a client machine to search a database provided by a server and retrieve records that are identified as a result of such a search.

This standard is intended for systems supporting information retrieval services for organizations such as information services, universities, libraries, and union catalogue centers. It addresses connection oriented, program - to - program communication. It does not specify a user interface.

Who uses the standard?

Z39.50 was originally a pre-Web ancestor of SRU-CQL, developed primarily for library and information related systems. It is mostly used for cross-searching bibliographic databases, although it has been extended to cover non-bibliographic media. Z39.50 is widely used in library environments and is often incorporated into integrated library systems and personal bibliographic reference software. Interlibrary catalogue searches for interlibrary loan are often implemented with Z39.50 queries.

The standard's maintenance agency is the Library of Congress.

Further background information

Z39.50 is stateful, connection-oriented and defines the interactions between two machines only. The recently developed applications that permit a client to search multiple servers in parallel are built on top of Z39.50 and use multiple concurrent Z39.50 connections to multiple machines. Z39.50 does not specify an applications program interface (API) to the services of the protocol but deals only with the interactions between the client and server machines. In addition, Z39.50 neither addresses possible issues involved in user interfaces of the client nor issues involved in database management at the server.

The basic architectural model that Z39.50 uses is as follows: A server houses one (or more) databases (collections) with records. A set of access points (indices) is associated with each database and can be used for searching.

This standard describes nine operation types: Init, Search, Present, Delete, Scan, Sort, Resource-report, Extended-services, and Duplicate Detection. A SEARCH request produces a set of records, called a "result set", that are maintained on the server. The result is a report of the number of records comprising the result set. The client using PRESENT requests can subsequently retrieve records from the result set. The PRESENT request offers elaborate options for controlling the contents and format of the records that are returned and indicates specifically which records from the result set are to be retrieved.

Z39.50 Profiles

The Z39.50 standard defines a range of services useful in information retrieval applications and for each of them it provides choices and options for parameters in individual protocol messages. Since there are many objects used in conjunction with the standard (e.g., attribute sets), the result is a comprehensive information retrieval protocol that offers flexibility to select services, parameters, and objects for specific applications. In general an implementation does not support the complete standard, but rather a conforming subset corresponding to specific relevant requirements. Consequently, interoperability between implementations is not always optimal.

To guide the use of the Z39.50 standard in applications and manage to improve interoperability, developers define profiles. Such profiles define a subset of specifications from one or more standards (e.g., selected services and required values for specific parameters) and associated objects to be used in specific applications. The objective of profiles is to improve interoperability between systems that conform to a specific profile. Thus, the implementation in that case means to configure a Z39.50 client and/or Z39.50 server to conform to one or more profiles.

There are several motivations for creating profiles. They can be built, for example, to solve interoperability problems with existing Z39.50 implementations within a specific community (e.g., libraries) or across two or more communities (e.g., library and museums). Thus, profiles can be characterized, as a response to community needs; they provide a solution path towards improved interoperability in specific applications and domains.

Furthermore, when there is a completed profile, customers can use it to aid in purchasing decisions. For example, a library can reference a profile in a Request for Proposal. Thus, a profile provides the details necessary for developers and vendors to build and configure Z39.50 clients and servers.

The Z39.50 Maintenance Agency monitors profile development in response to application and community needs and maintains a list of the profiles[32].

A widely known profile in the library domain is the *Bath Profile*. The Bath Profile was an attempt to remedy problematic situation because of the abstract syntax of Z39.50. For example, if the client specifies an author search, it is up to the server to determine how to map that search to the indexes that it has. On the one hand, this allows Z39.50 queries to be formulated without required knowledge about the target database. But, on the other hand, that means that results for the same query can vary widely among different servers; one server may have an author index, another may use its index of personal names, whether they are authors or not; another may have no name index and fall back on its keyword index; and another may have no suitable index and return an error.

Currently

As aforementioned, Z39.50 is a pre-Web standard and there were several attempts to update it to fit better in the current environment. Therefore, the protocol SRU[33] has superseded it, replacing the communication protocol with HTTP, removing much of the complexity but preserving the benefits of the query syntax.

[32] http://www.loc.gov/z3950/agency/profiles/

[33] http://www.loc.gov/standards/sru/

BlogForever deliverables which refer to the standard

- Description of Work, Task 3.2
  Z39.50 was mentioned as one of the essential required standards to achieve compatibility with a large set of libraries and other information services: "Z39.50 support is essential due to widespread use of the protocol library environments and integrated library systems".

Relevance and applicability for the BlogForever platform

After detailed research in the current state of the art and requirements analysis the Consortium decided to implement SRU instead.

References

http://standards.jisc.ac.uk/

http://www.dlib.org/dlib/april97/04lynch.html

http://en.wikipedia.org/wiki/Z39.50

http://www.oclc.org/research/activities/srw.html

4.5.4 SRU (Search / Retrieve via URL)

Type of standard

- Standard for search and retrieval

URLs

http://www.loc.gov/standards/sru/

Purpose of the standard

The Search / Retrieve via URL (SRU) protocol is a search and retrieval protocol that uses Internet and web facilities to carry the messages between user and target. The SRU protocol was developed as a way of increasing the level of compatibility between library systems and other information sources, adding compatibility with current web standards. The aim was an easier integration of information sources between libraries and digital information sources available on the Internet.

Who uses the standard?

SRU as a superseder of Z39.50 is widely used in the library community. For example, the European Library uses SRU as a search gateway to 47 European national libraries from a single interface, providing metasearch functionality across their resources. However, SRU is also used in several venues beyond accessing library catalogues.

Further background information

SRU was developed by the Library of Congress. The development was strongly based on lessons learned in the use of previous protocols, and particularly Z39.50, and the intention to address the occurred issues. Much of the functionality of SRU is derived from the older protocol, however, only the most useful was brought over, and in a simplified form.

The primary goal was the use of standard Internet protocols and communication formats for information interchange. Thus, possible obstacles to implementation by information providers outside the traditional library community could be overcome. The World Wide Web communicates

using hypertext transport protocol (HTTP) and HTTP Secure (HTTPS) and, therefore, the adoption of these protocols for communication eliminates the need for implementation of specialized protocols.

Meanwhile, Extensible Markup Language (XML) had quickly evolved into a widely used information interchange format, and the SRU developers adopted it as the basis for information exchange.

So SRU is XML-based and very flexible and the most common implementation is SRU via URL, which uses the HTTP GET for message transfer. Other versions, however, can be run over the web's SOAP protocol (SRU via SOAP), which supports more web service features, and over HTTP POST (SRU via POST), which avoids some length and character set restrictions that are currently present with HTTP GET. The records returned in response to a search can be in a well-defined XML format.

Since the changes in the information retrieval protocol were basically designed to allow greater integration on the web, the initial name of the protocol was Search /Retrieve Web Service with the initialization SRW. Eventually the terms SRW and SRU were used to distinguish the methods available for web based communication: SRW communication uses SOAP-based access, while SRU uses the Representational State Transfer (REST) approach. But the actual protocol operation is the same regardless of the communication method, and the current version of the protocol uses SRU to refer to both methods. However, the literature continues to contain references to both SRW and SRU.

There are three basic operations in SRU

- *explain*, which provides an XML description of the functionality of the service, including supported access points, record sets and features,
- *searchRetrieve*, which performs searches and retrieves records (similar to standard keyword searches and record requests),
- *scan*, which provides a list of available terms in an index (similar to browse lists).

The first version of the protocol was released in 2002 and the current SRU version is 1.2.

A key component of the Search/Retrieve operation is the query. SRU creators developed a query syntax that is both rich and simple−and well suited to getting the most out of library metadata. That query language is the Contextual Query Language[34], or CQL as it is usually called.

BlogForever deliverables which refer to the standard

- Deliverable D.4.4
  "External machines are able to query the repository using the standardized querying syntax of SRU and retrieve metadata in MARC or DC formats."
  Also, it is included in repository features RF45 and RF59 descriptions.

Relevance and applicability for the BlogForever platform

SRU protocol has been implemented in Invenio.

References

http://capping.slis.ualberta.ca/cap10/MichaelSilver/interop_and_sru.pdf

http://archive.ifla.org/IV/ifla72/papers/102-McCallum-en.pdf

[34] http://www.loc.gov/standards/sru/specs/cql.html

4.6 Web-archiving standards

These standards are used for creating web-accessible content in an archived state, representing the final form of a capture which can be disseminated over the web to a user agent (web browser). As such the standards could be described as standards for file formats, or more accurately "wrapper" formats for an aggregation of archived content. These standards were developed specifically to meet the requirements of the International Internet Preservation Consortium, a body that has since 2003 been developing standards that enable the creation of international web archives.

4.6.1 ARC / WARC

Type of standard

- File format
- Aggregate archive file

URLs

http://archive-access.sourceforge.net/warc/

http://bibnum.bnf.fr/warc/WARC_ISO_28500_version1_latestdraft.pdf

http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml

Purpose of the standard

The WARC (Web ARChive) format specifies a method for combining multiple digital resources into an aggregate archival file together with related information. The WARC format is a revision of the Internet Archive's ARC File Format [ARC_IA] format that has traditionally been used to store "web crawls" as sequences of content blocks harvested from the World Wide Web. The WARC format generalizes the older format to better support the harvesting, access, and exchange needs of archiving organizations. Besides the primary content currently recorded, the revision accommodates related secondary content, such as assigned metadata, abbreviated duplicate detection events, and later-date transformations.

Who uses the standard?

The WARC file format was designed to support the requirements of members of the International Internet Preservation Consortium.

Heritrix is used by the Internet Archive, the UK Web Archive, the Library of Congress, Harvard University Library, Government of Canada Web Archive, Web Information Collection and Preservation – WICP (China), Netarkivet.dk, Finnish Web Archive, BnF - BnF Web Legal Deposit (France), Slovenian Web Archive, Portuguese Web Archive, Web Archive Switzerland, and many others.

It is likely therefore that the content of these institutions' web archives is stored in WARC, which is the default format for Heritrix.

Further background information

ARC was adopted by IIPC members as a storage and exchange format. An ARC file consists of a sequence of URL records and each of these starts with a header that contains metadata. The

metadata include information about the harvesting technical context coming from the HTTP protocol exchange between the crawler and the host, followed by the corresponding to harvested URL file.

WARC was an extension of arc as an effort of the consortium to accommodate larger information. It was intended to be introduced as a format for web archives to cover the needs of institutions for an international standard format that would provide trusted repository and exchange of data.

WARC is an open standard and has been accepted as an ISO standard in 2009 (ISO 28500:2009).

WARC files are used to store and preserve web archive content in an open way, facilitating best practices, system interoperability, and long-term web content preservation. Furthermore, there is a quite large library of open source tools to access and work with this file format

WARC is a container for the results of a crawl. It is data-format agnostic, it essentially just encapsulates all the character-encoded bytes that the crawler gets over HTTP - both file contents and server responses. It doesn't particularly care what the contents are - HTML/XML (valid or not), JPEG, GIF, - they are simply character-encoded.

Post-processing - parsing, file extraction, - needs to be done as a next-step to make any particular sense of it. Presumably this action is carried out by the Wayback Machine.

BlogForever deliverables which refer to the standard

No mention made yet.

Relevance and applicability for the BlogForever platform

BlogForever originally opted not to use WARC as a storage format, but considers that storage of crawled and parsed blog content in XML is the preferred approach.

The BlogForever project conducted an internal report on WARC in November 2012, mostly written by Karen Stepanyan and incorporating some consultation with UoL partners. This report was written with a particular aim in mind. The specific question at the time was whether WARC would be suitable for transferring data from the spider to the repository.

The report is attached as Appendix A. It indicated there are numerous technical challenges to recasting a BlogForever parsed-XML crawl as a WARC file. At the time, the report questioned whether it would be feasible to implement WARC for this specific purpose of data transfer.

We use the same 2012 report now to answer a different question: can, and should, WARC be implemented in BlogForever for interoperability purposes?

To use WARC as anything other than the receptacle for the outputs of a crawl would seem to be counter-productive. If one already has broken the data down into discrete files, then no purpose is served by putting them back into a WARC, better to use a standard file-encapsulation method - ZIP, Bagit etc. Likewise if data files have already been parsed for metadata, data structures, etc - that post-processed data belongs in an appropriate data container - XML, RDBMS, etc.

With regard to rendering the contents of a WARC, it can no doubt be done 'on-the-fly', since serialising a WARC file to STDOUT is sending results very like those that Web server would normally deliver, or that a Web client would normally receive.

In a BlogForever context, WARC would be a potentially useful container to store all the crawler's results for post-processing; and, since it will be the richest record we have, should also be preserved as-is for future requirements foreseen or unforeseen.

The BlogForever repository should be delivering post-processed content - either complete HTML files extracted from the WARC, or sub-elements/content-blocks (title, content, metadata) parsed out of the HTML.

The present report concludes that there is a potential lack of technical compatibility here, which affects the prospects of interoperability between BlogForever and members of the IIPC, or indeed any institution that is committed to using Heritrix, WARC and Wayback Machine to build their collections of archived websites. The issue is whether a BlogForever crawl could be reused by any such institution, and whether such institutions would be in a position to reciprocate by hosting / storing BlogForever crawls.

However, at time of writing, a potential way to address this situation is being considered by the Project manager, the Invenio team and CyberWatcher.

References

http://web.hanzoarchives.com/bid/30720/Open-Standards-are-Important-in-Commercial-Web-Archiving

Web archives long term access and interoperability: the International Internet Preservation consortium activity, Catherine Lupovici, 2005.

4.7 Other standards

4.7.1 Encoded Archival Description

Type

- Character encoding standard

URLs

http://www.loc.gov/ead/

Purpose of the standard

Encoded Archival Description (EAD) is an encoding standard for machine-readable finding aids such as inventories, registers, indexes, and other documents created by archives, libraries, museums, and manuscript repositories.

EAD is increasingly being used to enable archives to publish or share their archival records. EAD includes some elements for describing digitised versions of archival materials. Multimedia objects can be described in simple terms within an EAD record, but those using EAD may prefer to link to more detailed records described using another schema.

Who uses the standard?

EAD is used in archives, museums, and special collections. It is used, for example, within the UK's Archives Hub and the Online Archive of California (OAC).

Further background information

EAD uses the Standard Generalized Markup Language (SGML). An XML version has also been developed. The EAD 2002 Schema is available in two syntaxes: Relax NG Schema (RNG) and W3C Schema (XSD).

BlogForever deliverables which refer to the standard

- Deliverable D3.1
  Section 5.4.3 mentions EAD as a possible way of expressing rights metadata.

Relevance and applicability for the BlogForever platform

In one sense, EAD lends itself to describing websites, in that one of the strengths of EAS is the way it preserves the hierarchical relationships existing between levels of description. As such, EAD might offer a small opportunity for interoperability if the service were to be used by an institution which catalogues its web collections using EAD.

One potential partner for interoperability is The Archives Hub [35], which stores descriptions in EAD. It is a JISC-funded service based at Mimas, a National Data Centre supporting world-class learning and research. It brings together descriptions of archives for research and education, enabling users to search across over nearly 200 repositories. However, it should be noted that these EAD catalogues are at a very high level, and tend to be descriptions of entire collections of resources (not individual websites or blogs).

The capability for authoring EAD catalogues is not built into Invenio. One reason for this could be that BlogForever is oriented in the direction of library description (e.g. bibliographic metadata) than archival description. Another reason may be that the DTDs for EAD are not widely used or supported except for specialist collections.

References

http://www.jiscdigitalmedia.ac.uk/guide/metadata-standards-and-interoperability#

4.8 Conclusions

Standards play a key role when interoperability is considered from the technical perspective, since they represent the common language that facilitates the interoperation among two or more different environments. This section presents the outcome of an extensive literature survey on the technical standards that are adopted commonly in order to enable interoperability scenarios.

The focus on the survey was on the standards that are used from the specific communities of digital libraries, web archives and digital preservation. The development of most of these standards derives indeed from these communities.

For each of the standard, a brief and comprehensive description is presented about the purposes that are served, the functionality and the main implementers. Therefore this section provides a coherent and useful guide about the most commonly used standards that are met in interoperability scenarios in the aforementioned domains.

[35] http://archiveshub.ac.uk/

5 A simple approach to consider interoperability

In the following, we propose a simple approach to consider interoperability of the BlogForever system (or any other digital library) in a concrete usage context.

Interoperability describes how two (or more) systems operate together (Geraci, 1991). However, the term system is not limited to software systems, and, therefore, interoperability issues can arise on different levels (e.g. technical, semantic, and organisational) (Anon., 2004). Interoperability depends strongly on the purpose of the intended interoperation between two systems. It can only be considered and accomplished successfully if the purpose of the interoperation is sufficiently defined. Nevertheless, interoperability aspects can change over time, for example, if new technologies are established and, therefore, it is necessary to be maintained.

The following approach facilitates establishing and assessing interoperability. It consists of five steps that should be processed sequentially. An extensive documentation of the considerations and decisions made in each step can be used afterwards for revision and, thus, for interoperability maintenance. The section is structured into four parts: In the beginning, an overview about the five steps of the approach is given in section 5.1, including the description of each step. Afterwards, the limitations, caused through the simplicity of the approach, are revealed in section 5.2. The example in section 5.3 demonstrates the application of the approach in a fictive scenario in order to further illustrate its use. Finally, conclusions are drawn in section 5.4.

5.1 General overview: Five steps

The following section describes the five steps of the proposed approach. Ideally, the steps should be performed sequentially. However, ambiguous or missing information can make it necessary to return to previous steps.

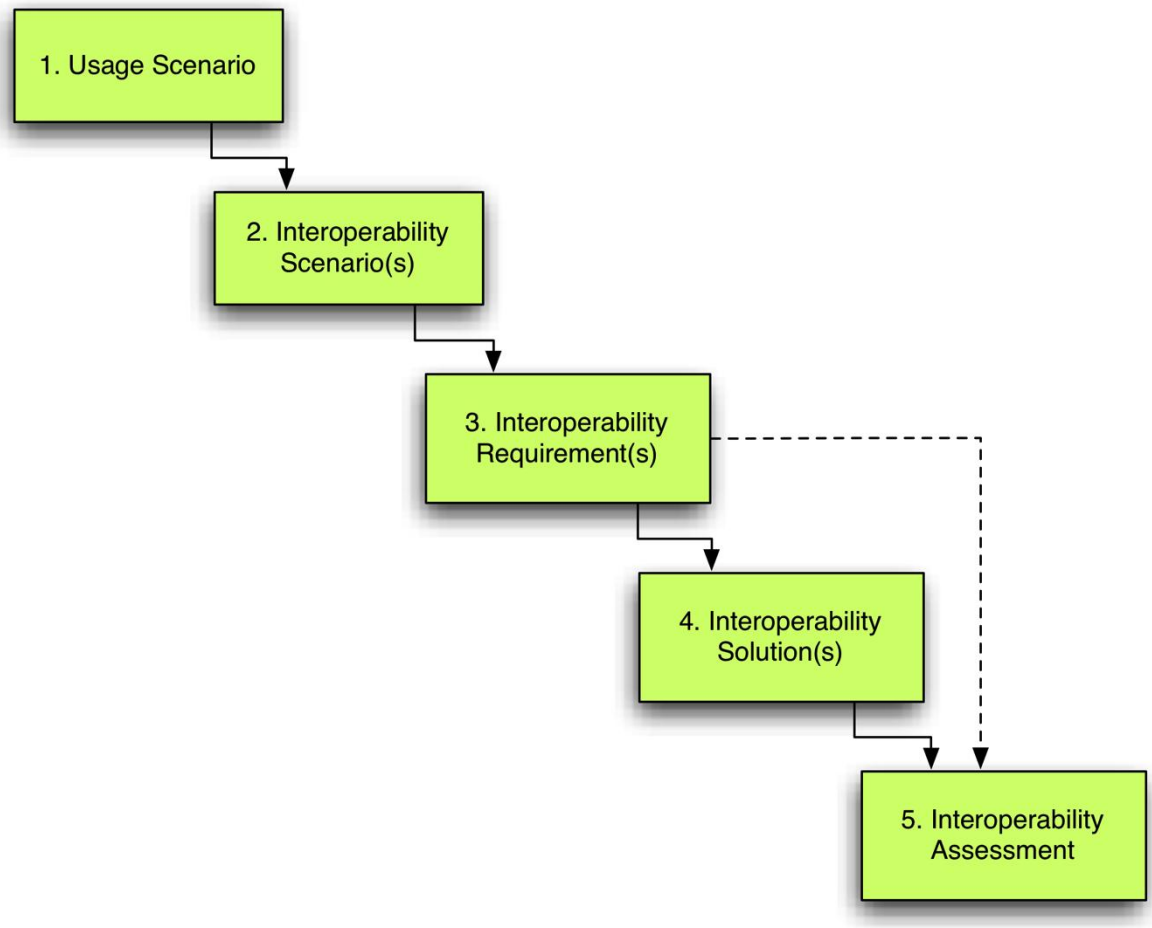

Figure 4: Five steps approach

Step 1 - Usage Scenario: To consider interoperability prospects of a system, one has to understand completely the system before. Therefore, the current (or intended) usage of the system has to be analysed and documented. The usage scenario describes the usage of the system in a specific context. Thereby, the description should cover the areas of organisation, content, user, functionality, policy, quality, and architecture.

Step 2 - Interoperability Scenario(s): The interoperability scenarios are derived from the usage scenario. An interoperability scenario emerges if the system has to interoperate with another system in order to provide the services described in the usage scenario. Thereby, a service itself can state also an interoperability scenario. The system can interoperate with various other systems for several reasons. Therefore, an overview about existing and intended interoperability scenarios should be created in order to facilitate the identification of synergies, and the prioritization of solutions in subsequent steps.

Step 3 - Interoperability Requirements: The interoperability scenarios lead to requirements that have to be fulfilled in order to establish interoperability. Thereby, the requirements are not limited to the capabilities of the system itself. Requirements can also address the usage context, and, thus, further specify the context in which interoperability can be. A complete fulfilment of the requirements enables the usage scenario from the perspective of interoperability.

Step 4 - Interoperability Solutions: Interoperability requirements that address the capabilities of the system create subsequently need for solutions. Several solutions may address a single requirement, and a single requirement may address again several interoperability scenarios. Thus, additional aspects (e.g. synergies) can be taken into consideration in order to make a decision about the solution.

Step 5 - Interoperability Assessment: The proposed solutions have to be evaluated with respect to requirements. If the requirements can be assessed as fulfilled, the interoperability is addressed sufficiently.

5.1.1 Usage scenario

The following section describes the recommended aspects to analyse and describe a usage scenario. Additionally, a template is provided that facilitates the documentation of a usage scenario.

A usage scenario describes the actual deployment of the BlogForever platform in a specific context. The description should cover the following aspects: organisation, content, user, functionalities, policies, quality, and architecture (Athanasopoulos et al., 2011).

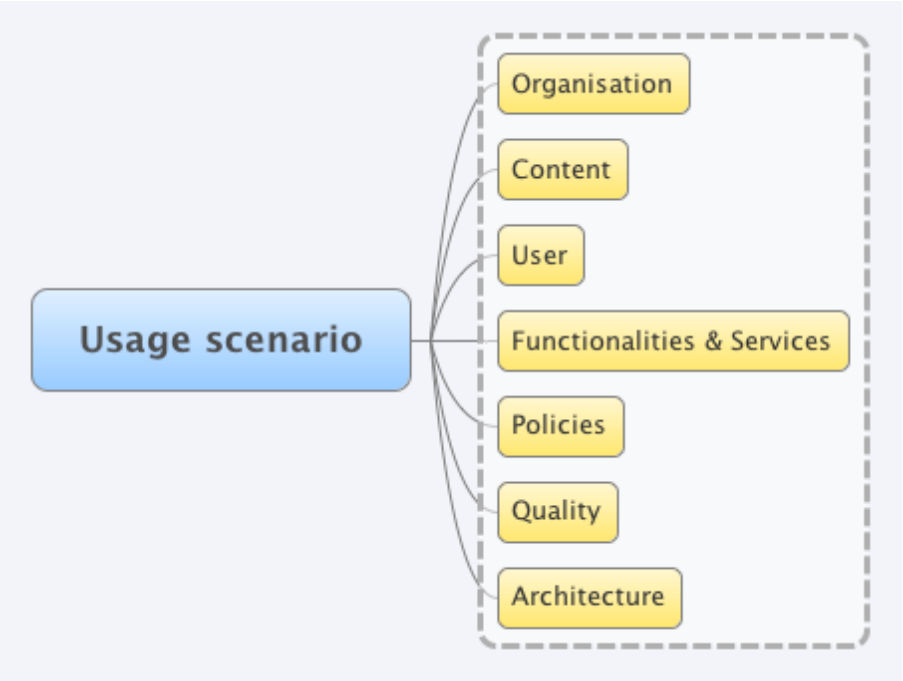

Figure 5: Usage scenario aspects

The architecture aspect describes the IT infrastructure of the system. It comprises of hardware and software components (Candela et al., 2011, pp.52-54). The architecture is an overview about the technical infrastructure. It constitutes the technical foundation and context of further considerations.

The organisation aspect comprises organisational decisions and impacts the other aspects. It defines the subordinate mission and goal of the system. The organisational decisions should be used in step 2 to reason considerations about the necessity of specific interoperability scenarios (Athanasopoulos et al., 2011).

The content aspect describes the available information objects. The information objects are not limited to primary objects that should be preserved but comprise also any kind of metadata and annotation (Candela et al., 2011, pp.39-41).

The user aspect describes roles and groups of users in the system. Single users or user groups can be differentiated if they have different tasks, rights, perspectives, etc. in the system. Typical user roles are administrator, manager, and end-user of a system (Candela et al., 2011, pp.42-43).

The functionality aspect describes services or functionalities that are provided by the system. Thereby, the services are not limited to the end-user (or customer) but comprise also services for internal users, administrators, and managers (Candela et al., 2011, pp.43-48).

The policy aspect describes conditions, rules, terms, and regulations that govern the operation of the system. Thereby, the consideration should not be limited to explicitly documented policies but should also reveal implicit rules that are often based on experience (Candela et al., 2011, pp.48-50).

The quality aspect characterises qualitative standards and requirements of the system. Thus, quality should be considered regarding every other quality-related aspect, e.g. content quality, quality levels of the provided functionalities, etc (Candela et al., 2011, pp.50-52). Given that quality is always considered regarding another aspect (e.g. regarding content), the quality aspect is integrated into the other dimensions in the following templates. Thereby, quality can be associated not only with each class of content or functionality but also with specific information objects or services. Some of these parameters are quantitative and objective in nature and can be measured automatically, whereas others are qualitative and subjective in nature and can only be measured through user evaluations (e.g., focus groups).

5.1.1.1 Templates

The following templates should facilitate the consideration and documentation of the usage scenario.

1/6 Architecture

Description	The Architecture concept refers to a Digital Library System and represents the mapping of the overall service offered by a Digital Library, and characterised by Content, User, Functionality, Policy and Quality, on to hardware and software components.
Software components	*[Software are the programs and applications that belong to the IT architecture of the digital library system.]*
Hardware components	*[Hardware components are the collection of physical elements that belong to the IT architecture of the digital library system.]*
Architecture quality	*[Describes the quality of the architecture, whether, for example, the system has a distributed architecture, and whether it is based on standards.]*
Others/ Comments	

2/6 Organisation

Explanatory description	The blog archive as can be considered as an organisation itself (not only software). It pursues the goal of providing a library service. The concept should not be confused with the organisation or institution that runs the blog archive even if there are overlaps and dependencies. The blog archive as an organisation should have a mission and a goal.
Mission of the blog archive	*[The mission is the long-term purpose or objective of the archive. For example: Preservation of the Greek Blogosphere for future generations.]*
Goal of the blog archive	*[The goal specifies aims that have to be achieved to fulfil the mission. For example: Harvesting, archiving, and preservation of Greek speaking blogs, and blogs under the top-domain "gr".]*

3/6 Content

Explanatory description	Content aggregates all forms of information objects that the blog archive collects, manages, preserves, and delivers. It encompasses a diverse range of information objects, including primary objects, annotations and metadata. In the following, the types of primary objects, metadata, and annotations should be documented.
Primary objects	*[Primary objects are the objects that are archived / preserved. For example: Blogs, Blogposts, Videofiles, Audiofiles, etc.]*
Metadata	*[For example descriptive metadata, administrative metadata.]*
Annotations/ User generated content	*[Additional content or data that are generated by the user of the archive.]*
Content quality	*[Describes the archived content. Examples for content quality are accuracy, completeness, timeliness, or granularity.]*
Others/ Comments	

4/6 User

Explanatory Description	User includes all notions related to the representation and management of actor entities within the blog archive. It encompasses such elements as the rights that actors have within the system and the profiles of the actors with characteristics that personalise the system's behaviour or represent these actors in collaborations. Example roles or groups are manager, end-user, administrator, or curator.
Groups	*[A group is a collection of users with a given set of permissions assigned to the group (and transitively, to the users).]*
Roles	*[A role is a collection of properties like rights and responsibilities, and a user effectively inherits those properties when he acts under that role.]*
User interface quality	*[Examples for user interface quality are accessibility and completeness.]*
Others/ Comments	

5/6 Functionalities & Services

Explanatory Description	Services are offered by the blog archive to its different users, whether they are individual users or user groups. Functionalities are more granular and can be aggregated to services. A good starting point to identify functionalities and services are the areas Access, Discovery, Manage, Configure, Acquire, Browse, Search, Visualize, and Collaborate.

Services	*[Services are offered by the blog archive to its different users, whether they are individual users or user groups.]*
Functionalities	*[Functionalities are more granular than services and can be aggregated to services.]*
Functionality quality	*[Describes the quality of the functionalities, and services. Examples for functionality quality are precision, recall, or scalability.]*
Others/ Comments	

6/6 Policies (Extrinsic/Intrinsic)

Explanatory Description	The Policy concept represents the set or sets of conditions, rules, terms and regulations governing every single aspect of the blog archive service including acceptable user behaviour, digital rights management, privacy and confidentiality, charges to users, and collection formation. Policies can be intrinsic (defined by organisation itself) or extrinsic (imposed by a superior rule, e.g. national law), implicit or explicit, prescriptive or descriptive, enforced or voluntary. Note: Conditions, rules, terms, and regulations can be overlapping concepts.
Conditions	*[Conditions are things that must be satisfied to enable specific behaviour, procedures, or transactions.]*
Terms	*[Terms are things we agree to do or not to do.]*
Rules	*[Rules restrict the possible behaviour or procedures.]*
Regulations	*[Administrative restrictions (or rules) that have the effect of a law, and are imposed by authorities.]*
Others/ Comments	

5.1.2 Interoperability scenarios

The usage scenario in the former step describes the specific deployment of the BlogForever system itself. Now, an interoperability scenario considers the interoperation with another system. The deployment of a system has only a single usage scenario but can aim on various interoperability scenarios. The interoperability scenarios comprise existing and intended interoperability.

In order to break down the complexity, the analysis of interoperability scenarios should follow the steps:

1. Identification of interoperability scenarios and associated components.
2. Evaluation of intersections based on the interoperability scenarios.

The steps are explained in the following subsections.

5.1.2.1 Identification of interoperability scenarios and associated components

Four main components can be distinguished for an interoperability scenario (Athanasopoulos et al., 2011):

- Provider: The system that provides a resource that is used by the consumer system.
- Consumer: The system that uses the resources that is provided by the provider system.
- Resource: A specific resource that is provided to the consumer system by the provider system.
- Task: The intended usage of the resource in the consumer system.

Each interoperability scenario has to be described with respect to the aforementioned four components. Therefore, the following template should be used. An identifier or a name should be given to each interoperability scenario, and each component. The identifier or name is required because it is used in further steps to visualize relationships, and, thus, to identify synergies and conflicts.

Descriptor/ID of the interoperability scenario

Consumer system	*[Identifier / name and description of a specific system or a class of systems.]*
Task / Aim	*[Identifier / name and description of the intended usage of the resource in the consumer system.]*
Resource(s)	*[Identifier / name and description of the resource(s) that has to be transmitted.]*
Provider system	*[Identifier / name and description of a specific system or a class of systems.]*
Comments	*[Additional comments or restrictions.]*

The BlogForever system represents either the provider system or the consumer system in such an interoperability scenario. The corresponding system can be any other socio-technical system. Furthermore, it can also be a class or type of identical systems in order to enable interoperability with every system that belongs to this class. In this case, characteristics and behaviour of the class should be defined clearly.

Next to the detailed descriptions of the different interoperability scenarios, a list of used components should be maintained. A mind map can be used that consists of the three main branches "systems", "tasks / aims", and "resources". Thereby, systems are not separated into provider and consumer systems because a single system can take both roles. Furthermore, tasks and resources can also be aggregated to classes.

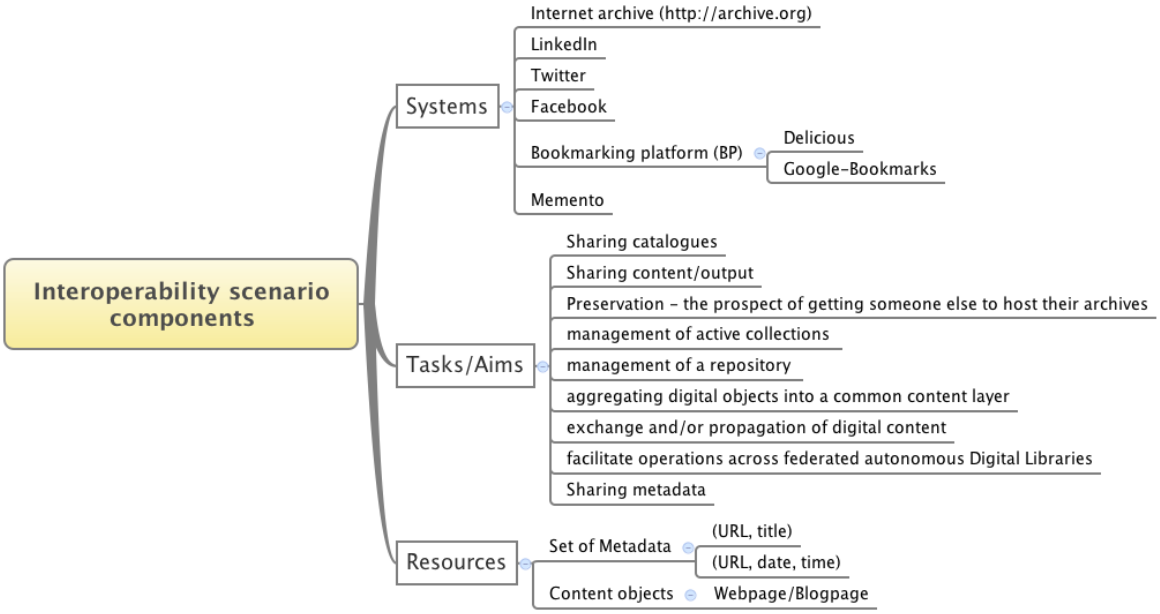

Figure 6: Simple example for a hierarchy of interoperability scenarios components

Figure 6 shows a simple example of a mind map that lists the components of interoperability scenarios. Note that the bookmarking systems of delicious and google are subsumed in the class of bookmarking platforms. Additionally, two different sets of metadata are subsumed in the class Set of Metadata.

5.1.2.2 Evaluation of intersections based on the interoperability scenarios

In case of several interoperability scenarios, tables of intersections can be created. Such maps provide a first impression about the diversity of the interoperability scenarios. Table 2 demonstrates a simple example for table of intersections. Each column represents one system or class of systems that the deployed BlogForever system should interoperate with. Each row represents a resource or class of resources that appears in an interoperability scenario. The intersection point of a resource and a system indicates if at least one interoperability scenario exists that contains both. It is additionally indicated with a letter and a colour if the system acts as a provider (P, red colour) or as a consumer (C, blue colour). A class of resources (e.g. Set of Metadata) aggregates the information of the contained resources. Similar tables can be created for the combination of resources and tasks, or tasks and systems.

Table 2: Resource intersections based on interoperability scenarios

Resources \ Systems	Internet archive (http://archive.org)	LinkedIn	Twitter	Facebook	Bookmarking plattform	Memento
Set of Metadata		C	C	C	C	C
(URL, title)		C	C	C	C	
(URL, date, time)						C
Webpage/Blogpage	P					

The tables of intersection allow a visual exploration of the diversity of interoperability scenarios. They can support planning and decision for further consideration of interoperability scenarios. However, it is not possible to apply a general recommendation or rule for decisions on these tables because they do not contain any information about the individual importance of the scenarios.

5.1.3 Requirements

Requirements have to be identified for each interoperability scenario that should be established. Thereby, the requirements should be analysed for each scenario separately. The description of an interoperability scenario (created in step 2) should be, therefore, extended with a list of associated requirements. Additionally, a detailed description has to be created for each requirement.

Requirements can be identified in different categories. The use of categories facilitates the requirements identification process because it counteracts the tendency of focusing on a specific kind of requirements (e.g. just technical perspective). This can be further facilitated if experts from different disciplines or with different backgrounds are included in the identification process. The following template uses three categories "organisational", "semantic", and "technical" that are derived from work of dl.org. However, other distinctions or additional categories (e.g. security, legal, etc.) are possible as well.

Descriptor/ID of the interoperability scenario

Consumer system	*[Identifier / name and description of a specific system or a class of systems.]*
Task / Aim	*[Identifier / name and description of the intended usage of the resource in the consumer system.]*
Resource(s)	*[Identifier / name and description of the resource(s) that has to be transmitted.]*

Provider system	*[Identifier / name and description of a specific system or a class of systems.]*
Comments	*[Additional comments or restrictions.]*
Organisational requirements	*[Organisational requirements address the processes, and policies that have to be adjusted to enable the scenario. For example, a special approval could be necessary that contents are without privacy concerns before they can be exchanged.]*
Semantic requirements	*[Semantic requirements address the vocabularies or terms that have to be adjusted to enable the scenario. For example, different systems may use different terms to indicate copyright and licence information.]*
Technical requirements	*[Technical requirements address the necessary hardware and software that has to be deployed to enable the scenario. For example, an Internet connection and communication protocols may be necessary to exchange data.]*

Each requirement needs an identifier and an explanation. The identifier is used in the description of an interoperability scenario to connect it with the associated requirements. The explanation of a requirement comprises (at minimum) an explanation of the requirement and a measure to assess the fulfilment of the requirement. More information, like a degree of necessity (e.g. essential, recommended, optional) or additional constraints, can be necessary or supportive with an increasing complexity of the requirements situation. The following template covers the minimum required information for a requirement description.

ID of the requirement

Description	*[Detailed description of the requirement.]*
Assessment / Measures	*[A measure that indicates the fulfilment of the requirement.]*
Comments	*[Additional comments or restrictions.]*
Required for	*[Interoperability scenario(s) that the requirement belongs to.]*

After the identification of the requirements for each interoperability scenario has been finished, the list of identified requirements should be consolidated. Thereby, duplicated requirements should be merged and the field "Required for" has to be updated. An overview table (see Table 3) of requirements and interoperability scenarios should be created in order to visualize the complexity and support the identification of similarities. The dark colour of a cell indicates that the particular interoperability scenario requires the corresponding requirement. Different colours could be used to indicate the degree of necessity if this information has been identified.

Table 3: Overview of requirements-scenario-intersections

Requirements \ Interoperability scenario	IS1	IS2	IS3
Req1		■	
Req2	■		
Req3	■	■	■
Req4		■	

5.1.4 Solutions

Each requirement has to be satisfied in order to establish the corresponding interoperability scenario(s). However, several solutions may address a single requirement, and in some cases a single solution may cover more than one requirement. Therefore, potential solutions should be considered regarding their impact on different interoperability scenarios.

Firstly, solutions have to be identified and described. Therefore, each requirement should be examined regarding potential solutions. Each solution should be documented with a description and the information about the requirements that aims to satisfy. Additional information like costs and risks of the solution can be supportive in the further decision process, and should be, therefore, added if available.

ID of the solution

Description	*[Detailed description of the solution.]*
Comments	*[Additional comments or restrictions.]*
Solves	*[Requirement(s) that the solution solves.]*

After the identification of solutions has been finished, the list of identified solutions should be consolidated in order to eliminate duplicates. Additionally, an overview table of requirements and solutions (see Table 4) should be created.

Table 4: Overview of solutions-requirements-intersection

	Requirements	Req1	Req2	Req3	Req4

Solutions				
S1				
S2				
S3				
S4				

The visual exploration of this table supports consideration of the impact of different solutions or combination of solutions. Examples of such impacts can be:

- Solutions are interchangeable if they cover the same requirements.
- A solution is interchangeable with a bunch of other solutions if they cover the same requirements.
- A solution is redundant if all the corresponding requirements are solved by another requirement.

Decisions about the implementation of solutions require more aspects of consideration (e.g. feasibility, political decisions, etc.) that cannot be included here because of their complexity. However, the documentation in the former steps allows to go back, and to update the considerations (e.g eliminating interoperability scenarios) with minimal effort for a repeated analysis.

5.1.5 Assessment

After the selection of solutions and their implementation, the overview table of the intersections between requirements and interoperability scenarios created in step 3 is reused to assess interoperability. Therefore, each requirement will be tested regarding the measure defined in the requirement description. If the test result was positive, the requirements row in the overview table (see Table 5) is updated in a way that former dark cells (which indicated for which interoperability scenario a requirement was necessary) are changed into green (to indicate the fulfilment of the requirement). Similarly, red colour is used if the test result was negative (to indicate that the requirement is not fulfilled). The column of an interoperability scenario shows whether it is established (only green and white cells) or not (red cells included). Additionally, the table shows the requirements that have not yet been tested (remained grey cells).

Table 5: Assessment of interoperability

Interoperability scenario / Requirements	IS1	IS2	IS3
Req1			
Req2			

Req3			
Req4			

5.2 Limitations

In order to reduce the complexity and provide a simply adopted methodology, the proposed approach has the following limitations/simplifications:

- Interdependencies among the requirements are ignored.
- Interdependencies among solutions are ignored.
- Potential benefits that may come up from specific solutions, which are not relevant to any current requirements or problems but could be beneficial in future (e.g. an agreement to a specific standard could prevent the appearance of future interoperability issues later), are not taken into account.

5.3 Scenario example

A fictional usage scenario of the Atlantis University Library (AUL) is presented and explained in this section[36]. The aim of this usage scenario is the application and the validation of the presented approach to interoperability. Moreover, the usage scenario will be utilized subsequently for the development of interoperability scenarios.

Organization

Mission of the Blog Archive	The mission of Atlantis University Library Blog Archive (AUL-BA) is to provide a cost-effective long term preservation repository for blog content in support of teaching and learning, scholarship and research in the Atlantis University.
Goal of the Blog Archive	In order to achieve its mission, the Blog Archive has to harvest, preserve, disseminate and reuse blog collections relevant to the topics of interest of Atlantis University Departments.
Operation of the Blog Archive	AUL-BA is operated by the librarians and administrators of the Atlantis University Library. All costs are covered by the budget of the library.

Content

Profile	Blogs relevant to the topics of interest of the Atlantis University Departments. These include: • Economics, • Business Administration, • Marketing, • Technology Management

[36] Another example is given in section 6.4.

	• Applied Informatics, • International and European Studies, • Education and Social Policy, • Balkan Slavic and Oriental Studies, • Music Science and Art. The relevant blogs are limited to blogs under the top level domain of Atlantis ".ay".
Primary objects	Primary objects in the AUL-BA are • Harvested blog pages stored as HTML, • Related text files (e.g. CSS files), • Related media objects (audio, video, etc.) as files. The blog objects are further structured according to the BlogForever Data Model[37].
Metadata	Descriptive, Provenance and Administrative Metadata: • Descriptive metadata are bibliographic metadata like title, author, creation date, etc. • Provenance metadata are digital preservation information, e.g. about the object's life cycle and history in the digital library, • Administrative metadata are technical metadata about content files and information about intellectual property rights. • Structural metadata describe how the components of an object are organised and provide links between content, e.g. relevant content, parent content, etc. • Rights Management metadata are including information regarding content rights.
Annotations / User generated content	Atlantis University Library users are encouraged to add comments to the Blog Archive, as the software platform provides this feature.
Others / comments	N/A

Users

Groups & Roles	Library personnel are distinguished in • Librarians are the content managers of the Blog Archive. They support all end-users. They have permission to administrate end users (create new users, change permission, delete users), curate archived objects and metadata, as well as

[37] See BlogForever deliverable D2.2: Weblog Data Model

	do any kind of information retrieval (search, browse, etc). • Administrators are responsible for maintaining the IT infrastructure. Thereby, they support the librarians. They have full access to the whole system are not responsible for the selection or management of the archived content. End-user are distinguished in • University Students (Bachelor & Graduate) are the end-users of the Blog Archive. They use it for learning purposes. They have the permission to search and access archived content. • Academics are the end users but also the content managers as they can suggest content to be included / excluded in the Blog Archive. They use it in support of teaching and learning. They have the permissions to propose additional blogs for archiving, and to search and access content. • Researchers (e.g. PhD candidates) are the end users but also the content managers and they can suggest content to be included / excluded in the Blog Archive. They use it in support of scholarship and research. They have the permissions to propose additional blogs for archiving, and to search, access, and export archived content. • 3rd party users are citizens with access to University Library Services. They are the end users of the Blog Archive, using it for learning purposes. They have the permission to search and access archived content.
Others / comments	N/A

Services and Functionalities

Services	The AUL-BA offers the following services to its users: • Maintain collections of blogs organized by topics, relevant to the University (Librarians). • Provide users with the ability to create new collections and/or alter the blogs included in existing collections (Academics, Researchers). • Provide users with the ability to use archived blog content for learning / teaching purposes (End-users). • Provide users with the ability to use archived blog content for research (Researchers).
Functionalities	In order to provide the services, the AUL-BA has the following functionalities: • Harvest sets of blogs, • Preserve harvested blogs, • Analyze and separate entities in blogs (e.g. posts, author, comment, date, etc.),

	• Full text search in blogs, blog posts, comments, and blog pages, • Web interface to view, search, and browse archived blog content, • Export functionalities to reuse archived blog content, • Export of collections of blogs, and blog posts including related comments and media objects, • Translation of archived content into various languages, • User comments for blogs, blog posts, and comments, • Creation of user collections of archived content, • Linkage to the original URL of blogs, and blog posts.
Others / comments	N/A

Policies (Extrinsic / Intrinsic)

Conditions	• Only the Library personnel curate collections. Academics and researchers can provide suggestions for collection management and adding / removing content. • Fair use of archived blog content is acceptable for all users.[38] • No charges are necessary to access the AUL-BA. • End-users have to be registered to access the AUL-BA. • The preservation of blogs is dependent from the implicit or explicit permission of the blog author.
Terms / Rules	• All Atlantis University personnel are obliged to have their blogs included in the Blog Archive. • Unrestricted access to Blog Archive APIs is available within the networks of Atlantis University only. 3rd party access is granted only after explicit licensing by the Atlantis Library Board of Directors. • A permission to preserve blogs is assumed unless the blog author objects explicitly. • In case of a blog author request, blogs are removed from the AUL-BA if the consistency of the archive is not at risk.
Regulations	• Berlin Declaration on Open Access to Scientific Knowledge[39] • The Atlantis Intellectual Property Law forbids the reuse of intellectual property (e.g. blog publications) for commercial reasons up to 20 years beginning with the death of the IPR

[38] http://www.ala.org/Template.cfm?Section=copyrightarticle&Template=/ContentManagement/ContentDisplay.cfm&ContentID=26700

[39] http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/

	owner (the blog author).
Others / comments	N/A

Architecture

Software Components	The BlogForever platform which consists of the two components: • The blog spider, and • The blog repository.
Hardware Components	The hardware infrastructure consists of three elements: 1. **The Blog Spider** server which is responsible for crawling all the necessary characteristics of the blogs designated for preservation. 2. **The Blog Repository** server which is responsible for storing the blogs permanently and enables further access and analysis. 3. **The Backup Storage** server which is responsible for keeping safe backup copies of the Archive.

5.3.1 Interoperability Scenario (IS) 1 – Federated Search

The following scenario aims on the participation of the AUL-BA in federated search of AUL.

Federated search[40] is an information retrieval technology that allows the simultaneous search of multiple searchable resources. Federated search is used in AUL because of the multitude of systems deployed. The library operates the following systems:

- Online public access catalogue (OPAC)
- Institutional Repository
- Electronic Journals
- Blog Archive (AUL-BA)

Using federated search, web users can search all these systems simultaneously from a single web interface. The process consists of the following steps:

(1) Transform a query and broadcast it to the participating systems,

(2) Merging the results collected from them,

(3) Presenting the results to the user in a uniform way.

(4) The user selects a desired result and is redirected to the relevant system.

The following interoperability scenario defines the requirements for the participation of the AUL-BA in the federated search of AUL. The elements of the interoperability scenario are:

Consumer system	Altantis University Library Federated Search Engine

[40] http://en.wikipedia.org/wiki/Federated_search

Task / Aim	Search Blog Archive to find relevant results to query
Resource(s)	Blog Archive Search Results
Provider system	Atlantis University Library Blog Archive
Comments	N/A
Organisational requirements	OR1 – Legal policy for external search access OR2 – Responsible person/role for external search OR3 – Service quality of the external search
Semantic requirements	SR1 - Shared set of common conceptualization (e.g. title, author, date) SR2 - Searchable entities and URI defined SR3 - Defined public / private collections
Technical requirements	TR1 - Search protocol / API to search the AUL-BA TR2 - Defined export format for search results

5.3.1.1 Organisational requirements

OR1 – Legal policy for external search access

Description	It has to be defined who is allowed to search which parts (e.g. restrictions regarding the time period) of the AUL-BA. This is necessary to avoid unintended or forbidden access.
Assessment / Measures	A documented polices states clearly who is allowed to search what in the AUL-BA.
Comments	N/A
Required for	IS 1 - Federated search

OR2 – Responsible person/role for external search

Description	It has to be defined who is responsible for matters of external search in the AUL-BA. The responsibility should include • to promote and to answer questions about the possibility of external search access,

	• to react in case of unintended shutdown, • to maintain the service or initiate maintenance activities.
Assessment / Measures	It is clearly stated which role or person is responsible for the external search functionality.
Comments	N/A
Required for	IS 1 - Federated search

OR3 – Service quality of the external search

Description	It has to be defined what service quality level the external search functionality has to fulfil.
Assessment / Measures	The service quality is defined clearly.
Comments	N/A
Required for	IS 1 - Federated search

5.3.1.2 Semantic requirements

SR1 – Shared set of common conceptualization (e.g. title, author, date)

Description	Well-defined and commonly agreed conceptualisations are necessary for the communication between the consumer system and the AUL-BA. A set of concepts / terms that can be used to query the AUL-BA has to be defined explicitly.
Assessment / Measures	Terms and concepts that can be used for search queries on the AUL-BA (e.g. title, author, date, period) are defined explicitly.
Comments	Definitions of terms and concepts should be either reused from existing standards or defined using a standard description language.
Required for	IS 1 - Federated search

SR2 – Searchable entities and URI defined

Description	It has to be defined what entities/objects (e.g. blogs, posts, comments, pages) will be delivered based on a search query. Each entity must be addressable with an URI in order to enable further requests, access, or usage on the entities returned to search query.

Assessment / Measures	A set of entities returned by search queries is defined, and each returned entity has an URI.
Comments	N/A
Required for	IS 1 - Federated search

SR3 – Defined public / private collections

Description	The collections hosted in the blog archive must be set as public (available for federated search) and private (dark archive). Different permissions should be defined for each collection and/or item in the archive.
Assessment / Measures	Blog archive collections and items have clearly defined permissions.
Comments	N/A
Required for	IS 1 - Federated search

5.3.1.3 Technical Requirements

TR1 - Search protocol / API to search the AUL-BA

Description	A protocol or API has to be defined and implemented for communication between the consumer system and the AUL-BA. The protocol or API has to provide a way for the consumer system to send queries to AUL-BA and get search results in return.
Assessment / Measures	A protocol or API is defined and implemented in AUL-BA that covers the possibility of search queries on the AUL-BA
Comments	N/A
Required for	IS 1 - Federated search

TR2 - Defined export format for search results

Description	The format of the search results has to be defined explicitly (e.g. encoding, structure) in order to enable the consumer system to process and present the search results correctly.
Assessment / Measures	The format of results returned on search results is comprehensively and explicitly defined.
Comments	N/A

Required for	IS 1 - Federated search

5.3.2 Solutions

ID	S1 - OpenSearch
Description	One of the most common ways to support federated search is via the OpenSearch protocol http://www.opensearch.org/ which can be used to share search results between different systems.
Comments	The correct implementation of Open search can be tested by validating the OpenSearch endpoint XML description and the results xml of the AUL-BA in order to check if everything complies with the XML Schema of the OpenSearch protocol.
Solves	TR1 - Search protocol / API to search the AUL-BA SR1 – Shared set of common conceptualization (e.g. title, author, date)

ID	S2 - SRU
Description	**SRU** is a standard XML-focused search protocol for Internet search queries, utilizing CQL (Contextual Query Language), a standard syntax for representing queries. http://www.loc.gov/standards/sru/
Comments	The correct implementation of SRU can be tested by validating the SRU endpoint XML description and the results xml of the AUL-BA in order to check if everything complies with the SRU protocol standards.
Solves	TR1 - Search protocol / API to search the AUL-BA SR1 – Shared set of common conceptualization (e.g. title, author, date)

ID	S3 - RSS
Description	RSS is an established web standard to share website updates in a standardized format.
Comments	Any Blog Archive search result page automatically has an RSS feed that can be used by the federated search engine to regularly check for new search results matching the search query.
Solves	TR1 - Search protocol / API to search the AUL-BA SR1 – Shared set of common conceptualization (e.g. title, author, date)

5.4 Conclusion

In this section, a simple approach to consider interoperability of the BlogForever system (or any other digital library) has been proposed. It has the meaning of a guideline for responsible managers in a concrete usage context. Therefore, it consists of five general steps that should be conducted to (a) analyse the general usage scenario, (b) describe the desired interoperability scenarios, (c) deduce the necessary requirements, (d) identify the possible solutions, and (e) measure the interoperability fulfilment. Several templates have been proposed to facilitate the structuring and documentation in the different steps. Additionally, recommendations have been made on how to identify synergies between the interoperability scenarios based on the related requirements and possible solutions. The description of the approach is accompanied by a fictive example to illustrate its application in the context of BlogForever. Another example is also given in section 6.4.

The strength in the approach lies in its simplicity. It gives the responsible managers a flexible guideline that can be fast be understood and applied. However, the simplicity limits the potential outcome. Therefore, it is recommended to extend the approach with further templates, aiming on a flexible set of analysing tools for interoperability in digital libraries. Additionally, the documentation in a real scenario will probably become extensive, and should therefore be facilitated by IT support.

6 BlogForever succession plan

The preservation strategy for digital materials should, ideally, include a succession plan specifying what would happen to the materials should the organisation responsible for them become no longer able to maintain the roles necessary for their preservation. The BlogForever Description of Work proposes the possible delivery of a succession plan. This is expressed in the following:

> *"Means for reliable transfer of content from BLOGFOREVER to other digital repositories are to be evaluated at a technical level and liaise with relevant research efforts in order to establish the successful undertaking of the succession plan, if need be one."* [41]

Since there is no existing BlogForever Archive[42] at the present time, it is difficult to discuss concrete low level details concerning the transfer of materials when both source and target archives are not specifically defined. On an abstract technical level, solutions have already been proposed for content level interoperability to support safe transfer or exchange of material through the use of accepted description standards (for example, see those in Sections 4.3, 4.4, 4.6), exchange protocols (e.g. see Section 4.5), and the translation of information packages (e.g. TIPR[43]).

Even before we need to worry about the exchange of information packages, however, a succession plan is dependent on strategies supporting organisational interoperability: this organisational level planning is seldom discussed in the literature in detail. The purpose of the discussion here is to fill this gap. Here the proposal is to align the investigation of succession plans with a set of recommended steps that could be implemented to promote organisational level interoperability. This is intended to inform future BlogForever Archives in developing and reviewing their succession plan in preparation for transferring the stewardship of a collection to another organisation when their archive is no longer able to continue. We propose that the details of safe transfer can only be developed as part of a longer term consideration of information value, skills and roles in relation to several potential successor organisations.

We adopt the business model as a starting point: more specifically, we review the guidelines for succession planning in the business context (Section 6.1) and map it to the repository context (Section 6.2). The approach is not specific to blog collections that adopt the preservation infrastructure of the BlogForever project, rather, it proposes a general strategy for succession planning that is transferable to other digital repository contexts where digital materials are collected for well defined purposes.

6.1 Succession plan in the business context

The objective of a business *succession plan* or *exit plan* is, fundamentally, to identify who will take over the key roles in a business (if anybody) when those who have been carrying out the roles cannot continue[44]. The discontinuation of the business under current structure and/or management might take place for a number of reasons including:

- bankruptcy and insolvency
- de-registering, sale, or winding up of a solvent business

[41] Page 10, Part A, WT3: Work package description, Description of Work, BlogForever Project – Grant no. 269963.

[42] In this report, we will use the term blog repository to mean any repository containing blogs. Blog collection on the other hand will refer to a repository of blogs developed with a set of repository objectives in mind with respect to blogs. A BlogForever Archive is a blog collection developed using the BlogForever Software/Services and the recommended BlogForever policies as a guideline.

[43] For example, http://www.ijdc.net/index.php/ijdc/article/view/145

[44] http://www.business.gov.au/Information/Succession-plan-template-and-guide/

- retirement, departure, or death of owner and/or key staff

Depending on the circumstances, the business may be faced with closing down or continuing under new ownership and/or management. A smooth transition is only possible if an effective succession plan is already in place before the need for succession arises (Rothwell, 2010). This involves making explicit the current business details and succession target, for example,

- business name and/or registration number;
- business structure (e.g. sole trader, partnership, trust, or company);
- the scope of the succession plan (e.g. agents, roles and positions covered);
- type of succession planned (for example, is it a partial or a complete);
- details of the successor(s);
- the timeframe of the plan implementation; and,
- any restrictions regarding the succession or successor (e.g. legal considerations).

The succession plan would, ideally, also, make explicit the proposed organisational structure to be adopted after succession, a list of positions/roles that need to be filled, target skills to be maintained and developed in association with each role, necessary training for successors, changes that have to take place on any register (i.e. names, identifiers, structures), legal considerations that need to be addressed (contracts, terms, agreements, conditions, wills and testaments), insurance, timetable for detailed succession stages, and assessment and management of succession risk. A succession plan must come with related financial information, such as current market value, sales conditions, required payout, detail of shares, transfer or sale taxes. The succession plan is usually accompanied by supporting documents as evidence of the above details. For example,

- Legal documentation of business details (e.g. official documents detailing business name(s), business type(s), business and/or company registration number(s), goods and services tax registration, insurance policies, leases)
- Business finances (e.g. market evaluation of the business; retirement payout; sale details; buyout details; tax records)
- Legal obligations (e.g. partnership agreement; buy-sell agreement; succession agreement; will and testaments; contracts; payout agreement; licenses)
- Resumes of potential successors

In the current report we are proposing a succession plan for a blog repository that becomes unable to continue. This could lead to closing down the collection or finding one or more organisations who will take responsibility for the collection, and might involve arranging reliable transfer of the holdings. In the next sections, we demonstrate how the steps in the business context might be mapped to the repository context.

6.2 Succession plan in the context of a blog collection

The proposal for mapping the business succession plan (Section 6.1) to the repository context is presented in Figure 7. The left hand side of the diagram represents requirements for succession planning that arise in the business domain. This is translated into the corresponding requirements on the right hand side that might be considered in the case of a given blog repository.

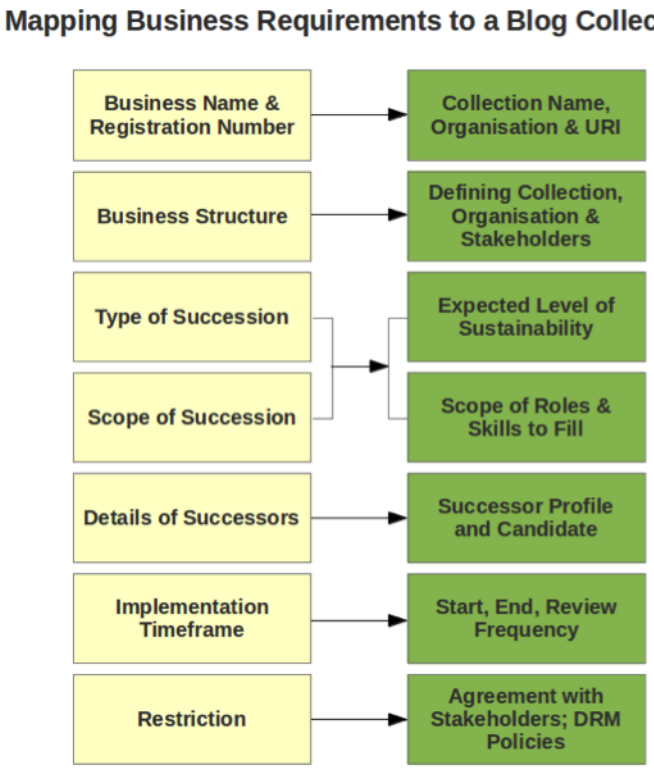

Figure 7: Facets of a business succession plan (left hand side) associated with facets of a blog collection (right hand side)

6.2.1 Business details versus repository details

Just as any legal names and registration numbers of a business need to be specified in a business succession plan, the name and associated information that identifies the collection of blogs must be specified in a blog collection succession plan. The blog collection is identified by a name of its own and/or by the organisation that is responsible for its maintenance. Even if the collection is an aggregated collection shared across many organisations, it is expected that the collection would have an identifiable collection name. Likewise, a BlogForever blog collection of the future is expected to be accessible online, and, therefore, will be expected to have a URI (or equivalent) which would serve the same function as a business registration number. After succession, these details need to be traced, updated and documented. This is essential for successful transfer of repository services delivered to the stakeholders but also for tracking provenance (see Figure 8), an element to be considered in the preservation of digital information. The information can be documented in a number of ways. To increase ease of succession, it is recommended that such information be exposed using a widely accepted method for sharing data and resources that supports machine accessibility, such as OAI-ORE[45] resource map in RDF[46], and Linked Open Data[47].

[45] http://www.openarchives.org/ore/

[46] http://www.w3.org/RDF/

[47] http://linkeddata.org/

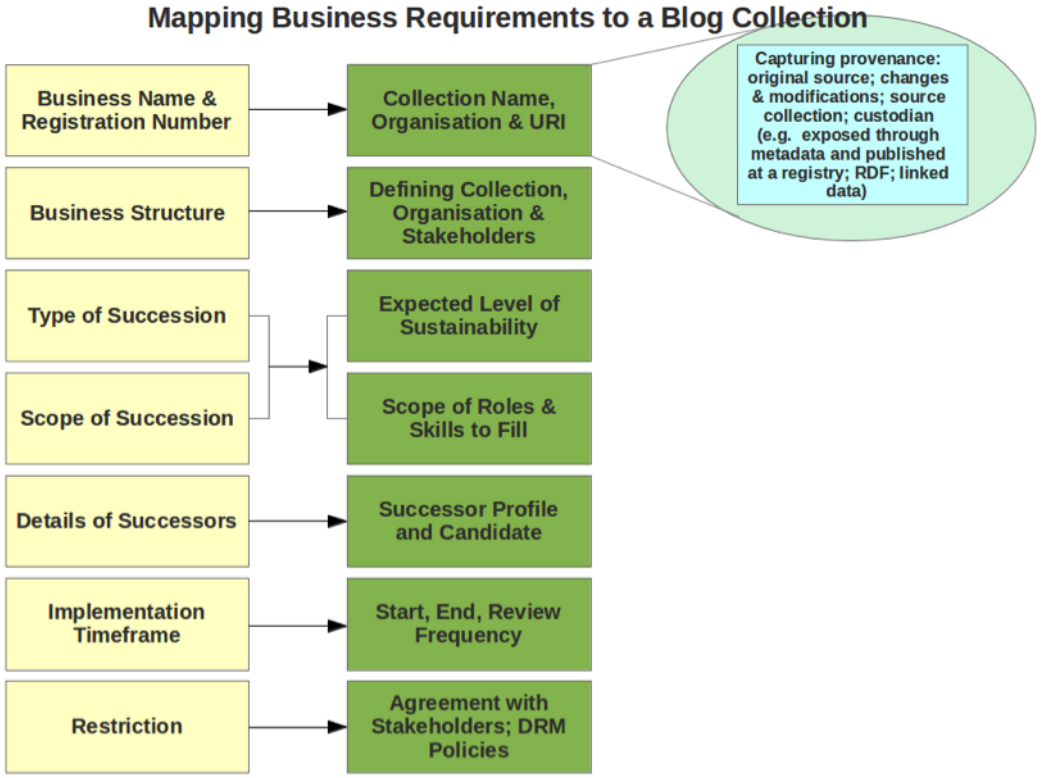

Figure 8: Capturing and exposing provenance.

6.2.2 Defining business structure versus describing the collection, organisations and stakeholders

In a business situation, the succession plan (Section 6.1) specifies the structure of the business (e.g. sole trader, partnership, trust, or company). This should be clearly specified also in a BlogForever repository to represent the body of stakeholders of the repository. Each known stakeholder (and/or designated community) of the repository must be described (see Figure 9) in terms of the sector to which they belong, the value of the repository holdings to that stakeholder, and, perhaps also, the risks to the stakeholder that would result from the loss of the repository holdings.

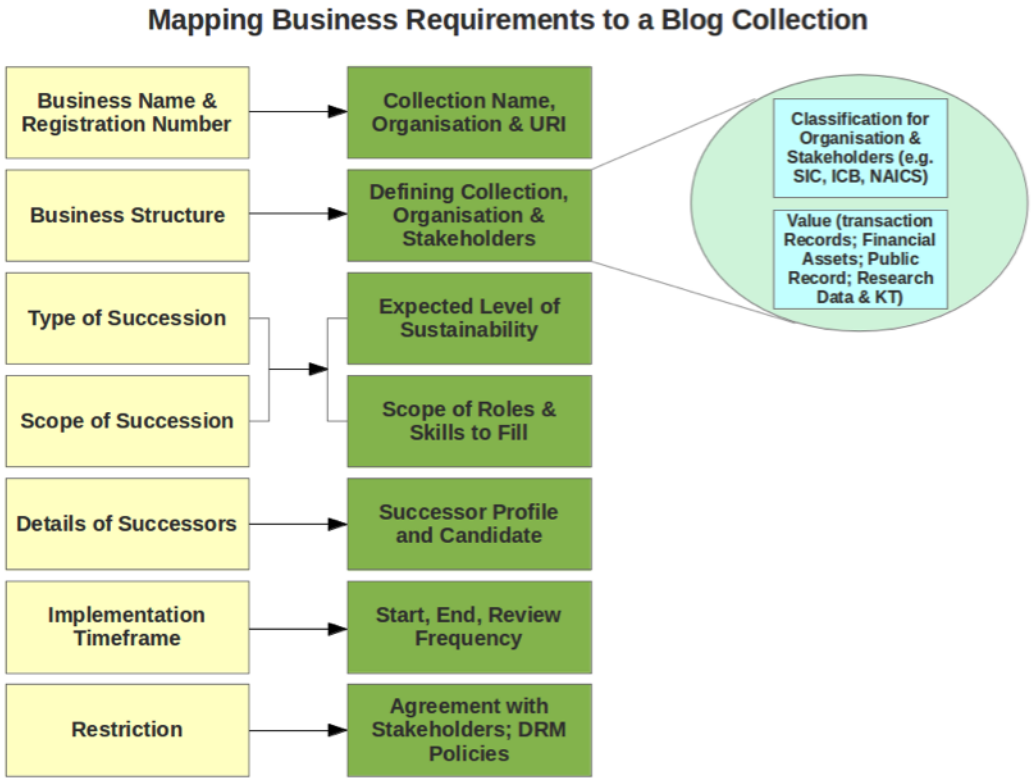

Figure 9: Describing the stakeholders of the repository.

It is recommended that sector description follows a well known standard such as SIC[48] or NAICS[49], and the value of the holdings be identified through a periodic review of repository usage within each stakeholder sector (this may be established through a combination of surveys, questionnaires and interviews, as well as, other types of market analyses). A market value analysis is a component in the business succession plan and it is also recommended within the repository context. There are several advantages of maintaining a sustained analysis of the value of repository holdings:

- it allows the diagnosis weaknesses and threats with respect to the organisation, making it easier to take necessary steps to prevent disasters or plan for succession;
- it makes the immediate financial obligations (e.g. liabilities, debts, taxes) and expected revenues clear (e.g. see legal considerations, Section 6.2.6);
- it exposes the network of stakeholders with whom the repository has relationships and agreements, and, to whom the repository has responsibilities (also relevant to Section 6.2.6);
- it increases the probability that a successor might be found, by providing a profile to which potential successors can be matched, and, by highlighting relevant sectors and the value of the collection to generate common purposes (see Section 1.2.5);
- to support the case for the proposed level of expectations that might need to persist in the new repository (e.g. see Section 6.2.3).

Market value analysis can be conducted from several perspectives: desk analysis of blog usage with respect to stakeholders and target candidate successor organisations; surveys conducted at organised training and network events. A description of an example survey has been included in Section 6.3 for reference.

[48] http://www.companieshouse.gov.uk/infoAndGuide/sic/sic2007.shtml

[49] http://www.census.gov/eos/www/naics/

6.2.3 Type of succession versus levels of expectations

In a business succession scenario, the type of succession to be carried out is addressed. The aim is to define whether the succession will be partial or full with respect to the successor's responsibility in relation to the business. Likewise, there can be five different approaches to succession dependent on levels of expectations[50] (see pale green boxes of Figure 10):

1. **No succession, the resource is abandoned:** there are steps to take even within this situation, such as notification to stakeholders and dissolving of responsibilities.
2. **Continued maintenance:** the resource continues to be looked after and maintained, but it does not develop in terms of new content or additional features.
3. **Continued development:** the resource continues to grow and/or develop. This will involve building on the original collection and/pr services. Either way this will require significant further funding.
4. **Integration:** the resource is integrated into a third party's existing collection. This may not involve further funding, but it will require forward planning in terms of interoperability and use of standards.
5. **Re-purposing:** the material is re-used in another resource. Depending on the type of media in the collection, re-purposing might involve creating variations based on file type, file size, audio quality, clip duration, image resolution or colour mode for use elsewhere; or it could mean reworking the metadata to meet the needs of a new audience.

Depending on which of these approaches are being adopted, the level of interoperability expected in the transfer of material between the two archives will differ. The first and second approach places no expectations on the integration of the resources into another repository and also does not place any expectations on further development of the collection or associated services beyond continued access to existing resources. The last approach, in contrast, places not only expectation on the integration of resources but also the integration of services across the collections of both the source and target organisations. Regardless of which of these paths is chosen, the integration could occur at the level of archival information packages (AIP), or, at the level of hosting location, application programming interface (API) and exchange protocols (see grey boxes in Figure 10). These levels of interoperability have been substantially investigated in previous sections.

[50] Borrowing concepts from http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/sustainability-of-digital-collections#st2

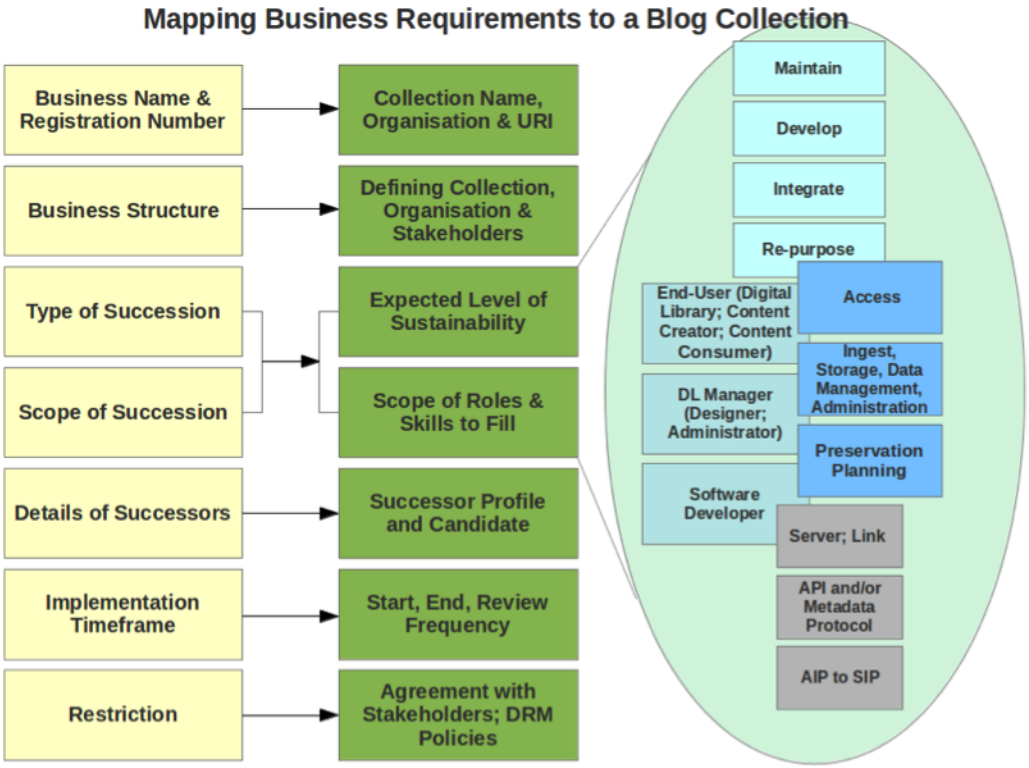

Figure 10: Succession levels and required skills and roles.

6.2.4 Scope of succession versus skills and roles to be filled

In the eventuality of the owner or key staff quitting the business, the succession plan consists of a plan on how to fill these roles. This often involves forward thinking in terms of identifying persons that might fill key roles and developing their skills to meet the challenges when succession takes place. Likewise, in the repository, once the level of expectations is understood, the resources required to support collection sustainability can be derived from the functional entities of the repository. In the language of the OAIS model[51], depending on the levels of expectations, this would culminate in a selection of roles with respect to Producer, Manager, Consumer, and Archive (Ingest, Storage, Data Management, Administration, Preservation Planning, Access) and their associated sub processes as required skills (see yellow boxes in Figure 10). In parallel, the roles (digital librarian, content creator, content consumer, system designer, administrator, software developer) of the Digital Library Reference Model[52] need to be fostered.

[51] http://public.ccsds.org/publications/archive/650x0m2.pdf

[52] http://www.cs.ucsb.edu/~anika/D3.2bDigital_Library_Reference_Model.pdf

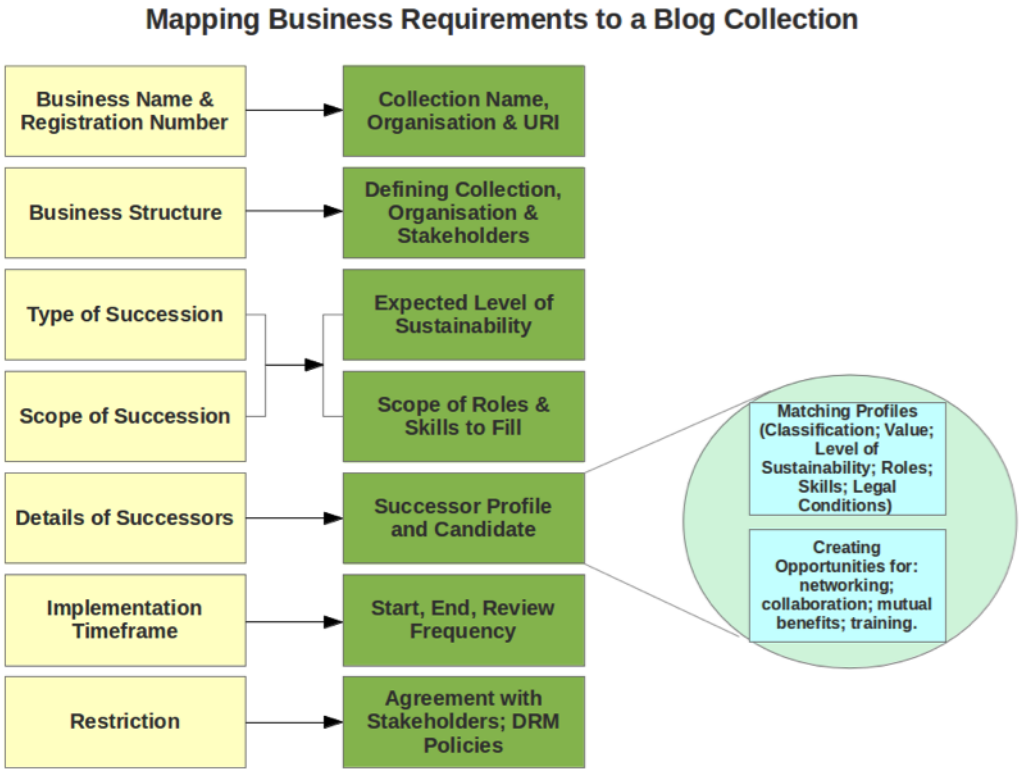

Figure 11: Identifying successors and generating willingness to succeed.

6.2.5 Details of successor versus successor profiling and identification of candidates

Within a business where key roles are identified, risks assessed and replacement personnel are planned and developed, target individuals can be detailed. In a repository, however, the challenge in sustaining an existing collection lies in finding one or more organisations *willing to take on the responsibility* of managing the inherited holdings to meet the requirements of the collection with respect to stakeholder requirements and legal obligations. The requirements and obligations that can be met, subject to the extent of effort that any succeeding organisation, if any, would be willing to put into the upkeep and preservation of the material, need to be made explicit to the stakeholders.

Unless the collection is abandoned, even if the expectation threshold is chosen to be minimal at the collection maintenance level, the responsibility to meet the requirements of digital preservation is still non-trivial. For example, concerns range from maintaining availability (e.g. providing adequate technical support for uninterrupted web service), security procedures (e.g. user authentication, management of access privileges, virus check), digital rights management (e.g. monitoring correct use of digital assets), and, accessibility to content (e.g. migrating content to viable formats, updating metadata in existing records, correcting errors). This involves keeping up to date with emerging technologies, and maintaining staff capable of managing administrative functions. The financial burden that results from taking on responsibility for these activities often leads to abandonment.

As part of a succession plan, it is essential that the existing repositories generate the basis for securing a commitment to preserve the material from one or more competent succeeding organisations. To achieve this, the following essential steps are recommended (see Figure 11):

- Identification of a number of potential successors (providing details of Section 6.2.1) by matching the classifications, collection usage, expectations, and roles identified in Sections 6.2.2, 6.2.3, and 6.2.4, in view of the legal condition identified in Section 6.2.6. This will form a Usage scenario.

- Explicit and transparent description of gaps between the usage scenarios with respect to the source repository and those of the potential successor organisations.
- Proposal for bridging the gaps by early initiatives to create common purposes.

The last of these could be facilitated by regular training/networking events that allow the repository to collaborate and liaise with successor candidates. These connections should be in place long before the need for succession is likely to arise.

6.2.6 Legal considerations

In the case of a business, there are various forms of legal obligations, contracts, and agreements that the business enters into as part of the business process (see Section 6.1). For example, this could consist of stipulated restrictions on potential successors; insurance contracts; partnership agreement; buy-sell agreement; succession agreement; will and testaments; employment contracts; payout agreement; and, licenses. Similar legal obligations arise when a repository is established in relation to the stakeholders: i.e. end-users, repository managers, and software developers. These legal considerations will naturally influence the identification of successors as the legal policies and obligations may conflict with succession candidate policies. To resolve the conflict, the repository should involve stakeholders at all levels in the succession process (see Figure 12 - for example, by allowing content creators to opt out of the new repository agreement).

6.2.7 The overall time frame and frequency of review

In Figure 13, we have outlined the stages involved in a succession plan. This succession plan is initiated from the time of establishing the blog collection. It starts with identifying candidate successors for the collection. This involves applying the step 6.2.5 detailed above. As there is no exact match between two organisations, effective solutions are likely to be developed by creating opportunities for synergies and training exchanges between source and successor organisations. Each of the steps outlined in Sections 6.2.1, 6.2.2, 6.2.3, 6.2.4, 6.2.5, and 6.2.6 should be reviewed on a regular basis for source and candidate organisation as details may change over time. The frequency of this review depends on the projected funding and profit margin of the repository in question. In addition, it is recommended that risk assessment regarding the probability of each step failing and the severity of its eventuality be performed at the same time as the success plan is reviewed (see orange boxes at the bottom row of Figure 13).

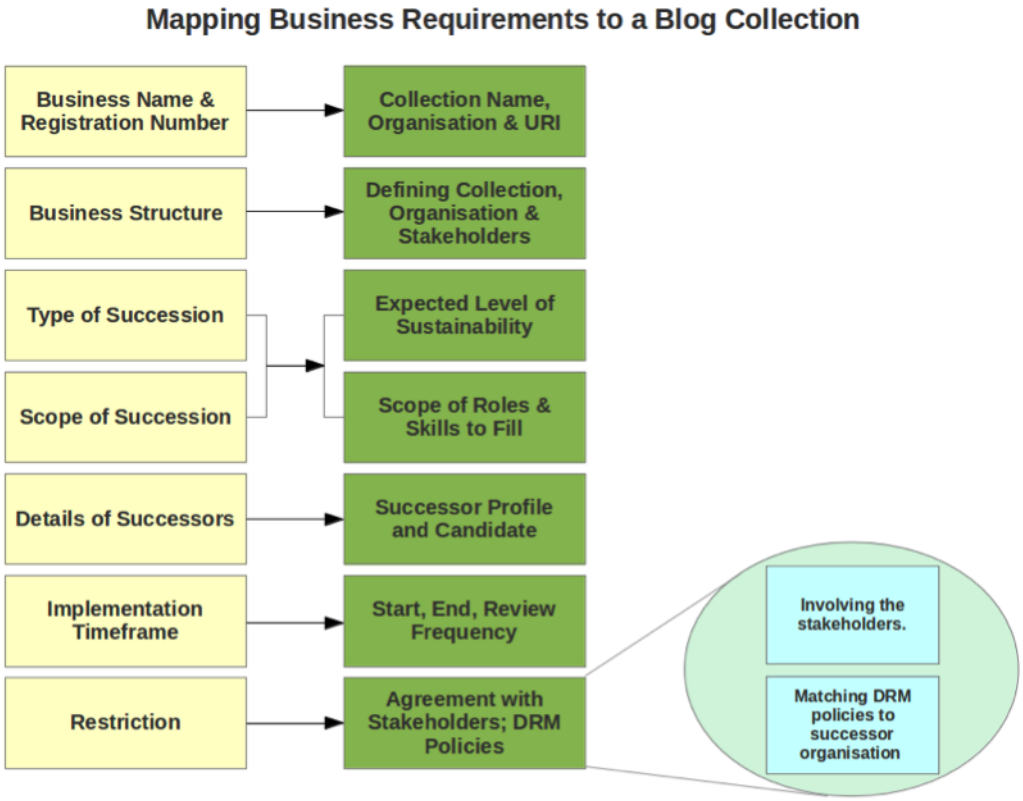

Figure 12: Matching and resolving legal requirements.

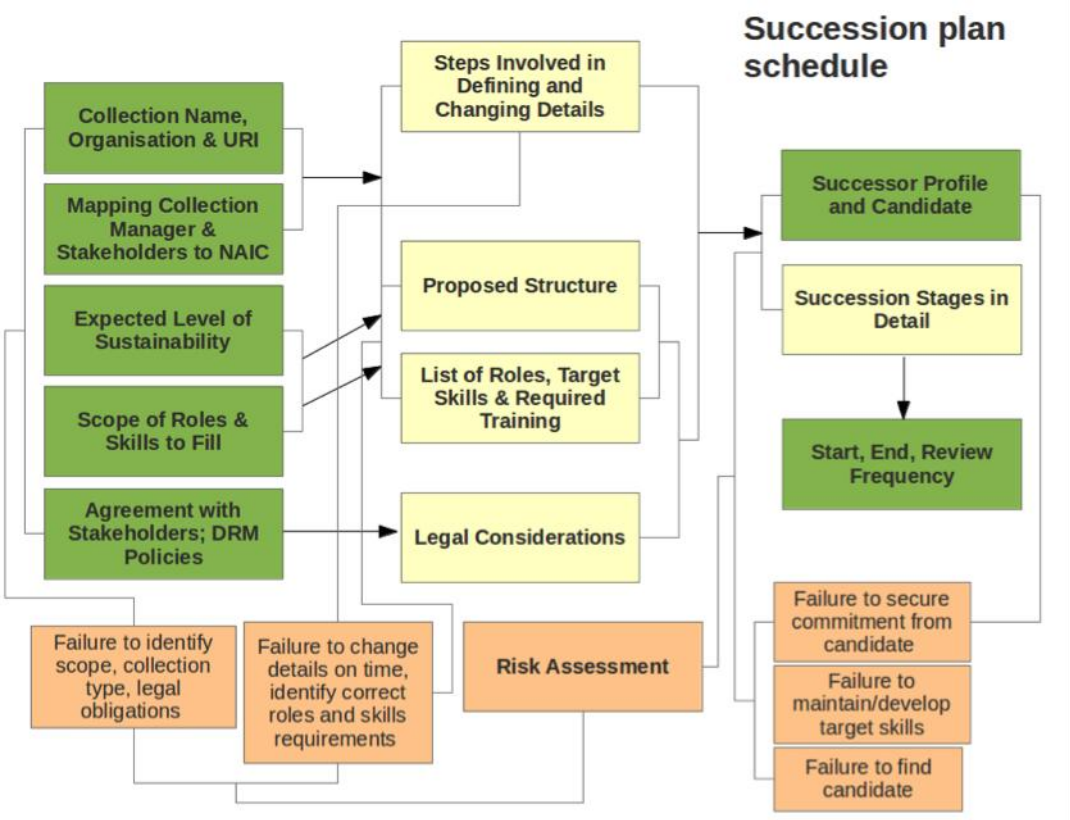

Figure 13: Succession plan schedule and risk management (identified risk in orange boxes). The arrows indicate the flow of activities in succession planning. Where there are no arrows, these connect events that are concurrent are not sequentially ordered.

6.3 Market Analysis of Blog Usage and Value: an Example

The purpose of this section is to demonstrate feasibility of blog market analysis and to highlight the increasing value that blogs have as communication channels for organisations that might need to be considered in a succession plan.

There are several guides on carrying out product market analysis (e.g. Aaker, 2008). Here we are not trying to present a thorough analysis of product value with respect to repository blogs. This section is only intended to demonstrate how one might carry one out a review of blog usage and value with respect to not only the repository but also the stakeholders (e.g. content owners) of the blogs and potential successor organisations.

The approach we adopted here consisted of two strands of investigations: one based on desk research of usage with respect to several types of blogs (the list is provided in Appendix B) and the other based on a short survey consisting of questions (the questions are provided in Appendix C) formulated from an analysis of the desk research. The desk research was used to identify key types of blog usages within the relevant domains; the subsequent survey was used in an attempt to validate the identified types of usage and perceived value.

We looked at three types of blogs: fifteen corporate blogs, four government blogs, and three personal blogs. The emphasis on corporate blogs was deliberate to complement other studies within BlogForever that have focused on academic or personal blogs. We used the list circulated by sitepoint (www.sitepoint.com), reviewing "15 Companies That Really Get Corporate Blogging"[53], as our main source of corporate blogs. In an actual digital repository market analysis, the blogs in repository plus related blogs (e.g. other stakeholder blogs or blogs in the same market – as defined by the classification of Section 6.2.2) should be targeted. The success of the blogs (that is, how much it is used, how frequently, by how many people, in how many ways), who uses them and how fast it is growing are all central factors in the market analysis[54].

The trends that our research reveals is that blogs have become widely used and are no longer confined to private use only. Many companies have their own blogs, which they use for various different purposes. It has been suggested that corporate blogs could be defined as websites where an organisation publishes and manages contents in order to achieve its goals (Lee et al., 2006). This definition however fails to take into account the many different ways in which companies do take advantage of the use of Blogs.

There are two main categories of corporate blogs, which are internal and external blogs. External Blogs can be read by the general public and internal blogs are only visible to people working in the company (Jackson et al., 2007). For obvious reasons, we were only able to gain an understanding of the different ways in which companies use external blogs. In a real world succession plan, where prior relationships and a level of trust have been established with other organisations through networking and training events (see Section 6.2.5), more insight into internal blog usage may become available,

The use of Blogs for Small and medium sized enterprises (SMEs) has various benefits. Corporate Blogs have shown to increase credibility with customers and stakeholders. They facilitate direct contact and make opinions more visible. Having a third party discussing a company's service or product gives the information more credibility. In this way blogs function as a combination of "distributed expertise, real-time collective response to breaking news and public opinion barometer" (Drezner & Farrell, 2004). A recent study (Kang, 2010) showed that Americans and Europeans trust the opinions of "average people" more than authorities. People place credibility in the form of dialogue and blogs tend to be perceived as free from authoritative gatekeepers (Shetty, 2012).

[53] http://www.sitepoint.com/15-companies-that-really-get-corporate-blogging/

[54] http://www.netmba.com/marketing/market/analysis/

For SMEs Blogs are an affordable way to increase the company's visibility and communicate with customers and stakeholders. SMEs can get bloggers to review their products and services on their private blogs. Companies thereby create mutually beneficial relationship with bloggers. The company benefits because their product or service is discussed on a public platform and the blogger benefits as she/he is able to create new and interesting content on her/his blog.

SMEs also use their blogs for information and documentation. Some government blogs upload minutes of meetings[55], making these available to the public. This makes the company seem more transparent by giving the public an insight into the company's day to day actions and the company's values and ideals. By archiving older posts and materials in the blog's archive the company can document processes and results and thereby successes and failures.

Corporate blogs are very often used to communicate and co-operate with stakeholders and customers. SMEs can create contact with experts in their field. Customers can comment on posts thereby giving the SMEs regular feedback on services and products. In this way blogs help to create important networks.

For example, we have studied the 37signal's blog (see Table 6) which has 100,000 RSS subscribers indicating the blog's success. The company develops apps for making communication in businesses easier. They took a very smart approach and created two blogs. One blog for talking about their ideals, vision and interests and another one for talking about their products. As one company they have created two blogs in order to communicate with two different audiences. When looking at all these different blogs it becomes clear that companies use blogs to give themselves a more approachable image, making them look more open, by giving customers, clients and stakeholders the opportunity to get involved with the content that they create and make available. Some of blogs provide a sense of concern for accountability, for example, the BBC blog (Table 6) uses their blog to explain editorial decisions.

Good and successful corporate blogs tend to create a lot of new contents, so a lot of data, often using not just text but videos, sound files and graphics. Losing this content would lead to viewing the content and the history of the company out of context.

Table 6: Sample characteristics of successful corporate blogs

Name	Website	Category	Author	Notes
Marriott Hotel	www.blogs.marriott.com	Accommodation and Food Services	One (CEO)	posts about the company's recent activities. Direct interaction with the CEO
37signals	http://37signals.com/svn/, http://37signals.com/news	Retail Trade	Multiple (employees)	focused on industry news and insights as well as product updates
BBC the Editors	www.bbc.co.uk/blogs/theeditors	Arts, Entertainment, and Recreation	Multiple	aims to explain the editorial decisions and dilemmas faced by the teams running the BBC's news service - radio, TV, and interactive.

[55] For example, http://www.biglotteryfund.org.uk/wales/about-big/our-people/england-committee-members/england-committee-meeting-minutes-and-agendas

Southwest airlines	www.blogsouthwest.com/blogsw	Transportation and Warehousing	Multiple	used for press releases and posts by former employees, friendly relaxed approach
General Motors	http://fastlane.gmblogs.com	Manufacturing	Multiple	If you make cars talk about cars

Further observations from sitepoint.com on the blogs of Table 6 and the results of our investigations in relation to government blogs and popular personal blogs are presented in Appendix B.

Based on the observations from the desk research we carried out on blog usage (outlined above), we developed a questionnaire of fourteen topics addressing blog usage, risks of loss, and interest in collaborative efforts to tackle shared concerns (see Appendix C, for a list of questions), to be distributed to targeted individuals responsible for organisational blogs. We did not want to send out the survey through e-mail lists or social media at this time, for two reasons: first, we wanted to test what we could achieve through a qualitative analysis of responses from contacts we know on a professional basis (because this may be one of the likely scenarios that will arise in the future with respect to digital repository managers trying to attract candidate successor organisations), and, second, the resources allotted for the investigation into succession planning did not afford the time necessary for a qualitative analysis of large survey response datasets.

Unfortunately, this meant that we were not able to collect many responses: only 4 people from different organisations (Academic, Government, and Commercial sector) returned answers to the survey. They all had 1-5 blogs within their organisation. People in all the associated organisations are encouraged to use and engage with the blogs. The two participants working in the commercial sector used their blogs both internally and externally. The government organisation only uses their blog for external activities. The loss of the blog would only lead to legal implications for the academic organisation. All the other participants said that "they would be sad to see their blog go but it would have no legal impacts for their organisation". Two participants, however, added that the loss of the blogs would make communication with stakeholders and customers a lot more difficult and possibly more expensive. Only one of the participants has taken steps to preserve the content of their blogs by backing it up. Sadly, only one of the participants said that they would be interested in joining a cooperative of stakeholders of blogging communities related to their organisation, in order to preserve their blogs.

To summarise, the survey shows that organisations tend to use blogs actively as communication channels but most of them have not thought about the impact that the loss of the blogs might have on their organisation.

6.4 Succession plan as an interoperability scenario

The BlogForever repository succession plan is aligned with the approach to developing interoperability scenarios introduced in earlier sections of this report. The steps are as follows:

1. Develop source organisation Usage Scenario
2. Identify target successor candidate organisations (see Section 6.2.5) and their Usage Scenarios.
3. Analyse gaps between source and target scenarios.
4. Prioritise best match organisations to develop solutions for the gaps that exist.
5. Take steps to implement solutions.

In Figure 14 we illustrate that the steps of Section 6.2 can be matched against the dimensions of Usage Scenario: Organisation, User, Quality, Functionality, Architecture, Content and Policies. An example of Usage Scenario for a fictitious source organisation is presented in Table 7 to

Table 12. Quality is not discussed in a separate table as it is related to expectations regarding all aspects of the archive. This links back to the level of expectations discussed in Section 6.2.3. Depending on the level of expectation (maintenance; development; integration; and re-purposing) placed on the succession, the demand on interoperability regarding each usage scenario dimension will differ.

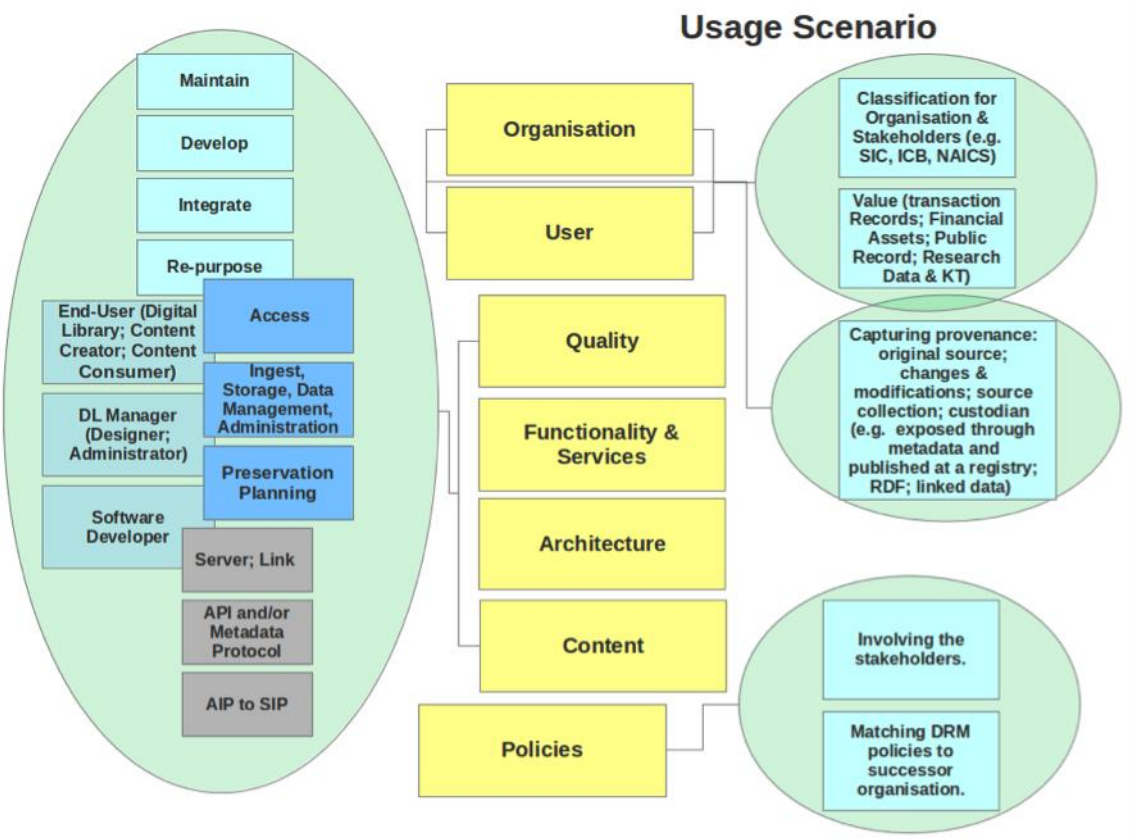

Figure 14: BlogForever succession plan as an interoperability usage scenario.

Table 7: Usage scenario: example description of source organisation.

Organisation	Example
Mission of organisation	The mission of Crazy Cat Archive is to provide a cost-effective long term preservation repository for blog content in support of the business activities of Crazy Cat Co.
Goal of organisation	To sustain collection value for stakeholders with respect to: 1. Transactions 2. Data, Information and Knowledge 3. Public Relations & Knowledge Transfer

Identity of organisation: 1. Ownership 2. Name 3. URI	Ownership : Joe Blogs Name: Crazy Cat Archive URI: www.fictitiouscrazycatarchive.com
Type of organisation: 1. Structure of the organisation 2. Classification	Structure: Partnership. Classification: Finance and Insurance - Commercial Banking (NAICS code 52211) http://www.census.gov/eos/www/naics/

Table 8: Usage scenario: example content of source organisation.

Content	Example
Primary objects	Blogs maintained/owned by Crazy cat Co. Blogs that discuss Crazy Cat Co. Products and services across the globe but with emphasis on blogs in the UK. Possibly blogs of other financial institutions in the UK after acquiring permission.
Categorised content	Textual content of the blogs. Images at local directories used within the blog. Audio at local directories used within the blog. Video at local directories used within the blog. Tags and categories generated for the posts at the time that the blog was active. Links used within the blog (not target content).
Metadata	Descriptive metadata (this includes Title, author, creation/publication date, last modified date, language, geographic region). Technical metadata for categorised content specified in the document "Crazy Cat Archive Technical Metadata v05042013" (e.g. URI: http://www.fictitiouscrazycatarchive.com/CCA_technical_metadata.xml). Provenance metadata in PREMIS (this includes source information describing name and URI of previous and current custodian, and source blog) Rights metadata in PREMIS (ownership, access, modification, distribution, and copyright). Category and tags generated by users when blog was active. Category and tags generated by users while interacting with Crazy Cat Archive.

Table 9: Usage scenario: example users of the archive at source organisation.

Users	Example
End-user	Banking product consumers. Share holders. Potential bank customers. Employees and management of the bank. Crazy Cat Archive staff.
Blog Archive Manager	Systems administration team.
Software developer	Configuration and adaptation of BlogForever software.

Table 10: Usage scenario: example functionalities and services of the source archive.

Functionalities and Services	Examples
Functionality and services	Organisation of content by selected metadata (see Table 8): that is, • descriptive metadata • technical metadata • provenance metadata • rights metadata Export of selected content (blogs, posts, comments and/or associated content) as authenticated documents. Harvest of company owned blogs. Search and harvest of blogs that mention Crazy Cat Archive. Tracking and harvesting selected content of other financial institution upon negotiated permission. Plugin for expressing negotiated permission. Provision of archival functions: • Ingest and storage of harvested material (in particular, translation of ingested material into Archival Information Package (AIP). • Update and management of associated data. • Dissemination of AIPs as dissemination information packages. • Transparent preservation planning.

Table 11: Usage scenario: example architecture of source archive.

Architecture	Example
Software components	Spider/crawler. Ingest. Storage (AIP storage; two copies of entire blog one offline, one location off site). Data management. Preservation planning and action. Administration (coordination of the above).
Hardware components	Number of servers to support process performance and harvest functions. Storage hardware.

Table 12: Usage scenario: example policies of source blog archive.

Policies	Example
End-user terms, conditions, agreements	Terms and agreements with content contributors (e.g. URI: http://www.fictitiouscrazycatarchive.com/conditions_of_use)
DRM Policy	Blog are only archived with blog owner permission. Contributors are bound by End-user terms that allow archiving content for preservation and research purposes. Not intended for general distribution. Blog are only accessible by Crazy Cat employee, management, shareholder s in its archived form.
Legal issues in relation archive staff	Pay-out agreement; employment contract.
Business structure	Buy-sell agreement; partnership agreement.
Legal obligation with respect to other bodies	Insurance; licences; will and testament.

To develop the interoperability scenario, assume that we have identified Archive 1, 2, and 3 (referred to as A1, A2, A3, respectively). Each possible solution for the interoperability scenario that defines a succession plan is developed as a candidate successor organisation. Assessment is based on how well they match and the cost involved in bridging the gap. A simplified example of interoperability requirements is presented in Tables OR1 - OR4, SR1, and TR1.

Organisational Requirements

OR1 – Shared organisational profile

Description	Organisations with similar classification and structure are intuitively more likely to support a succession scenario.

Assessment / Measures	Comparison of SIC, ICB, or NAICS classification
Comments	Example quantification of measure: Exact match scores 0; share parent 1; no shared classification 2
Required for	IS - Succession plan

OR2 – Shared roles and skills between source and successor organisation.

Description	Skills and roles required for the archive must ideally exist in the candidate organisation.
Assessment / Measures	Check list of functionality (see Table 10) and existence of responsible person for the functionality.
Comments	Add 1 for each time candidate fails to meet role or skill.
Required for	IS - Succession plan

OR3 – No legal impediment to the succession.

Description	And candidate organisations must be clear of all legal restriction with respect to succession (e.g. rights of source organisation must be transferable to successor organisation) and DRM policies must be acceptable by source, successor and other stakeholders (for example, content contributors).
Assessment / Measures	Check list of policies (see Table 12) and existence of procedure for transferring legal rights and responsibility.
Comments	Example quantification of measure: Exact match scores 0; share parent 1; no shared classification 2
Required for	IS - Succession plan

OR4 – Access to archival holdings of source organisation must be transferable.

Description	Method must exist for transferring the access to archival holding from the source archive to the new successor archive.
Assessment / Measures	Is the same information package used by the source and successor organisation? Is there an exchange protocol, standard, or translation tool to support the transfer of material into the new archive, and/or to enable federated search across the two collections?
Comments	Example quantification of measure: Exact match scores 0; share parent 1; no shared classification 2

Required for	IS - Succession plan

Semantic Requirements

SR1 – The Information Package (IP) of the source organisation must be understandable by the successor organisation.

Description	The information package of one repository will differ from that of another. There must be steps taken to close the gap so that information disseminated to the new stakeholders makes sense.
Assessment / Measures	Does the agreement of IP model already exist? If not, does a translation tool or map to an exchange standard exist?
Comments	Example quantification of measure: Agreement already exists 0; map to exchange standard exists 1; custom translation tool exists 2; a mapping must be created 3
Required for	OR4

Technical Requirements

TR1 – The metadata presented at dissemination must be consistent.

Description	Exchange protocol across collection of the new organisational structure must exist. It must consistent across all items returned and must allow the organisation of items with respect to specified conditions on selected metadata.
Assessment / Measures	Does the agreement of metadata model already exist? If not, does a translation tool or map to an exchange standard exist?
Comments	Example quantification of measure: Agreement already exists 0; map to exchange standard exists 1; custom translation tool exists 2; a mapping must be created 3
Required for	OR2

An example interoperability assessment is presented in Table 13. This example scenario suggests that the best solution might be A1. However, in an ideal approach, all of the top solutions should be pursued within the limits of the resources available at the archive to mitigate risks associated to the failure of A1 as a successor (e.g. unable to develop/maintain the skills required – this is especially a concern given its high cost score with respect to OR2).

Table 13: Example assessment: the larger the score the higher the cost and less desirable as a solution.

	A1	A2	A3
OR1	0	0	0
OR2	5	1	7
OR3	3	2	1

OR4	0	3	2
SR1	0	1	1
TR1	1	4	2
Total	9	11	13

6.5 Conclusions

In this part of the "BlogForever D3.2: Assessment of Prospects of Interoperability", aspects to be considered for the development of a succession plan for future BlogForever repositories were investigated. This resulted in a range of steps recommended to be covered for the smooth transition of repository stewardship (Section 6.2).

We have also provided an example Market Analysis (Section 6.3) in an attempt to determine the value of blogs. The analysis provided is intended to be a limited pilot test case. It needs further exploration before any firm conclusions can be drawn on its suitability for market analysis of future blog repositories. The purpose of the study presented here is to demonstrate feasibility of blog market analysis and to highlight the increasing value that blogs have as communication channels for companies and governments that might need to be considered in a succession plan. This would be key to diagnosing emerging threats that might lead to the repository being unable to continue, to recording immediate financial obligations, to identifying candidate successor organisations that share concerns, and, to generating commitment from candidate successor organisations to support, maintain and preserve the blogs.

We have also tried to align succession planning to an interoperability scenario as a special case. We have tried to follow the steps for interoperability scenario development to highlight the key issues that might arise and need addressing (Section 6.4).

A key component at the heart of a succession plan, unlike other interoperability scenarios that have been presented in this report, is the relative weight that is placed on organisational interoperability. This manifests in two ways: the strategy involved in generating commitment from a successor organisation to buy-in to maintaining, developing, integrating and re-purposing the blog collection, and, the demand on periodic review and continued risk assessment to revise the possible solutions and implementation steps.

The proposal here is intended to be a foundational guideline to be used by future BlogForever repositories and therefore does not take into consideration fine-grained levels of processes that might differ across different repository contexts. It is subject to continued development.

7 Conclusions

Aim of this report was the evaluation of interoperability prospects of the BlogForever platform with third parties. The future adoption and deployment of the BlogForever platform, e.g. by libraries, should be facilitated. Therefore, the report

- Reviewed interoperability models in the existing literature,
- Conducted a Delphi study to identify crucial aspects for the interoperability of web archives and digital libraries,
- Examined interoperability standards and protocols regarding their relevance for the BlogForever,
- Proposed a simple approach to consider interoperability in specific usage scenarios, and
- Presented an approach to develop a succession plan that would allow a reliable transfer of content from the current digital archive to other digital repositories.

Interoperability is still an ambiguous concept that can be interpreted either in a narrow sense that focuses on technical aspects, or a more comprehensive perspective that include non-technical aspects like organisational and legal constraints. Section 2 summarized several conceptualizations that have been proposed about interoperability models and levels and provided an extensive overview of the current literature on interoperability.

A Delphi study, conducted within the framework and the needs of the project, about aspects of interoperability of web archives and digital libraries was presented in section 3. The study reveals remarkable insights regarding current problems, limitations, needs and challenges that are encountered in today's interoperations (or efforts to this direction) among systems of the web archiving and digital library communities. It contributes to the limited so far empirical research for interoperability, presenting the current barriers but as well suggestions for future approaches, and can be a useful study for the three involved communities: the web archiving, the digital library and the digital preservation community.

An extensive review of the standards that support, assist or establish interoperability was provided in section 4. Standardisation is probably the most essential aspect of interoperability since standards can be the bridge between two or more different environments. The review provides a useful guide about the most commonly used standards that address the needs for interoperation mainly in the domains of digital libraries and web archives since these are the most relevant systems for BlogForever to interoperate with.

Section 5 presented a 5-step approach to consider and configure the enabling of interoperability of the BlogForever system (or any digital library) with another potential information system. This approach offers a useful and concrete guideline for managers, not only to realize future interoperations with BlogForever, but also to use as a basis to build upon their own interoperability methodology tailored to their own environments and needs. The description of the approach proposes several templates to assist the documentation of the steps. Furthermore, an example is given to facilitate the understanding of the application of the approach in the context of BlogForever.

Section 6 presented a tangible approach to develop a succession plan that would allow a reliable transfer of content from the current digital archive to other digital repositories in the case that a future BlogForever Archive is unable to continue for any reason. Concepts from the business model are employed for succession planning, and are adapted to the digital repository context, to suggest steps for achieving organisational interoperability. The succession plan development is also elaborated upon using the interoperability scenario development framework.

The results of this report inform on the one hand further development of the BlogForever platform, e.g. through revealing the relevance of the WARC standard for interoperability prospects. On the other hand, future deployment of the platform in real life scenarios is supported through the presented guidelines and approaches. Additionally, most of the findings are not just applicable for the BlogForever project but inform also the web archiving and digital preservation community in general.

8 References

Aaker, D.A., 2008. *Strategic market management*. John Wiley & Sons.

Anon., 2004. *European Interoperability Framework for Pan-European eGovernment Services*. Luxembourg: Office for Official Publications of the European Communities.

Arms, W.Y. et al., 2002. A Spectrum of Interoperability: The Site for Science Prototype for the NSDL. *D-Lib Magazine*.

Athanasopoulos, G. et al., 2011. *D3.4 Digital Library Technology and Methodology Cookbook*. DL.org.

Athanasopoulos, G. et al., 2011. *Digital Library Technology & Methodology Cookbook: An Interoperability Framework, Best Practices & Solutions*.

Banos, V., 2011. *OAI-PMH validator and data extractor*. [Online] Available at: http://www.oaipmh.com [Accessed 10 April 2012].

Banos, V. et al., 2012. Technological foundations of the current Blogosphere. In *International Conference on Web Intelligence, Mining and Semantics (WIMS'12)*. Craiova, Romania, 2012.

Camlon H. Asuncion, M.J.v.S., 2010. Pragmatic Interoperability:A Systematic Review of Published Definitions. In *IFIP*., 2010.

Candela, L. et al., 2011. *D3.2b The Digital Library Reference Model*. DL.org.

Commission, E., 2004. *EIF - European Interoperability Framework for pan-European eGovernment services*. [Online] Available at: http://ec.europa.eu/idabc/en/document/2319/5938.html [Accessed 20 April 2012].

DELOS, A.N.o.E.o.D.L., 2005. *D5.3.1: Semantic Interoperability in Digital Library Systems*. Deliverable. DELOS, A Network of Excellence on Digital Libraries.

DL.org, 2008. *Coordination Action on Digital Library Interoperability, Best Practices & Modelling Foundations*. Deliverable.

DOI, n.d. *The DOI System*. [Online] Available at: http://www.doi.org/ [Accessed 17 April 2012].

Drezner, D.W. & Farrell, H., 2004. Web of Influence. *Foreign Policy*, pp.32-40.

Duval, E., 2002. Metadata Principles and Practicalities. *DLib Magazine*.

Geraci, A., 1991. *IEEE Standard Computer Dictionary. A Compilation of IEEE Standard Computer Glossaries*. Piscataway, NY, USA: IEEE Press.

Gomes, D., Miranda, J. & Costa, M., 2011. A Survey on Web Archiving Initiatives. In Gradmann, S., Borri, F., Meghini, C. & Schuldt, H., eds. *Proceedings of the 15th international conference on Theory and practice of digital libraries: research and advanced technology for digital libraries (TPDL'11), LNCS 6966*. Berlin, Heidelberg, 2011. Springer-Verlag.

Gottschalk, P., 2009. Maturity levels for interoperability in digital government. *Government Information Quarterly*, pp.75-81.

Gradmann, S., 2007. *Interoperability. A key concept for large scale, persistent digital libraries*. Briefing paper. Digital Preservation Europe.

Grotke, A., 2008. *2008 Member Profile Survey Results*. International Internet Preservation Consortium.

HANDLE.NET, 2010. *Technical Manual Version 1.1, CNRI*. [Online] Available at: http://hdl.handle.net/4263537/5043 [Accessed 17 April 2012].

HANDLE.NET, n.d. *HANDLE.NET Systems Fundamentals*. [Online] Available at: http://www.handle.net/overviews/system_fundamentals.html [Accessed 17 April 2012].

Hans-Werner Hilse, J.K., 2006. Implementing Persistent Identifiers. In *Consortium of European Reasearch Libraries(CERL)*. London, 2006.

Herbert Van de Sompel, P.H.B.-A., n.d. *OpenURL syntax description*.

Higgins, S., 2007. *Digital Curation Centre: What are Metadata Standards*. [Online] Available at: http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards [Accessed 26 April 2013].

IEEE, 1990. *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries.* New York, NY: Institute of Electrical and Electronics Engineers.

Innocenti, P. et al., 2011. Towards a Holistic Approach to Policy Interoperability in Digital Libraries and Digital Repositories. *The International Journal of Digital Curation*, pp.111-24.

Internet Memory Foundation, 2010. *Web Archiving in Europe*.

Invenio, 2012. *Invenio #903 Real OpenURL link resolver*. [Online] Available at: http://invenio-software.org/ticket/903 [Accessed 17 April 2012].

Jackson, A., Yates, J. & Orlikowski, W., 2007. Corporate Blogging: Building community through persistent digital talk. In *Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS'07)*. Hawaii, 2007.

Jacobsen, G., 2007. Webarchiving Internationally: Interoperability in the Future? In *World Library and Information Congress: 73rd IFLA General Conference and Council*. Durban, South Africa, 2007.

Kang, M., 2010. *Measuring Social Media Credibility: A Study on a Measure of Blog Credibility*. Institute for Public Relations, www.instituteforpr.org.

Kumar, S., 2009. *Interoperability Protocols and Standards in LIS*. [Online] Available at: http://www.slideshare.net/alibnetweb/interoperability-protocols-and-standards-in-lis [Accessed 26 April 2013].

Lagoze, C. & van de Sompel, H., 2001. The Open Archives Initiative: Building a low-barrier interoperability framework. In *ACM/IEEE-CS Joint Conference on Digital Libraries*. Roanoke, 2001.

Landeta, J., 2006. Current validity of the Delphi method in social sciences. *Technological Forecasting & Social Change* , pp.467-82.

Lannom, L., 2008. *Handle System Workshop Introduction, Handle System Workshop*. Brussels.

Lawrence, S..F.C.E.G.D.P.G.F.F.N.R.K.A.K.C.L.G., 2001. Persistence of Web References in Scientific Research. *IEEE Computer, Volume 34, Number 2*, pp.26-31.

Lee, S., Hwang, T. & Lee, H.-H., 2006. Corporate blogging strategies of the Fortune 500 companies. *Management Decision*, pp.316-34.

Linstone, H.A. & Turoff, M., 2002. Introduction. In H.A. Linstone & M. Turoff, eds. *The Delphi Method: Techniques and Applications*. Addison Wesley. pp.3-12.

Manso, M.-Á., Wachowicz, M. & Bernabé, M.-Á., 2009. Towards an Integrated Model of Interoperability for Spatial Data Infrastructures. *Transactions in GIS*, pp.43-67.

Morris, C., 1938. *Foundations of the Theory of Signs*. Chicago: Univ. of Chicago Press.

National Diet Library, 2010. *18th CDNLAO Annual Meeting: CDNLAO Questionnaire Survey on Web Archiving (Q2-Q7)*. Japan.

Okoli, C. & Pawlowski, S.D., 2004. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, pp.15-29.

Paskin, N., 2010. *Digital Object Identifier (DOI) System*. Taylor & Francis.

Rothwell, W.J., 2010. *Effective Succession Planning: Ensuring Leadership Continuity and Building Talent from Within*. 4th ed. New York: AMACOM.

Shetty, S.V., 2012. Public Relations and New Media: Stakeholders: Perspective on Corporate Credibility through Blogging. In *Forty Years of Media and Communication in Asia: Retrospect, Introspect and Prospects*. Concorde Hotel, Shah Alam, Malaysia, 2012.

Skumolski, G.J., Hartman, F.T. & Krahn, J., 2007. The Delphi Method for Graduate Research. *Journal of Information Technology Education*.

Tolk, A., Diallo, S.Y. & Turnitsa, C.D., 2007. Applying the Levels of Conceptual Interoperability Model in Support of Integratability, Interoperability, and Composability for System-of-Systems Engineering. *Journal of Systemics, Cybernetics and Informatics*, pp.65-74.

Tolk, A. & Muguira, J.A., 2003. The Levels of Conceptual Interoperability Model. In *Proceedings IEEE Fall Simulation Interoperability Workshop*. Orlando, Florida, 2003. IEEE Press.

A. UW's report on WARC

The brief report below was written by Karen Stepanyan (UW) after some consultation with UoL project partners and circulated in November 2012.

Web ARChive (WARC) is an archive file format that allows combining multiple digital resources into an aggregate archival file [1]. The WARC format represents a revised and generalised version of the earlier developed ARC format that is used by the Internet Archive for storing blocks of information harvested by crawlers [2]. The rationale for using WARC can be explained by the challenges arising from crawling and capturing large number of constituent data objects for the purpose of storage, management, and exchange. Unlike ARC files, WARC provides a mechanism for recording of HTTP request headers, arbitrary metadata, the allocation of identifier for every contained file, management of duplicates and migrated records, the segmentation of the records [3]. It also enables later transformation of the objects and migration [4].

The use of the WARC format is being more popular due to its use by the Heritrix web crawler from the Internet Archive and various other applications that enable creation and processing of WARC files (e.g. WARCreate Google Chrome extension, though not yet public). The benefits of using the format include: URL-based look-up and browsing; full-text search using Nutch/NutchWax or Hanzo Search; possible use of customised/extended GUI (e.g. browsing by subject, collection or alphabetically) using Wayback [3].

Yet, while effective with working with various types of digital objects, WARC files generated by web crawlers constitute records and content blocks associated with crawled resources. The harvested content represents a snapshot of the resource with records that are annotated by their URL, IP-address, archive-date, content-type, result-code, checksum, length and so on. Hence, identifying sections of the web page that correspond to specific conceptual elements of the blog (e.g. post, comment, author etc.) becomes challenging. While acceptable for services limited keyword search, this presents a problem for implementing faceted search.

Other limitations of WARC include the lack of support and time constraints for implementation. When used with evolving open source tools, using WARC raises challenges for subsequent maintenance.

Summary:

- WARC can accommodate a data model, but development and maintenance can be resource/time expensive.
- WARC is good for transferring and managing large number of files
- WARC is good when the content is to be presented via existing tools for browsing/searching WARC files (this is not the case for BlogForever).

Given the above, the use of WARC in BlogForever can benefit from the perspective of: [a] transferring files and [b] providing an alternative way of exploring collected data (other than Invenio). Given the time required for developing tools for generating WARC, this is unlikely to be feasible within the given time frame of the project.

References:

[1] A. Arvidson, G. Mohr, and M. Stack. (2007, 08.10.2012). The WARC File Format (0.16 ed.). Available: http://archive-access.sourceforge.net/warc/warc_file_format-0.16.html

[2] M. Burner and B. Kahle, "Arc file format," 1995.

[3] H. Hockx-Yu. (2009, 10.10.2012). Web Archiving Tools: An Overview. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.177.3019&rep=rep1&type=pdf

[4] S. Strodl, P. P. Beran, and A. Rauber, "Migrating content in warc files," in The 9 th International Web Archiving Workshop (IWAW 2009) Corfu, Greece, September/October, 2009 Workshop Proceedings, 2009.

B. List of Blogs Examined

Corporate Blog: description from http://www.sitepoint.com/15-companies-that-really-get-corporate-blogging/

- Dell (http://direct2dell.com/): "Though Dell's corporate blog rarely strays from Dell-centric news, the company posts with a great conversational voice, often breaks news on their blog (which keeps people coming back), and listens and responds to customers. Dell also posts regularly (1-2 posts per day at least) which keeps content fresh and encourages repeat visits."
- Lenovo (http://lenovoblogs.com/): "The great collection of blogs from computer maker Lenovo demonstrate that the company really understands blogging. Lenovo intersperses posts about its product line with musings about business, design, life, and technology. Definitely don't miss the Design Matters blog, which should be a must-read for any designer."
- 37signals (http://www.37signals.com/svn/): "37signals is kind of the poster child for corporate blogging. Their 'Signal vs. Noise' blog has almost 100,000 RSS subscribers and there's a good reason: 37signals rarely blogs about their products anymore (they split off a separate product-only blog for that), but instead shares advice and insights about business, design, editorial, and other topics."
- Adobe (http://blogs.adobe.com/): "Adobe offers a huge collection of employee blogs, many of which are great reads. By allowing employees to blog, Adobe has empowered them to evangelize their products for them — many post tutorials, advice, reviews, and other great tid-bits promoting Adobe products — while not pigeon holing them into talking *only* about Adobe."
- BBC (http://www.bbc.co.uk/blogs/): "In addition to their news blogs, the BBC publishes a series of behind-the-scenes blogs. They're tremendously interesting, especially The Editors blog, in which BBC News editorial staff dissect the broadcaster's news coverage and the media industry in general."
- Southwest Airlines (http://www.blogsouthwest.com/blogsw) "Southwest Airlines' 'Nuts About Southwest' blog doesn't take itself too seriously — and that's a good thing. The company blogs about itself and the airline industry with a personal touch and has been producing a series of fun, behind-the-scenes videos that are both interesting and engaging."
- Sun Microsystems (http://blogs.sun.com/): "Like Adobe, Sun allows their employees to blog. They've been doing it for a long time, and their blog portal has over 4,500 bloggers covering over 110,000 posts. Some of their blogs, such as that of Web 2.0 and Web Services Evangelist Arun Gupta, have become quite popular on their own. That's 110,000 posts of promotional gold for Sun and they know it."
- Freshbooks (http://www.freshbooks.com/blog/): "Taking a page from 37signals, the team at Freshbooks uses their corporate blog to share advice and insights into their way of doing things. Slowly, and in large part due to their blog, Freshbooks is turning their users into true fans."
- Marriott International (http://www.blogs.marriott.com/): "Marriott on the Move is the official blog of Marriott Hotels, Resorts, and Suites Chairman and CEO Bill Marriott. Though a self-described technophobe, Marriott uses the blog to talk about his thoughts and opinions on all sorts of things related to being a hotelier. Marriott, who was recently featured on NBC Nightly News for a story on corporate blogging, says he blogs because it is 'a great way to communicate with [your] customers and stakeholders in this day and age.' We agree."

- Seagate (http://storageeffect.com/): "Penned by Seagate Global Marketing Manager Pete Steege, Storage Effect is a must-read blog for anyone in the computer storage industry. Beyond first looks at upcoming Seagate products, Steege mixes it up with musings about the industry and fun posts like a recent one about Batman's storage requirements."
- General Motors (http://fastlane.gmblogs.com/): "The GM Fastlane Blog is a great example of corporate blogging because GM has clearly realized that regurgitating press releases is not what blogs are made for. GM talks a lot on their blog about their cars and trucks and the design choices they make while creating them, but they also throw in interesting treatises on current hot-button issues, such as alternative energy."
- Quicken Loans (http://www.quickenloans.com/): "Quicken Loans publishes a handful of unique blogs — unique among corporate blogs in that they're not overly self-referential. Their What's the Diff? blog, for example, publishes stories about "things that make the difference in business and in life." The Quizzle blog, on the other hand, posts advice about how to understand the home loan market. It is all subtle marketing for Quicken Loans, but it is done in an informative and useful manner that will win points among customers."
- FiveRuns (http://blog.fiveruns.com/): "FiveRuns, who create products aimed at Ruby on Rails developers, also publish an excellent blog. Along with regular tutorials about how to do things with Rails and use their products, the FiveRuns team also posts weekly five question interviews with prominent members of the Rails community. Brilliant stuff."
- Accenture (http://www.accenture.com/Global/Accenture_Blogs/): "Consulting firm Accenture publishes 8 blogs that are definitely worth checking out. Rather than just blog about what Accenture can do for your business, the company has tapped some of its smartest employees to share insights on business, communications, technology, consulting, and hiring. A sample of recent posts: how to balance work and life, thoughts about Twitter's downtime, and musings on GTD theories."
- Amazon Web Services (http://aws.typepad.com/): "Amazon's Web Services blog is truly one of the great corporate blogs because it reads like a fan blog. You'd never know that Jeff Barr, the scribe behind the AWS blog, is a Web Services Evangelist at Amazon just by reading his posts."

Government Blogs:

- NASA (blogs.nasa.gov/cm/newui/blog/mainblogs.jsp): Family of 53 Blogs accessible through single web portal. different blogs by different employees administrators, Chief Technology Officer, Goddard Space Flight Center CIO, Astronaut. Cover different topics and some are written for specific audiences (kids, teachers, or general public). All blogs open for reader comment. Posts updated about once a week. Tone of posts is conversational and have a good length.
- Greenversations (environmental protection agency) (blog.epa.gov/blog): Multiple authors. Employees of the organisation. Different topics: green energy, sustainability and related issues. Lack of photos and visual materials. Very lively conversations underneath the posts.
- Army lives (armylive.dodlive.mil): Multiple authors. Topics of interest to Army's personnel, families and others who want to keep up with the doings of the service. Easy to use Facebook and Twitter buttons. Linking to flickr photo page and other sites of interest.
- TSA Blog (The Transportation Security Administration) (blog.tsa.gov): One author: Blogger Bob. One purpose: "to facilitate an ongoing dialogue on innovations in security, technology and the checkpoint screening process". Bob TSA policies defender. Posts articles in order to explain, justify and update the public TSA policies and activities. Posts generate many comments, in particular hostile wants. TSA has possibility to filter comments but seems to allow people to voice their opinions. Makes use of web to keep communication channels open.

Popular Blogs:

- Mashable (www.mashable.com): Technology and social media blog founded by Pete Cashmore in July 2005. Times noted mashable as one of the 25 best blogs in 2009. As of march 2013 it has over 3,200,000 Twitter followers and over 1,000,000 fans on Facebook. Divided into different news categories (Social Media, Tech, Businesses, Entertainment, US&World, Lifestyle) and 3 main headings: The new stuff, the next big thing, What's hot. Contents is easily shared over various social media channels. Site is very easy to navigate and has lots of visual content and little text. In the actual articles there are many links to other resources and websites. Big illustrations and pictures throughout the article. The tech section features a lot of product reviews, suggesting that the blog will get funding from the reviewed brands. There are links to retailers and the brand's website.
- Engadget (www.engadget.com): Multilingual technology blog network with daily coverage of gadgets and consumer electronics. Operates as a total of 10 blogs 4 written in English and 6 in a different language with independent editorial staff. Voted one of the best blogs in 2010 by TIME. Engadget was co-founded by former Gizmodo technology weblog editor and co-founder, Peter Rojas. Engadget is a member of Weblogs, Inc., a blog network with over 75 weblogs including Autoblog and Joystiq and formerly including Hack-A-Day. The English edition of *Engadget* operates four blogs which, like the international editions, have been assimilated into a single site with a sub-domain prefix. These include *Engadget Classic* (the original *Engadget* blog), *Engadget Mobile*, *Engadget HD* and most recently *Engadget Alt*.It launched in 2004 and its contents is updated several times a day. It often offers opinions within articles and also comments on the newest gossip in the tech world. Since its founding, dozens of writers have written for or contributed to *Engadget*, *Engadget Alt*, *Engadget Mobile* and *Engadget HD*, including high profile bloggers, industry analysts, and professional journalists. Darren Murph who became the World's Most Prolific Professional Blogger as recorded by Guinness World Records on July 29, 2010, is the site's Managing Editor and has written over 17,212 posts as of October 5, 2010. Google Reader, as well as many other RSS readers, has included *Engadget* as a default RSS feed, pulling the latest articles which appear at the top of all user's mailboxes. Other example of engadgets influence: In May 2007, Engadget published a story based on an email sent to Apple employees announcing that the company was delaying the launches of both the iPhone and Mac OS X Leopard. After the story ran, Apple's share price dropped 3%. Less than 20 minutes later the story was retracted after the email was discovered to have been a hoax perpetrated on Apple employees. Apple's shares eventually recovered and Ryan Block apologized for the mistake.
- Gizmodo (www.gizmodo.com): Is a technology weblog. It is part of the Gawker Media network run by Nick Denton and covers topics related to the technology industry, as well as other topics as broad as design, architecture, space, and science. The blog, launched in 2002, was originally edited by Peter Rojas, but he was recruited by Weblogs, Inc. to launch their similar technology blog Engadget. By mid-2004, Gizmodo and Gawker together were bringing in revenue of approximately $6,500 per month.

C. BlogForever Blog Usage & Value Survey

The following questions are the basis of an interview/survey being carried out on behalf of the BlogForever project (http://www.blogforever.eu) to develop sustainable strategies for the preservation and management of weblogs.

Part A. About yourself

1. What is your name?
2. What is the title of your position in your organisation?
3. What is your email address?

   Would you like to receive information on the results of this survey (delete as appropriate)?

   Yes

   No

Part B. About your Organisation

4. To what kind of organisation do you belong (please place a "x" next to all the items that apply to your organisation)?
   - Industry, Business, and/or Commercial
   - Government, State, and/or Local Council
   - Academic, Research, and/or Teaching
   - Other

     If "Other", please specify:
5. If your organisation belongs to an industrial organisation to which sector does it belong - these categories follow the The Industry Classification Benchmark# (please place a "x" next to all the items that apply to your organisation) ?
   - Basic Material
     - Basic Resources
     - Chemicals
   - Consumer Goods
     - Automobile & Parts
     - Food & Beverage
     - Personal & Household Goods
   - Consumer Services
     - Media
     - Retail
     - Travel & Leisure
   - Financials
     - Banks
     - Financial Services

- Insurance
- Health Care
- Industrials
  - Construction & Materials
  - Industrial Goods & Services
- Oil & Gas
- Technology
- Telecommunications
- Utilities
- Other

  If "Other", please specify:

6. Does your organisation have a website (delete as appropriate)?

   Yes

   No

   If yes, what is the URL of your organisation website?

Part C. Usage of blogs

7. Does your organisation own and/or maintain a blog as part of the activities related to the organisation (delete as appropriate)?

   Yes

   No

   If yes:

   - How many blogs are there?
   - It is used for activities (please place an "x" next to all items that apply):

     External to the organisation?

     Internal to the organisation?

     Both?

     Other?

     If "Other", please specify:

8. Are people in your organisation encouraged to maintain/visit organisational or independent blogs to carry out tasks related to organisational activities (delete as appropriate)?

   Yes

   No

9. Do your organisation and/or people in your organisation *purposefully* make use of blogs maintained by people outside of the company (delete as appropriate)?

   Yes

   No

10. How do you use blogs (please place a "x" next to all items that apply)?

- Do you contact other blog owners to ask them to discuss, review, or broadcast selected topics? If yes, how do you find these target blogs?
- Do you browse blogs as a source for up-to-date information and knowledge, and, as a collection of how-to guides? If yes, how do you find the blogs you use?
- Do you use blogs to:
    - communicate, co-operate and interact with stakeholders (customers, business partners, colleagues)?
    - talk about the organisation's mission, ideas, plans, and beliefs?
    - give the organisations a more human and approachable image, making sure everybody in the organisation is a) actively involved in adding content; and, b) engaged in talking to customers, colleagues and business partners?
    - make the organisational policy more transparent, increasing the company's credibility?
    - document the organisation's transactions, collaborations, and business ventures?
    - manage knowledge transfer within or beyond the company?
    - Other

      If "Other", please elaborate:

Part D. Risks assessment

11. What impact would the loss of blogs have on the organisation (Please give an example where possible)?
    - Legal (obligations; rights)?
    - Record-keeping (transactions; public)?
    - Knowledge base (source of knowledge; pipe-lining; valuable data)? Please give an example if possible.
    - Communication channels (social impact; trust)? Please give an example if possible.
    - Other

      If "Other", please specify:
12. List any steps you take to mitigate/manage these risks (e.g. backup and export)?

Part E. Expression of interest and commitment to blog preservation

13. Would you be interested in joining a cooperative of stakeholders, in blogging communities related to your organisation, aiming to preserve their blogs?

    For example, each member could provide part of the necessary resources (say, technical support, financial support, personnel time, expertise consultation, storage space, network, equipment) and/or collaborative effort (say, application for third-party support) to secure resources, in exchange for better access to channels of communication, customer reach, and/or blog preservation infrastructure. (delete as appropriate).

    Yes, I would be interested

    No, this would not be possible

    Depends

    If it "depends", what would be a deciding factor?

Part F. Permissions

14. Do we have permission to make all your responses to the following questions public for the purposes of improving approaches to the discovery, preservation, and management of blogs (delete as appropriate)?

    Yes

    No

    If 'No', which questions would you like to withhold from publication?

D. Delphi Study Questions & Detailed Results

A' Round Questions

All questions of the first round are formulated as open questions.

1. What in your view are the purposes of interoperability? What problems or opportunities are addressed with interoperability? Please reply with a descriptive answer, if possible using scenarios that describe the purpose, the partner institutions, and the systems that are involved.

   Think of problems that have been solved or problems that exist and require interoperability practices, problems that you either experience directly or you can identify. Additionally, think of benefits that occur from the interoperation between systems/institutions.

2. What are the main obstacles and limitations that prevent or hinder interoperability? *(technical, political, organizational, management, legislation or other barriers)*

3. What changes or developments in the landscape would, in your view, assist the interoperability of digital libraries and/or web archives (and how)?

   Think of technical changes/developments (e.g. standards, frameworks, services), political or legislation changes, new concepts etc.

4. What do you consider as future challenges regarding inter- operability of digital libraries and/or web archives?

   Think about important problems that have to be solved, obstacles to overcome, possible additional future barriers that may occur due to forthcoming changes in needs, technology, perspectives, legislation etc.

B' Round: Questions & Result Tables

8.1.1 Purposes

In the first round participants were asked about the general motivations for interoperability in the context of digital libraries and web archives and the responses were analysed.

We have identified the following purposes/motivations for interoperability. Please indicate to what extent you agree or disagree.

	Strongly disagree	**Disagree**	**Agree**	**Strongly agree**	I can't say / I don't know
Uses					
Federated Search *Ability to search over several web archives or digital libraries from a single point or with a single query (regardless of data accessibility)*	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*
Federated Access *Ability to access data (view, copy, print etc.) of several web archives or digital libraries from a single point or with a single query*	0% *(0)*	0% *(0)*	17% *(1)*	83% *(5)*	0% *(0)*
Exchange *(e.g. Exchange of data to create or complement specific collections about a particular event/topic)*	0% *(0)*	0% *(0)*	50% *(3)*	33% *(2)*	17% *(1)*
Replication *Aiming at data redundancy in order to reduce the risk of data loss and improve reliability*	0% *(0)*	17% *(1)*	50% *(3)*	33% *(2)*	0% *(0)*
Boundaries					
National	0% *(0)*	0% *(0)*	67% *(4)*	33% *(2)*	0% *(0)*
Organisational *(among organisations of same type or different type e.g. among libraries, between libraries and web archives etc.)*	0% *(0)*	0% *(0)*	50% *(3)*	50% *(3)*	0% *(0)*
Information in focus					
Primary objects *(digital objects)*	0% *(0)*	0% *(0)*	67% *(4)*	17% *(1)*	17% *(1)*
Metadata	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*

Table 14: Results on Section "Purposes for Interoperability"

8.1.2 Barriers

Another important point of our research is to reveal what prevents or hinders the current efforts for interoperability. The analysis from responses led to the identification of several issues that can be possible barriers of interoperability in digital libraries and web archives. Following, we want you to assess whether you acknowledge the statements as barriers or not and evaluate their impact, firstly, on an individual institution/organisation, and, secondly, on the web archiving and digital library community as a whole.

The following have been identified as barriers to interoperability. Please rank their impact from the point of view of **(a) a single institution/organisation** and **(b) the web archiving and digital library community as a whole**.

	Organisation					**Community**				
	Not a barrier	**Mild barrier**	**Moderate barrier**	**Extreme barrier**	*I can't say / don't know*	**Not a barrier**	**Mild barrier**	**Moderate barrier**	**Extreme barrier**	*I can't say / don't know*
Standardisation										
Lack of agreed standards	0% *(0)*	33% *(2)*	17% *(1)*	33% *(2)*	17% *(1)*	0% *(0)*	17% *(1)*	0% *(0)*	50% *(3)*	33% *(2)*
Competition among the current standards	33% *(2)*	0% *(0)*	17% *(1)*	33% *(2)*	17% *(1)*	17% *(1)*	0% *(0)*	17% *(1)*	33% *(2)*	33% *(2)*
Lack of global identifiers	17% *(1)*	17% *(1)*	17% *(1)*	17% *(1)*	33% *(2)*	0% *(0)*	17% *(1)*	33% *(2)*	17% *(1)*	33% *(2)*
Tools & Implementation										
Lack of tools that implement current standards	0% *(0)*	17% *(1)*	33% *(2)*	33% *(2)*	17% *(1)*	0% *(0)*	0% *(0)*	33% *(2)*	33% *(2)*	33% *(2)*
Different implementation of the same standard	0% *(0)*	0% *(0)*	50% *(3)*	50% *(3)*	0% *(0)*	0% *(0)*	0% *(0)*	50% *(3)*	33% *(2)*	17%
Organisational										
Unwillingness of institutions to invest in standardising	17% *(1)*	0% *(0)*	33% *(2)*	50% *(3)*	0% *(0)*	17% *(1)*	0% *(0)*	17% *(1)*	50% *(3)*	17% *(1)*
Unwillingness of organisations to commit in collaboration/dependencies	0% *(0)*	17% *(1)*	33% *(2)*	50% *(3)*	0% *(0)*	0% *(0)*	17% *(1)*	17% *(1)*	50% *(3)*	17% *(1)*
Fear of the expected effort	0% *(0)*	0% *(0)*	50% *(3)*	33% *(2)*	17% *(1)*	0% *(0)*	0% *(0)*	33% *(2)*	33% *(2)*	33% *(2)*
Lack of know-how (in the organisation)	17% *(1)*	0% *(0)*	67% *(4)*	17% *(1)*	0% *(0)*	0% *(0)*	0% *(0)*	50% *(3)*	17% *(1)*	33% *(2)*
Lack of resources (in the organisation)	0% *(0)*	0% *(0)*	0% *(0)*	100% *(6)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*	67% *(4)*	17% *(1)*
Locked systems & no desire for interoperability	17% *(1)*	33% *(2)*	0% *(0)*	50% *(3)*	0% *(0)*	0% *(0)*	17% *(1)*	0% *(0)*	50% *(3)*	33% *(2)*
Legislation										
Limited or forbidden exchange of data outside national borders in some countries	17% *(1)*	0% *(0)*	17% *(1)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	0% *(0)*	67% *(4)*	33% *(2)*
Intellectual property laws	0% *(0)*	17% *(1)*	17% *(1)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*	67% *(4)*	17% *(1)*
Different approaches										
Definition of interoperability based on the targeted systems	0% *(0)*	0% *(0)*	33% *(2)*	17% *(1)*	50% *(3)*	0% *(0)*	0% *(0)*	33% *(2)*	17% *(1)*	50% *(3)*
Different perspectives and priorities between different communities	0% *(0)*	0% *(0)*	83% *(5)*	0% *(0)*	17% *(1)*	0% *(0)*	0% *(0)*	83% *(5)*	0% *(0)*	17% *(1)*

Table 15: Results on Section "barriers to interoperability"

8.1.3 Suggested solutions & Improvements

In the first round the participants were asked for certain changes that could assist the establishment of interoperability. They have been identified and grouped, and are referred to as solutions. Suggested solutions are of high importance and therefore to this part we ask the measure of two aspects: **effectiveness** and **difficulty**.

	Effectiveness					**Difficulty**						
	Not effective / Not a solution	**Somewhat effective**	**Effective**	**Very effective**	I can't say / I don't know	**Very easy**	**Easy**	**Average**	**Difficult**	**Very difficult**	**Not a solution**	I can't say / I don't know
Standardisation												
Consensus on current standards and conformity with them	17% *(1)*	0% *(0)*	17% *(1)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*	33% *(2)*	33% *(2)*	0% *(0)*	17% *(1)*
Initiatives/projects to necessitate the use of current standards	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	83% *(5)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*
Enhancement of current standards	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*
Global & well-defined standards	0% *(0)*	17% *(1)*	50% *(3)*	33% *(2)*	0% *(0)*	0% *(0)*	17% *(1)*	0% *(0)*	33% *(2)*	17% *(1)*	0% *(0)*	33% *(2)*
Development of new standards	0% *(0)*	50% *(3)*	50% *(3)*	0% *(0)*	0% *(0)*	0% *(0)*	33% *(2)*	33% *(2)*	17% *(1)*	0% *(0)*	0% *(0)*	17% *(1)*
Promotion of current and new standards	0% *(0)*	17% *(1)*	50% *(3)*	33% *(2)*	0% *(0)*	0% *(0)*	17% *(1)*	50% *(3)*	0% *(0)*	17% *(1)*	0% *(0)*	17% *(1)*
Tools & Implementations												
Development of tools that implement standards	0% *(0)*	0% *(0)*	67% *(4)*	33% *(2)*	0% *(0)*	0% *(0)*	0% *(0)*	33% *(2)*	50% *(3)*	0% *(0)*	0% *(0)*	17% *(1)*
Common APIs for search & retrieval	0% *(0)*	0% *(0)*	17% *(1)*	83% *(5)*	0% *(0)*	0% *(0)*	0% *(0)*	33% *(2)*	50% *(3)*	0% *(0)*	0% *(0)*	17% *(1)*
Central aggregation service	0% *(0)*	50% *(3)*	17% *(1)*	33% *(2)*	0% *(0)*	0% *(0)*	0% *(0)*	17% *(1)*	17% *(1)*	50% *(3)*	0% *(0)*	17% *(1)*
Knowledge sharing & providing												
Sharing experiences, best practices & successful stories	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	83% *(5)*	17% (1)	0% *(0)*	0% *(0)*	0% *(0)*	0% *(0)*
Consensus on best practices	0% *(0)*	17% *(1)*	17% *(1)*	67% *(4)*	0% *(0)*	0% *(0)*	17% *(1)*	33% *(2)*	0% *(0)*	33% *(2)*	0% *(0)*	17% *(1)*
Clear definitions & terminology about digital preservation	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	17% *(1)*	50% *(3)*	33% *(2)*	0% *(0)*	0% *(0)*	0% *(0)*
Foundation of central organisation that provides support for technical & legal issues	17% (1)	33% *(2)*	33% *(2)*	17% *(1)*	0% *(0)*	0% *(0)*	0% *(0)*	0% *(0)*	33% *(2)*	33% *(2)*	0% *(0)*	33% *(2)*
Legislation												
Clear legislation & policies for the exchange of data/metadata	0% *(0)*	0% *(0)*	33% *(2)*	50% *(3)*	17% *(1)*	0% *(0)*	17% *(1)*	0% *(0)*	17% *(1)*	33% *(2)*	0% *(0)*	33% *(2)*
Approaches												
Define interoperability from a Web Infrastructure perspective instead of a system-to-system perspective	0% *(0)*	0% *(0)*	50% *(3)*	33% *(2)*	17% *(1)*	0% *(0)*	17% *(1)*	33% *(2)*	0% *(0)*	0% *(0)*	0% *(0)*	50% *(3)*
Communities & people												
Better collaboration and stronger involvement of related communities to each other's activities to ensure everyone's needs are considered	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	17% *(1)*	67% *(4)*	0% *(0)*	17% *(1)*	0% *(0)*	0% *(0)*
Involvement of people with broader knowledge/experience, not individually confined to community aspects.	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	33% *(2)*	67% *(4)*	0% *(0)*	0% *(0)*	0% *(0)*	0% *(0)*

Table 16: Results on Section "Suggested Solutions & improvements for interoperability"

8.1.4 Future challenges

Trying to look further in the future and all the upcoming changes that are likely to happen, we asked participants to describe what they consider as future challenges for the interoperability. We included the answers that were related to:

- future changes (problems that will appear in the future or already exist but are likely to be magnified in the future)
- challenging aims that need to be also considered and achieved in the future as next steps.

The following have been identified as future challenges for interoperability. Please rank their urgency.

	Not a priority	Low priority	Medium Priority	High priority	Not a challenge
Achieving interoperability of content. To consider digital libraries and web archives also as big datasets that should interoperate not only in terms of URIs and metadata but also in terms of content.	0% *(0)*	0% *(0)*	17% *(1)*	83% *(5)*	0% *(0)*
New players are emerging in the field of web archiving. Therefore different systems, needs, and tools are emerging, the involved communities are increased and, as a result, interoperability may become more complex goal/affair.	0% *(0)*	17% *(1)*	50% *(3)*	33% *(2)*	0% *(0)*
Explosion of the volume of web data to archive. As a result of the combination of the increasing efforts to archive as much of the web as possible and the immense growth of the web.	0% *(0)*	0% *(0)*	50% *(3)*	50% *(3)*	0% *(0)*
The web resources become more and more complex. New and complex media and web resources (Web 2.0, Social Media, etc.) demand enhanced methods for web preservation.	0% *(0)*	0% *(0)*	0% *(0)*	100% *(6)*	0% *(0)*

Table 17: Results on Section "Further Challenges"

8.1.5 Interoperability Perspectives / Approaches

Through the analysis of all participants' responses, we identified additional interesting insights for further discussion. One of these, was the perception of two different perspectives to consider interoperability:

> ***System Interoperability****: The interoperability of systems is considered as the possibility of two or more systems to communicate and is defined based on which systems need to be interoperable.*
>
> ***Information Interoperability****: Interoperability of information is related to the information structure of the Web. Therefore, it is based on making the information itself (data/metadata/identifiers) usable in different environments, regardless of the compatibility between the environments*.

Please evaluate the following statements.

	Strongly disagree	**Disagree**	**Agree**	**Strongly agree**	I can't say / I don't know
System Interoperability is a valid and reasonable way to establish interoperability in the context of digital libraries and web archives	0% *(0)*	17% *(1)*	67% (4)	17% *(1)*	0% *(0)*
Information Interoperability is a valid and reasonable way to establish interoperability in the context of digital libraries and web archives	0% *(0)*	0% *(0)*	50% *(3)*	50% *(3)*	0% *(0)*
The benefits of System Interoperability are limited to the targeted systems *(i.e. the systems based on which interoperability was decided & designed)*	0% *(0)*	17% *(1)*	50% *(3)*	17% *(1)*	17% *(1)*
Information Interoperability is a broadly beneficial investment that involves various information environments	0% *(0)*	0% *(0)*	50% *(3)*	50% *(3)*	0% *(0)*

Table 18: results on section "Interoperability perspectives" (1)

	Much more advanced	**More advanced**	**Same**	**Less advanced**	**Much less advanced**	I can't say / I don't know
System Interoperability provides .... levels of interoperability among two or more interoperating systems than Information Interoperability	17% *(1)*	17% *(1)*	0% *(0)*	17% *(1)*	0% *(0)*	50% *(3)*
	Much more difficult	**Somewhat more difficult**	**Equally difficult**	**Somewhat easier**	**Much easier**	
Information Interoperability is ... to achieve than System Interoperability	0% *(0)*	33% *(2)*	0% *(0)*	50% *(3)*	0% *(0)*	17% *(1)*
	Much more sustainable	**More sustainable**	**Equally sustainable**	**Less sustainable**	**Much less sustainable**	
Information Interoperability is ... than System Interoperability	0% *(0)*	50% *(3)*	33% *(2)*	0% *(0)*	0% *(0)*	17% *(1)*

Table 19: results on section "Interoperability perspectives" (2)