Interoperability of web archives and digital libraries: A Delphi study

Hendrik Kalb
Institute for Business
Informatics
Technische Universität Berlin
Berlin, Germany
hendrik.kalb@tu-berlin.de

Paraskevi Lazaridou
Institute for Business
Informatics
Technische Universität Berlin
Berlin, Germany
paraskevi.lazaridou@tu-berlin.de

Edward Pinsent
Academic & Research
Technologies
University of London
Computer Centre
London, UK
edward.pinsent@london.ac.uk

Matthias Trier
Department of IT Management
Copenhagen Business School
Copenhagen, Denmark
mt.itm@cbs.dk

ABSTRACT

The interoperability of web archives and digital libraries is crucial to avoid silos of preserved data and content. While various researches focus on specific facets of the challenge to interoperate, there is a lack of empirical work about the overall situation of actual challenges. We conduct a Delphi study to survey and reveal the insights of experts in the field. Results of our study are presented in this paper to enhance further research and development efforts for interoperability.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]; D.2.12 [**Interoperability**]

General Terms

Study, Theory

Keywords

Interoperability, Web Archives, Digital Libraries, Delphi Study

1. INTRODUCTION

This paper reveals the results of a Delphi study about the interoperability of web archives and digital libraries. The aim of the study is to provide a better understanding about crucial aspects of interoperability in this domain.

According to the Institute of Electrical and Electronics Engineers (IEEE), interoperability is the "ability of two or more systems or components to exchange information and to use the information that has been exchanged" [7, p. 114]. It has numerous facets including uniform naming, metadata formats, document models, and access protocols [16]. Interoperability in a narrow sense describes how technical systems interoperate. In a broader sense, it also comprises social, political, and organisational factors [9].

Research about interoperability of web archives and digital libraries often considers technical and semantic aspects of interoperability, e.g. protocols and standards. However, conceptual models for interoperability also comprise other aspects. The European Interoperability Framework differentiates between organisational, semantic, and technical interoperability [12]. This perspective has been adopted by the DL.org to describe and consider interoperability of digital libraries [4]. Other publications mention the semantic level under the term content level [3]. A similar perspective distinguishes between technical / basic, syntactic, functional / pragmatic, and semantic as levels with increasing abstraction [10]. Similarly, [25] describes with an increasing capability for interoperation the levels of technical, syntactic, semantic, pragmatic, dynamic, and conceptual interoperability. A specific perspective for digital libraries distinguishes the levels of gathering, harvesting, and federation [3]. The maturity of inter-organisational interoperability can be assessed on the levels of computer, process, knowledge, value, and goal interoperability [9]. While the different models indicate interoperability as a subject with various facets, only a few studies can be found that inquire into current web archives and digital libraries for interoperability issues (see Section 2). Thus, there is a risk that research and development to improve interoperability is mainly driven by personal assumptions, beliefs, or experiences of the researcher, and therefore fails to address the real needs of the community.

Our survey aims to gain insight into areas that have not been surveyed and derives from people who are highly involved and have personal experience. Our aim is to examine a theoretical framework of interoperability in both web archives and digital libraries with the assistance of people who have their own experiences and thoughts on the

topic. The survey does not focus on specific technical developments and offers to the participant the opportunity to develop freely their thoughts. This study can be 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 to the involved communities, is the sharing of the valuable outcome of an enlightening virtual discussion from experts about interoperability.

The remainder of this paper is organised as follows: Section 2 reviews related studies about web archives and digital libraries, and focuses on their relation to interoperability aspects. In section 3, we reveal the chosen method for our study including a short introduction into the Delphi method in general. The first and second round's results of our study are presented in section 4 and 5 before we discuss them in section 6. Finally, we draw our conclusions in section 7.

2. RELATED WORK

In this section, we review published surveys on web archives and digital libraries regarding their insights into interoperability aspects.

Several surveys have been conducted in the domains of web archiving and digital libraries addressing issues regarding the web archiving activities. The majority put their focus on web archiving in general, examining the progress that has been made, the maturity so far, the problems encountered and the practices used in the field of web archiving. However, interoperability was out of their scope. Aspects such as legislation and national legal authorities, permission access, common tools and standards and cooperation of institutions for common developments that are also addressed, are of course related to our topic of research but not in a direct way. In particular, the International Internet Preservation Consortium (IIPC) carried out a survey among its members, basically a profile identification, and got responses from 35 of them. The survey examined the maturity of web archiving, the scope, the tools used for harvesting, curation and access, legal limitations by their countries and access restrictions [11]. Another survey on 16 national libraries focuses on how they attempt to justify their web archiving activities [23]. The Internet Memory Foundation ran a survey on European institutions aiming to obtain a clearer understanding of problems encountered in the field of Internet archiving with the help of a panel composed of 74 participants from national libraries, audiovisual and institutional archives. This survey addressed several aspects such as the status of web archiving, legal aspects, access restrictions, policies and priorities regarding the scope and the types of archiving [1]. The 18th Conference of Directors of National Libraries in Asia and Oceania (CDNLAO) presented a report with the participants' answers about web archiving in this region. The questions were about cooperation, access and preservation policies, tools in use, and the legal framework [20]. Later, a survey 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 [8].

However, to the best of our knowledge, only a few surveys put their focus specifically on interoperability issues. A study regarding the future interoperability in web archiving was presented by [14] as a survey on national libraries. In this study, 37 participants responded to questions regarding 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 they could partner with to solve such problems. The motivation of this survey, and also one of the questions, was the belief that interoperability between all national web archives should be a long-range goal, and the majority agreed on that. Based on the results, it is clear that a great challenge and need for the national libraries is to make legal deposit, copyright and related legislation adapted to the world of the Internet so that the digital part of national heritage can be preserved for future generations. Additionally, some comments of participants regarding preferable engagement in partnerships revealed a need for cooperation with institutions that have to offer technical and collection expertise along with a commitment to preservation issues [14].

Another survey, that focused exclusively on interoperability aspects, but specifically in the digital library sphere, was conducted by the DL.org Policy Working Group [13]. This experimental survey on policy interoperability of digital libraries was carried out among a selected sample of digital libraries, digital repositories and federated services, and received 26 responses (15 completed). This survey addressed how the policies, strategies, frameworks and plans of the digital libraries affect or are affected by interoperability. Their findings revealed that existing policies of the organisations have been revised according to those of other organisations with regard to policy exchange and reuse only in the areas of Preservation, Access, Collection Development and Metadata. Furthermore, even if respondents expressed interest to interoperate with other public or private organisations, just few of the stated policies regulate such interactions. The authors identified in the results an indication for approaching policy interoperability not only from a technical but also from an organisational and semantic perspective [13]. Within the same project, another survey [26] was run regarding quality interoperability in digital libraries, since quality and interoperability are two aspects that affect each other. The results revealed to what extent the respondents use validation tools to check compliance of metadata, format or communication protocols and how complete they consider their metadata is. They also identified some barriers to metadata creation, like the complicated and contradictory guidelines. According to this survey, most of the respondents consider interoperability as a mainly technical issue.

3. METHOD

In this section, we outline the underlying method of our research. We aim to identify 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 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.

The Delphi Method grew out of the need for a technique able to obtain the most reliable consensus of a group of experts [21]. 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 [17]. For example, several Delphi studies are ranking-type and aim to extract a consensus opinion on the importance of specific issues, but others emphasise differences of opinion in order to develop a set of alternative future scenarios [21].

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 [18]. The responses are then collected and analysed. Based on the analysed results, a 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. The Delphi method is an iterative process and each subsequent questionnaire is developed based on the results of the previous questionnaire. The number of the 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 [24]. However, the participants are usually given at least one opportunity to revise their original answers upon examination of the group responses [18].

The Delphi study in our research consists of two rounds. A purposive sample of seven international experts from the web archiving and digital library communities was created. While the research team knew 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.

The aim of the first round was a brainstorming about the purposes, obstacles, possible solutions to overcome limitations, and other future challenges. Therefore, a questionnaire was created with four open questions (see Appendix A). Two researchers created the questions before an archivist reviewed them 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 text 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 are presented in section 4 and were used to design the second round.

The 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 and the possibility for further comments was created. The questions were created by two researchers according to the structure of the first round's results, and reviewed

afterwards by the archivist. 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. The questionnaire was sent to the participants at the beginning of April, and a reminder was sent after two weeks to participants that had not responded yet. The second round was completed by six of the seven participants. The responses were analysed and the results are summarised in section 5.

4. RESULTS FROM THE FIRST ROUND

In the following, we present the results of the first round of our Delphi study. We structured the results into categories represented by the following subsections.

4.1 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. However, each purpose may imply some specific requirements or a different context.

The identified purposes are further separated in three aspects. The first aspects describes the distinct uses for which interoperability is necessary. In this way, uses that motivate interoperability can be differentiated between (a) federated search, (b) federated access, (c) exchange, and (d) 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. In the traditional library, for example, it enables the user to search various printed and electronic collections through one interface [6]. The search query that the user types in a single interface is sent to multiple search engines. In this way, it is common that a selection or subset of search engines is generated instead of broadcasting the query to all search engines. The typical phases of federated search are resource representation, resource ranking, distributed search, and result merging [5]. 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 requires that just the location of the desired objects can be found even if it is distributed in distinct archives, **federated access** also enables the user to retrieve the data directly from a single point. This means that the data can be, for example, viewed or downloaded. We distinguish between federated search and federated access in order to emphasise 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 but describe different aims for the transfer of data between archives. The **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. One participant reported:

"collaborative constitution of collections or exchange of collections between institutions. For example, constitution of web archives collections for the 2012 Olympic games in London (IIPC project)."[1]

Replication on the other hand aims at data redundancy in order to reduce the risk of data loss and improve reliability. The preservation of digital information has to consider physical threats (e.g. natural events, age of the hardware), technological threats (e.g. format obsolescence), human threats (e.g. curational errors), and institutional threats (e.g. economic failure). Replication combined with regular auditing can help to reduce the impact of these threats [2]. While the specific reasons for replication were not further explained by the participants, the need for replication was mentioned in statements like the following:

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

The second aspect derives from the differentiation in the scopes of the above uses. Hence, it can also be understood as a specialisation of the purposes already described. In particular, the following refinements were made about inter-operation 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 aspect 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".

[1]For more information about the Olympics 2012 collection see also http://digital2.library.unt.edu/nomination/olympics2012/

4.2 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 latter 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 by 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 organisations 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 [14] where respondents indicated a 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 digital collections since their content is described and encoded in common standards. This particular point has been also investigated in [19] which examined digital longevity through standards and reached the conclusion that specific kinds of standards, even if not designed for digital longevity, are essential to this purpose to describe the functionality, the procedures and the concepts of a digital library or archive, to preserve the digital documents, to preserve the access to the content (metadata standards), and for interoperability.

4.3 Barriers to Interoperability

The second aspect of interoperability that we aimed to identify is the obstacles and limitations, or in other words, the barriers that hinder the establishment of interoperability. We grouped the identified barriers in five categories: (a) standardisation, (b) tools and implementation, (c) organisational obstacles, (d) legal problems, and (e) the approach to handle interoperability.

While various standards already exist, the current state of **standardisation** and compliance seems to be unsatisfactory. A lack of agreed standards has been reported. Similar to the lack of agreement, competition among the already existing standards has been reported.

However, even the agreements on standards do often 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 implemented differently in different contexts. More specifically, even if two archives apply the same schema, the content can be modelled differently and thus impede interoperability:

"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 a 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 standardising processes:

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

Furthermore, organisations may feel not able to provide or invest in 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 actually have 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 [14] and was later addressed by the same author in detail [15]. Apart from this, the copyright holders define significantly the level of access and intellectual property laws hinder an open or public 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 between 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."

4.4 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."

Standardisation: 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 would 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 also said to play 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 organisation 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."

4.5 Interoperability perspectives

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

System Interoperability (or system-to-system interoperability) that 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."

4.6 Further challenges

Part of our research was to examine interoperability with a view in 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 future challenges have been identified. 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."

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. For example:

"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)."

5. RESULTS FROM THE SECOND ROUND

In the following, we summarise the results of the second round of our Delphi study. In this 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. Due to constraints of questionnaire research, we focussed the second round on the five core aspects: purposes, barriers, suggested solutions, further challenges, and perspectives of considering and realising interoperability in web archives and digital libraries.

We asked for agreement regarding the identified **purposes** on a four point Likert scale from "Strongly disagree" to "Strongly agree". Each of the identified purposes was agreed by at least five of the six participants. Two times, a participant answered with "I can't say", and one participant

disagreed on replication as a purpose. In summary, we assess the purposes as verified. Minor trends can be identified in the differences of strong and normal agreement. Federated access and federated search got stronger agreement than exchange and replication. Also the interoperability of metadata got stronger agreement than primary objects.

The **barriers** were evaluated with four point Likert scale from "Not a barrier" to "Extreme barrier". Additionally, the participants were asked to evaluate them separately from the point of view of an individual organisation (e.g. a single library) and of the community as a whole. 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. Furthermore, 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*. The strongest barriers on the community level are *limited or forbidden exchange of data outside national borders*, *lack of resources (in the organisation)* and *intellectual property laws*.

Next to verifying the **suggested solutions** to overcome existing barriers, they should also be assessed regarding their efficiency in order to deduce recommendations which solutions may be prioritised. Efficiency consists of the ratio of the impact of the solution to the effort to realise the solution. The impact was measured through the evaluation of effectiveness on a four point Likert scale from "Not effective / Not a solution" to "Very effective". The effort was measured through difficulty on a five point Likert scale from "Very easy" to "Very difficult". All the identified suggestions for solutions were verified. One participant rated *consensus on current standards and conformity* and *foundation of central organisation that provides support for technical & legal issues* with "Not effective / Not a solution" but did not assess it as "Not a solution" in the difficulty measure. Therefore, we deduce that he assessed the solution as not effective but verified it as a generally possible solution. In order to provide recommendations about the identified solutions, the average effectiveness and average difficulty for each solution were calculated and plotted in a portfolio (see figure 1). Three clusters can be identified:

1. **Highly recommended solutions**: A line was drawn from "somewhat effective" and "very easy" to "very effective" and "difficult". Solutions in the sector above the line are considered as very efficient because their estimated effectiveness is higher than the required effort to accomplish. The most promising solution is
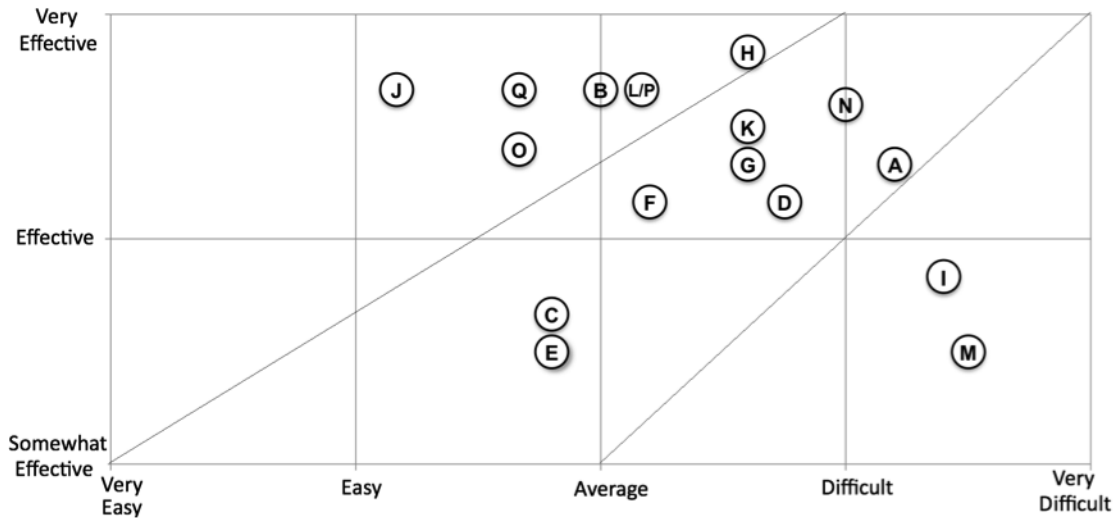
Figure 1: Portfolio of suggested solutions: (A) Consensus on current standards and conformity with them; (B) Initiatives / projects to necessitate the use of current standards; (C) Enhancement of current standards; (D) Global & well-defined standards; (E) Development of new standards; (F) Promotion of current and 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 and stronger involvement of related communities to each other's activities; (Q) Involvement of people with broader knowledge / experience, not individually confined to community aspects

thereby *sharing experiences, best practices & successful stories.*

2. **Recommended solutions**: A line was drawn from "Somewhat effective" and "average difficult" to "very effective" and "very difficult". Solutions that lie above this line and below the sector of highly recommended solutions 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**: Solutions in the third sector can not be assessed as efficient because their effectiveness is much lower than the probable 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.

Further challenges for the future were evaluated on a four point Likert scale from "Not a priority" to "High priority" and the alternative option of "Not a challenge". Each challenge was rated at least with "Low priority" by all participants, and, thus, the four identified challenges were verified. The average priority of each challenge was above medium priority. The increasing complexity of web resources was considered as the most pressing challenge.

The last part of the second round's questionnaire aimed at a comparative evaluation of system-to-system interoper-ability and information interoperability. None of the participants questioned the general applicability of the perspectives, and, therefore, it can be considered as verified by the participants that both are possible ways to establish interoperability. However, the answers to the comparative part were quite heterogeneous, and do not allow the identification of a clear trend.

6. DISCUSSION

In this section, we discuss the results of our Delphi study. As a first result 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 several threats [2]. 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, e.g. link analysis [22], as well as any interoperation demands for the ingest of new digital content into a web archive or digital library has not appeared in the participants' statements.

Additionally, we identified several benefits that are con-

nected 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 solutions are connected by nature because a solution (or improvement) addresses 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 from the beginning of this paper, 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. Thereby, 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 focuses on legal issues is not mentioned explicitly in the presented 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 [4]. 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 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 together 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.

7. CONCLUSION

The Delphi study, presented in this paper, revealed insights regarding current problems, limitations, needs and challenges that are encountered in today's interoperations (or efforts in this direction) among systems of the web archiving and digital library communities. The study was 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. It offered a unique contribution to the research field of interoperability, presenting the current barriers but also suggestions for future approaches, and can be a useful study for the communities of web archiving, digital libraries, and digital preservation.

However, a limitation has to be taken into consideration. The findings are influenced by the selection of experts. Therefore, the same questions may lead to different results with other experts. However, we did not aim on completeness, and we consider it unlikely that such results would be conflicting.

Finally, it should be emphasised again that further studies should be conducted in order to validate and to extend the understanding of current and future interoperability aspects for web archives and digital libraries.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Web Archiving in Europe. Technical report, Internet Memory Foundation, 2010.

[2] M. Altman, M. O. Adams, J. Crabtree, D. Donakowski, M. Maynard, A. Pienta, and C. H. Young. Digital Preservation through Archival Collaboration: The Data Preservation Alliance for the Social Sciences. *The American Archivist*, 72(1):170–184, 2009.

[3] W. Y. Arms, D. Hillmann, C. Lagoze, D. Krafft, R. Marisa, J. Saylor, C. Terrizzi, and H. Van de Sompel. A Spectrum of Interoperability. *D-Lib Magazine*, 8(1), Jan. 2002.

[4] G. Athanasopoulos, L. Candela, D. Castelli, K. El Raheb, P. Innocenti, Y. Ioannidis, A. Katifori, A. Nika, S. a. Ross, A. Tani, C. Thanos, E. Toli, and G. Vullo. Digital Library Technology & Methodology Cookbook: Interoperability Framework, Best Practices & Solutions. Technical report, 2011.

[5] T. T. Avrahami, L. Yau, L. Si, and J. Callan. The FedLemur project: Federated search in the real world. *Journal of the American Society for Information Science and Technology*, 57(3):347–358, 2006.

[6] A. Curtis and D. G. Dorner. Why Federated Search? *Knowledge Quest*, 33(3):35–37, 2005.

---

[2]http://blogforever.eu/

[7] A. Geraci, K. Freny, L. McMonegal, B. Meyer, J. Lane, P. Wilson, J. Radatz, M. Yee, H. Porteous, and F. Springsteel. *IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries*. IEEE Press, Piscataway, NJ, USA, 1991.

[8] D. Gomes, J. Miranda, and M. Costa. A survey on web archiving initiatives. In S. Gradmann, F. Borri, C. Meghini, and H. Schuldt, editors, *Research and Advanced Technology for Digital Libraries*, volume 6966 of *Lecture Notes in Computer Science*, pages 408–420. Springer Berlin / Heidelberg, 2011.

[9] P. Gottschalk. Maturity levels for interoperability in digital government. *Government Information Quarterly*, 26(1):75–81, Jan. 2009.

[10] S. Gradmann. Interoperability. A key concept for large scale, persistent digitallibraries. Technical report, 2007.

[11] A. Grotke. International Internet Preservation Consortium: 2008 Member Profile Survey Results. Technical report, 2008.

[12] IDABC. European Interoperability Framework for pan-European eGovernment Services. Technical report, Luxembourg, 2004.

[13] P. Innocenti, M. Smith, K. Ashley, S. Ross, A. De Robbio, H. Pfeiffenberger, and J. Faundeen. Towards a Holistic Approach to Policy Interoperability. *The International Journal of Digital Curation*, 6(1):111–124, 2011.

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

[15] G. Jacobsen. Web Archiving: Issues and Problems in Collection Building and Access. *LIBER Quarterly*, 18(3/4):366–376, 2008.

[16] C. Lagoze, H. Van de Sompel, M. Nelson, S. Warner, R. Sanderson, and P. Johnston. A Web-based resource model for scholarship 2.0: object reuse & exchange. *Concurrency and Computation: Practice and Experience*, 24(18):2221–2240, June 2010.

[17] J. Landeta. Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, 73(5):467–482, June 2006.

[18] H. A. Linstone and M. Turoff. *The Delphi Method: Techniques and Applications*. Addison Wesley, 2002.

[19] H. H. J. Lorist and K. v. d. Meer. Standards for Digital Libraries and Archives: Digital Longevity. In *NDDL '01 Proceedings of the 1st International Workshop on New Developments in Digital Libraries: n conjunction with ICEIS 2001*, pages 89–98. ICEIS Press, 2001.

[20] National Diet Library. CDNLAO Questionnaire Survey on Web Archiving. Technical report, 2010.

[21] C. Okoli and S. D. Pawlowski. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42(1):15–29, Dec. 2004.

[22] E. Reynolds. Web Archiving Use Cases. Technical report, 2013.

[23] R. Shiozaki and T. Eisenschitz. Role and justification of web archiving by national libraries: A questionnaire survey. *Journal of Librarianship and Information Science*, 41(2):90–107, 2009.

[24] G. J. Skulmoski, F. T. Hartman, and J. Krahn. The Delphi Method for Graduate Research. *Journal of Information Technology Education*, 6, 2007.

[25] A. Tolk, S. Y. Diallo, and C. D. Turnitsa. 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*, 5(5):65–74, 2007.

[26] G. Vullo, G. Clavel, N. Ferro, S. Higgins, R. van Horik, W. Horstmann, and S. Kapidakis. Quality interoperability within digital libraries: the DL.org perspective. In *2nd DL.org Workshop in conjunction with ECDL 2010*, Glasgow, UK, 2010.

APPENDIX

A. FIRST ROUND'S QUESTIONS

All questions of the first round were formulated as open questions.

- 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.

- What are the main obstacles and limitations that prevent or hinder interoperability?
  (technical, political, organizational, management, legislation or other barriers)

- 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.

- What do you consider as future challenges regarding interoperability 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.