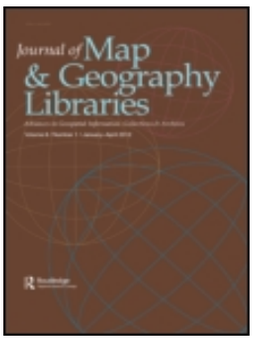

Routledge
Taylor & Francis Group

Journal of Map & Geography Libraries

Advances in Geospatial Information, Collections & Archives

ISSN: 1542-0353 (Print) 1542-0361 (Online) Journal homepage: http://www.tandfonline.com/loi/wmgl20

Starting Points for Lowering the Barrier to Spatial Data Preservation

Anita E. Locher

To cite this article: Anita E. Locher (2016) Starting Points for Lowering the Barrier to Spatial Data Preservation, Journal of Map & Geography Libraries, 12:1, 28-51, DOI: 10.1080/15420353.2015.1080781

To link to this article: http://dx.doi.org/10.1080/15420353.2015.1080781

Published online: 15 Mar 2016.

Submit your article to this journal

Article views: 55

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=wmgl20

Download by: [JH Libraries] **Date:** 09 September 2016, At: 10:48

Journal of Map & Geography Libraries, 12:28–51, 2016
Published with license by Taylor & Francis
ISSN: 1542-0353 print / 1542-0361 online
DOI: 10.1080/15420353.2015.1080781

Routledge
Taylor & Francis Group

Starting Points for Lowering the Barrier to Spatial Data Preservation

ANITA E. LOCHER
University of Zürich, Zürich, Switzerland, and the University of Barcelona, Barcelona, Spain

There is general agreement that spatial data adds particular difficulties to digital preservation due to, for example, the complexity of data models and semantics specific to individual thematic areas. However, there is a lack of literature providing an overview of the challenges and analyzing in particular the effort required to surmount these in combination with the potential added value gained through digital preservation.

The Delphi method was used to evaluate obstacles to archiving geographic vector and raster data serving as a basis for topographic base map creation, seen through the lens of data producers, providers and guardians. Two international Delphi groups were questioned on developments regarding geodata, and their influences on access and preservation.

The mentioned handicaps to preservation were of financial, managerial, legal, and technological in nature. The latter have a higher probability to be surmounted within at least 10 years than non-technological. The study shows that the lack of standardization and the use of proprietary formats is still a central problem. Furthermore, the consciousness about the value of geographic assets is considered most likely to rise early. As a good starting point for improving archiving of spatial data, we also suggest the controlled disposal of superfluous data as a measure to reduce cost.

KEYWORDS *digital preservation, spatial data, archiving, geodata, government data*

© Anita E. Locher
Address correspondence to Anita E. Locher, University of Zürich, Rämistrasse 71, Zürich, 8006 Switzerland. E-mail: anita.locher@geo.uzh.ch

INTRODUCTION

Starting from the assumption that digital legacy geodata can have permanent value, this article seeks to identify the challenges related to their preservation. Stakeholders such as archives, libraries, or the data creators are key actors in their preservation. However, the environment in which these institutions perform has changed significantly, since geographic information is now mainly produced in digital form, and processes must be adapted. Currently, attempts to preserve geodata are not based on standard procedures, and there is a need for improvements. In this respect, for an institution taking over responsibility for archiving in the long term, it is helpful to know the challenges of preserving digital legacy geodata and to compare them with recent trends that favor archiving and suggest ways of reducing challenges.

WHY PRESERVE GEODATA

The main argument for long term preservation in geosciences is longitudinal research (Sweetkind, Larsgaard, and Erwin 2006; Moran et al. 2009). This becomes evident searching geosciences journals especially in environmental science where researchers regularly use decades old data to explore processes involving change (Conway et al. 2013; Kennedy et al. 2008; Cushing et al. 2008; Leyk, Boesch, and Weibel 2006). Thus not only climate change research, but also any other kind of environmental change analysis can benefit from long term data preservation (Beruti et al. 2010; Erwin and Sweetkind-Singer 2009; Erwin, Sweetkind-Singer, and Larsgaard 2009; Harris 2001; Janée 2008; National Academy of Sciences. National Research Council and Committee on the Preservation of Geoscience Data and Collections 2002; Shaon and Woolf 2011; European Commission. High Level Expert Group on Scientific Data 2010). Understanding change on our planet can help predict natural events, even disasters (Beruti et al. 2010) and assist governments in better managing natural and human resources and planning. The digital nature of geodata adds under the right technological circumstances immediate utility to their historic value. This immediateness opens up the door for research questions that were previously difficult or impossible to answer for financial and timely reasons. Furthermore the interest of situating information in space makes geodata valuable. As more and more scientific domains recognize this, demand for geodata and legacy geodata is increasing (Beruti et al. 2010). The use of legacy geodata is not measurable yet, but its potential lets memory institutions hope that the value exceeds the cost of preservation. Figure 1 depicts the life cycle of data management, and shows that re-use is an integer part of the cycle which should be included in the global consideration of preservation challenges and benefits.

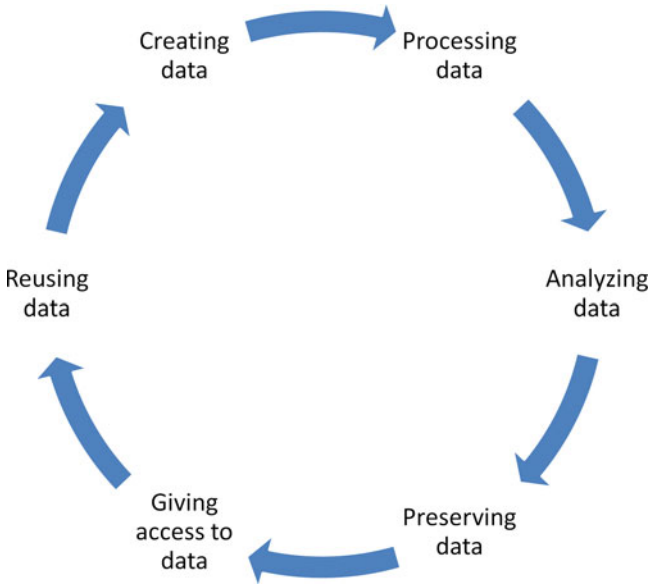

FIGURE 1 Data management life cycle adapted from UK Data Archive (http://www.lancaster.ac.uk/library/rdm/plan/data-lifecycle/).

PROBLEM STATEMENT

Any institution faces many challenges when taking on the role of a long term archive of geospatial data. State, national, and federal archives and libraries as well as data producers have taken responsibility for providing sustainable access to spatial information. The practical experience allows them to report their main problems such as keeping up with the frequency of database updates, data quantity, interdependence of data layers with base maps, semantics, quality of metadata at ingest, legal constraints, etc. (Bos et al. 2010; Morris, Tuttle, and Essic 2009; Fleet 1999; Geospatial Multistate Archive and Preservation Partnership [GeoMAPP] 2011). Although reports of these projects are valuable, they are mostly pragmatic and do not provide deep insights into the process by which the results were achieved or the decisions taken. Available literature about digital preservation challenges in general does not take account of the specificity of geodata or mentions only part of the challenges. There is a lack of literature giving an overview of the challenges that legacy geodata implies and, in particular, the effort required to surmount these in combination with the potential added value gained through digital preservation. Considering that ideal solutions requires a large amount of effort, time, and money; institutions that start a preservation process, in times of restricted budgets, might want to know where to start and which action is more sustainable within their financial or timely context.

The overall goal of this study is therefore to help institutions starting or revising a long term preservation process of geospatial data to identify which are the easiest and most effective measures they can take to improve the state of their process. We do this by distinguishing major barriers to archiving but also identifying positive drivers to sustainable access. Specifically, we want to know as follows:

1. What is the nature of the most burning issues with respect to preservation of geospatial data that we should focus on in the next years?
2. Which measures for improving preservation are easier to implement?
3. Does responsibility for archiving geodata lie with memory institutions, producers of geodata or other stakeholders? Are the most important challenges only to be addressed by a global community or can individual memory institutions have an impact?
4. Which issues have not been reported on or researched related to geodata preservation? Are there new/other challenges we have to be aware of?

This explorative research brings up issues and analyzes their critical impact; nevertheless, it does not give detailed answers to how to surmount them. In this study, we concentrate on geographic data that is used for the production of topographic maps.

RELATED WORK

We want to shed some light on the challenges reported so far. Going through the literature will allow us later on to find new issues in legacy geodata preservation. As the whole data management cycle influences preservation, there is no clear distinction between preservation issues and data management issues (see Figure 1). Nevertheless, some good practice for data management is central when it comes to long term access such as format migration, storage media refreshment, and quality metadata. An early technical evaluation of the issues regarding geodata was reported by the San Diego Supercomputer Center (Zaslavsky 2001). It questions how to maintain the relationship between topological objects in vector data and data sets of different scales covering the same area when migrating or standardizing. The proposed solutions are XML-based file formats and the use of standard metadata. One year later, Bleakly (2002) addressed the issues of geospatial digital archiving in the so-called SAND-report based on the knowledge gained in other domains. She described the different strategies of preservation and emulation generally used for any kind of digital content. Nevertheless, she recognizes the dominance of proprietary software, the variety of file formats, and the amount of data produced as specific challenges for geographic data (Bleakly 2002). Morris (2006) re-

turned to the problem of a lack of standards in vector file formats and addressed problems in content packaging. Lack of historic vector data due to overwriting by updates was a major issue for the GeoConnections Policy Advisory Node Working Group on Archiving and Preserving Geospatial Data (2005). This Canadian study gives a series of recommendations including maintenance of a version-capable database or taking snapshots. In 2009, Mcgarva et al. criticized the lack of versioning in GI systems or their incapacity to do so (Mcgarva, Morris, and Janée 2009). Nevertheless, the versioning capabilities of Geographic Information Systems (GIS) and awareness of the utility of using them has expanded in the last six years. Some challenges are more distinct to individual data sets such as the problem of the sheer amount of data when it comes to satellite images as reported by the European Space Agency (ESA) (Beruti et al. 2010). Nevertheless storage space can be a problem also for smaller institutions if ever increasing resolution of aerial photography and other sensor data is involved. On the nontechnological side, Fleet analyzed legal difficulties for third parties giving access to archived geospatial data or even receiving copies (Fleet 1999; 2003). In the context of the UK, paying contractual agreements between creator and deposit institutions could be made in order for the national library to continue offering access to legacy geodata and for the academic community to access it. While in Switzerland, for example, a new law triggered the archiving process (Ellipse 2013) and treats geodata at the same level as other government records to be transferred to the national archive. Nevertheless, the producing agent in Switzerland is still responsible for providing medium term continued access. Legal issues were also discussed in Canada where geospatial databases were examined from the perspective of a public archives obligation to maintain integrity and authenticity of records (Dingwall et al. 2005). In a study in the same year, GeoConnection came to the conclusion that policy was a main issue and suggested that building a business case for geodata products would be a good approach to face the lack of institutional policy on preservation. Building a business case was also a task of the GeoMAPP project and recommended in the Technology Watch Report (Mcgarva, Morris, and Janée 2009). Another suggestion was to create "authoritative responsibility centers that empower individuals with the ability to define and apply the information management principles" (GeoConnections Policy Advisory Node. Working Group on Archiving and Preserving Geospatial Data 2005). Steps to reach this goal include creating evaluation principles, standardization of formats, and metadata. National Oceanic and Atmospheric Administration (NOAA) gives a hint on the importance of semantics for science data and expresses the wish that communities would develop their respective standards which would merge in the following time to a general earth science standard. In 2008, semantics appeared in a list of problems to address associated with metadata (North Carolina Geographic Information Coordinating Council 2008). All these experiences are useful

sources for exploring challenges in preserving geodata and some potential solutions. Nevertheless, as the context of the studies varies widely, a comparison of the difficulty or the ease of surmounting the challenges is not possible. Taking experiences of accomplished preservation projects, we cannot know if all of the issues were reported or if we missed some challenges. This is why our research was conducted. Providing a unique context based on expert opinion, we want to raise issues and compare and analyze the challenges.

METHOD

Inspired by a study that explored interoperability issues for web archives by Kalb et al. (Kalb, Pinsent, and Trier 2013), we used the Delphi method to identify and classify trends that represent barriers or incentives for geospatial archiving and access. The goal of the Delphi method is to reach a consensus of opinion from a group of experts. We felt that this method was appropriate because the theme under investigation is so specific that it limits the potential participants in the study to small numbers. This ruled out a larger scale questionnaire which would in addition make it unrealistic to further explore the given answers. Alternative qualitative survey methods that allow gathering ideas include focus groups or individual interviews, both requiring participants to meet personally with the interviewer. Because of geographical and time constraints, this was not possible. Furthermore, these methods would not allow at the same time the quantification of obtained answers such as those practiced in this research. For a more complete overview of advantages and disadvantages of the different qualitative research methods, we refer to the very complete work of Przyborski (Przyborski and Wohlrab-Sahr 2014).

The Delphi method is typically used to explore issues that require expert knowledge and judgement. In our case, we involved experts from different countries (see Table 1) who are in contact with some of the mentioned deployments of geodata preservation, geodata producers, or memory institutions, we want to make sure all issues are found. All participants are named experts regardless of their status as customer, producer, or keeper of geodata. After answering questions and giving answers in each round, experts receive feedback and are asked to review or argue their opinion. This process is repeated until the differences of opinion become acceptably small for the contracting entity. The Delphi method was first used to gain quantitative results; nevertheless, often in the first round open questions are asked to extract the most important subjects that subsequent rounds would quantify or classify. A classic Delphi study is conducted blind, which means experts' names are not revealed to other group members. This technique allows "access to the positive attributes of interacting groups (knowledge from a variety of sources, creative synthesis, etc.), while pre-empting their

TABLE 1 Structure of the Delphi study and questions asked to each expert group

	Expert group 1	**Expert group 2**
Expert profile	Information providers (Archives, libraries with map collections, and Open access and government data specialists)	Technological experts (developer of GIS, developers of spatial and location services and geodata producers)
Number of participants	12 recruited / 10 participated (4 female, 6 male) Participation has fallen for the second round.	13 recruited / 11 participated (2 female, 9 male) Participation has fallen for the second round.
Questions of the first round	• What trends reduce the access numbers to legacy data. • What trends increase the access rate to legacy data. • What trends hinder or complicate archiving • What measures could improve the mentioned situation of geodata archiving. • If you could ask the future user one question, what would you ask?	• What technological developments favour archiving? • What technological constraints hinder archiving of vector data? • What measure could improve the mentioned situation of vector archiving? • What technological constraints hinder archiving of raster data? • What measure could improve the mentioned situation of raster archiving? • Technical developments that play a role in how smoothly a user can consume legacy geodata?

Questions of the second round	• Rank the factors that hinder access • In how many years will they disappear • Rank factors that increase access by their positive impact • Rank them by how easily they can be addressed. • Rank the factors that hinder archiving by their negative impact. • In how many years will they disappear? • Rank the factors that improve archiving by their positive impact? • Rank them by how easily they can be addressed.	• Rank the developments by how big their role in archiving is. • In how many years will the constraints for vector archiving disappear? • Rank the measures by their positive impact on vector data archiving. • In how many years will the constraints to raster archiving disappear? • Rank the measures by their positive impact on raster data archiving.
Participating institutions	Not-for-profit: Biblioteca de Catalunya (E), Cartoteca de l'Institut Cartogràfic i Geològic de Catalunya (E), Federal Archives of Switzerland (CH), Landesarchiv Baden-Württemberg (D), International Cartographic Association (AU). Academia: Technical University of Vienna (AU), Edina (UK), University of Barcelona (E), Centro de Ciencias Humanas y Sociales - Consejo Superior de Investigaciones Científicas (E). For profit: desideData Data Company (E)	Not-for-profit: Institut Cartogràfic i Geològic de Catalunya (E), gvSIG developer (E), Swisstopo (CH), Deutsches Zentrum für Luft- und Raumfahrt (D), Institut National de l'Information Géografique et Forestière (F). Academia: Centre for GeoInformatics, University of St Andrews (USA), Data Archiving and Networked Services (NL). For profit: ESRI (CH), Lokku ltd (UK), Google (USA), Ordnance Survey (UK)

negative aspects (attributable to social, personal and political conflicts, etc.)" (Fildes and Allen 2011).

For an explorative Delphi study, it is crucial that the selected experts cover the points of view of all stakeholders (Okoli and Pawlowski 2004). As in our case, based on the reports, on one hand we expected technical and on the other hand more organizational issues; we formed two international expert groups: members of group one were related to information providing and experts in group two had a technical background in geoscience or spatial data or software creation. For recruiting a mixture of personal contact and snowball sampling gave access to specialists. Table 1 gives a resume of the demography and geography of participants.

Group one was asked about trends in general that would influence archiving and access to geodata in a positive or negative way, and group two was asked about technological barriers and incentives. In order to take into account possible differences in knowledge, every expert was given the opportunity to rate his or her own answers according to the perceived understanding of the specific subject. The experts filled out an online survey created with Limesurvey that had open and closed questions. The first round took place in November 2013, and the second round in February–March 2014. The questionnaire contained an introduction with the most important definitions. First of all, we defined geodata as all data that are used or generated in the workflow of base map production. Legacy data was defined as all such geodata where at least one newer version or edition exists. In order to make sure preservation was understood in the sense of the definition as long term access and not only as conservation of the data, we also asked questions about access to legacy geodata. Table 1 summarizes the questions asked to the experts that were related to the barriers and incentives for archiving and legacy data access.

A token was assigned to each participant to help avoid double participation. No answer was compulsory, which is why some questions did not reach the minimum amount of responses, defined as four for each question. When the questionnaire was complete, the answers were extracted to a spread sheet. The questionnaire of the second round contained the ranking options and feedback on the median of the numerical questions. Where possible, the answers were weighted by how confident the expert felt with his or her own answer. We gave each position in the ranking a value from 1 to n, whereat n was the number of options to rank. The group ranking was then represented by the median of all values each option obtained. The answers to the first open questions about trends that favor or hinder archiving respond to the research questions one and four when we compare them to the literature. The ranking in the second round gives the answer to research question two. Finally, the nature of the expert responses gives hints to answer research question three.

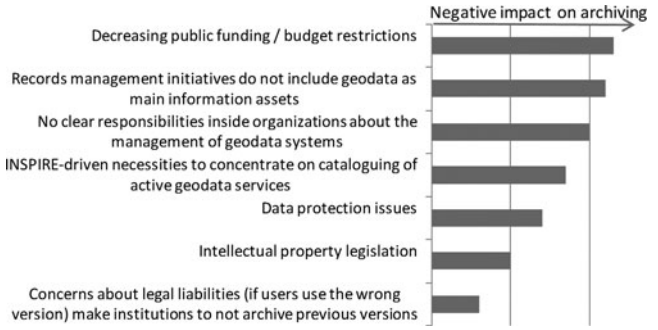

FIGURE 2 Challenges to spatial data preservation ordered by their negative impact on preservation.

RESULTS

General Factors That Hinder or Favor Preservation of Geodata

This section highlights what influences the preservation of legacy data in the opinion of the first Delphi group. Answers to questions about what current trends hinder archiving of legacy geodata were ranked in the second round by the same experts. The answers were then weighted and are shown in the graph above with the factor that has the biggest negative impact on geodata preservation at the top. As the ranking value by itself has no meaning, except in relation to the other values, there is no unit of measurement in this type of figure.

While experts were free to mention any type of hindering aspect to geodata preservation, their answers were of financial, managerial, and legal natures. In Figure 2, we can see that decreasing public funding and current records management practice that does not take into account geodata are the ones that are threatening preservation most. While in the literature the scientific community asked for authoritative responsibility centers (GeoConnections Policy Advisory Node. Working Group on Archiving and Preserving Geospatial Data 2005), the experts translated this into a need for recognition of the value of geodata and its proper management (answers 2 and 3 from the top in Figure 2). The legal issues such as intellectual property legislation and concerns about legal liabilities have lower negative impact. Nevertheless, it is still significant that three legal issues have been mentioned and none of technical nature when we foresaw the data protection issues having a technical as much as a legal aspect. The experts came up with a legal issue that has not been mentioned in the literature and reports: risk of using the wrong version prevents producing agents from archiving. The presence of legal concern in this Delphi group can be explained by the nature of the experts. As members of memory institutions, they are not the owners of the geodata, and therefore, they consider the difficulties

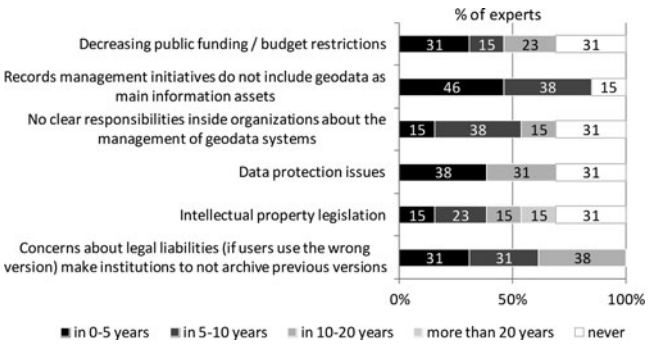

FIGURE 3 The percentage of experts that estimated these constraints for geodata preservation to be overcome in the indicated time frame.

in obtaining such data and or giving access to it. With respect to the third research question we find that, while public archives have little power over funding restrictions, any institution can decide by itself about including geodata as assets and implementing data management responsibilities in the company.

So which of these challenges will be overcome first? Figure 3 shows the estimations of the experts in Delphi group 1.

The lack of consciousness about the value of geographic assets in records management is considered to be most likely to change early. Indeed, the value of such data has been emphasized many times (Lazorchak et al. 2008; Erwin, Sweetkind-Singer, and Larsgaard 2009; North Carolina Center for Geographic Information and Analysis 2009; Shaon et al. 2012; Erwin and Sweetkind-Singer 2009; Harris 2001; Caruso et al. 2013). Awareness of the value of these assets seams to find its way into management practice. Furthermore, we observe that expert opinion diverges most about data protection issues: while one third thinks that these will not be an issue any more in about five years, another third believes that this problem will never disappear. The case of intellectual property legislation is very similar. This might reflect the differing legal situations in the countries of origin of the experts, or a simple lack of consensus about the real implications of such issues. Asking the experts gave us the means to distinguish which factors might be overcome sooner, so that an archiving institution can estimate the needs of acting or the eventual benefits of waiting for a better starting point for their preservation project.

The first Delphi group also commented on the positive impact of certain measures that can improve preservation of legacy geodata. The following diagram in Figure 4 shows the measures with the best effect on archiving to the right side. If the measure simultaneously appears in the upper part of the figure, it is also easier to address.

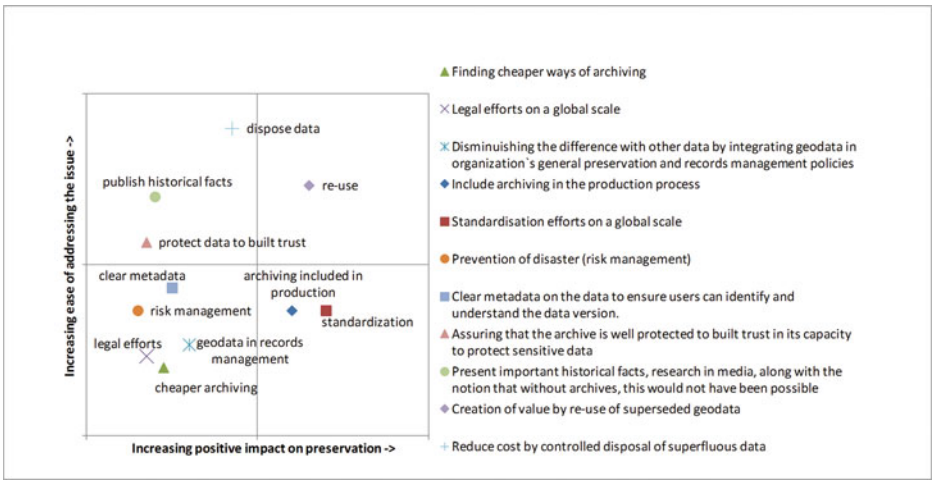

FIGURE 4 Measures to overcome the problems of geodata archiving in relation to how easy it is to address these issues.

We see that standardization efforts on a global scale would have the biggest positive impact on archiving legacy geodata, with the creation of value by re-use also being considered a very effective measure. Moreover, it seems to be an issue that is straightforward to address. Fostering re-use is a measure that each institution can contribute to on its own by offering access to the data and making it retrievable without an overload of organizational challenges as would be the case with global agreements. Nevertheless, because of legal issues, offering access is much easier for the producing agent than for a public archive. Including archiving in the production process of geodata would also have a big positive effect on preservation followed by the controlled disposal of superfluous data in order to reduce the cost of preservation. We want to draw some attention to the latter as it is the measure considered easiest to implement.

Due to the variety of services, geodata providers hold the same data in various forms. Once decided for an archival format, duplication should be avoided, and data in process stages in between so-called end-products can be deleted. Which process stages represent end-products must be defined by each institution. None of the reports and literature has mentioned disposal explicitly, but many projects include some kind of criteria for data appraisal that will limit the amount of data to archive. All other measures have been evaluated to be clearly harder to address. Prevention of disaster has the lowest impact value. This is not surprising, considering that it is a measure "post-archiving." If we have not solved the archiving challenge in the first place, there will be no data to protect from disaster. Building a business case for archiving geodata has not been mentioned by the experts as a measure to improve preservation as this has been in the technical reports.

Technological Factors That Influence Archiving of Legacy Geodata

As technology changes rapidly, we looked not only for the perceived challenges, but explored future trends by asking the second expert group for technological developments that will influence archiving. They gave the following options that we present ranked by their positive impact.

The use of standards both by the producer of geodata and the archive itself is obviously assessed as important. Standardization is an old issue, asked for in many reports, that is now perceived as one of the major trends. At the same level we find storage and retrieval capabilities. Here we see that retrieval is considered a necessary counterpart to archiving. This confirms a general tendency to define preservation as the capacity to give sustainable access in the long term (Blue Ribbon Task Force 2010; North Carolina Geographic Information Coordinating Council 2008). Where there is no retrieval, there can be no access. Easy access is also what triggers use and generates new value and awareness for the archive. In the middle field we find the use of cloud computing and linked data. As these trends still have quite a high impact on archiving, it is worth considering their implementation in preservation projects.

In Figure 5, we represent the technological constraints that hinder or complicate archiving versions of legacy vector geodata. The experts esti-

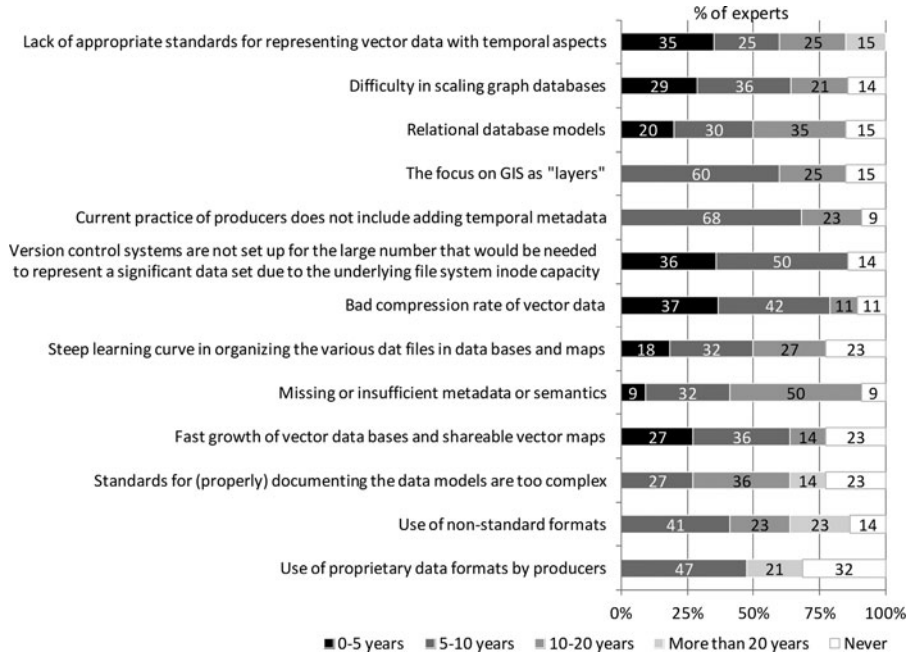

FIGURE 5 The percentage of experts that estimated these constraints for vector data archiving to be overcome in the indicated time frame.

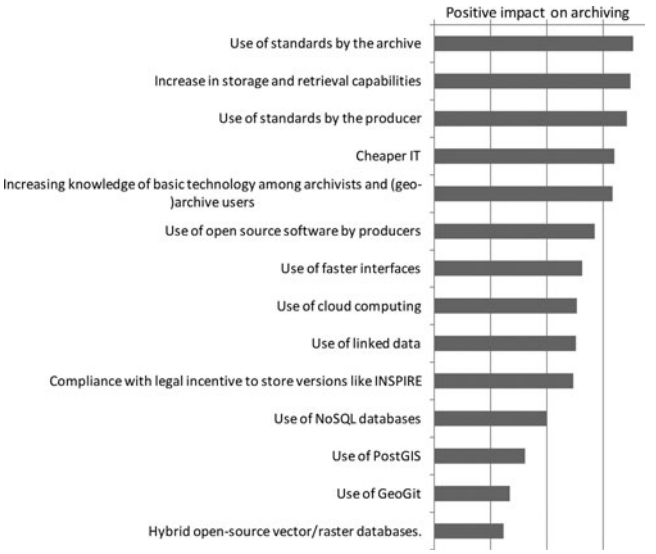

FIGURE 6 Technological developments with a positive impact on geodata archiving. The longer the bar the bigger the positive impact.

mated how many years it will take for the barriers to disappear or to be overcome. The numbers are the percentages of experts subscribing to this opinion.

We can see that the three factors that have the highest probability to be the first to disappear are the poor compression rate of vector data, the fact that the version control systems are not set up for a larger number of features, and the lack of appropriate standards for representing vector data with temporal aspects. The contradictory views on the use of proprietary data formats by producers are also striking. Almost half the experts think that the problematic use of proprietary data formats will disappear in 10 years, while on the other hand this factor is also associated with the highest number of experts who think this problem will never disappear.

In order to overcome these challenges, we asked experts what measures could be taken and what impact they would have on vector data archiving. Interestingly, the use of standard file formats by producers is also the measure that would have the biggest positive impact on geodata preservation as shown in Figure 6.

The use of open source software does not have as big an impact as the use of standard file formats. Open standard file formats are highly important for accessing and rendering files in the future, yet the use of open source software does not necessarily imply the generation of standard file formats. Nevertheless, both measures are related because the use of standard file formats would help developers to implement software solutions that work with them. We can also see that for some technical problems experts suggest

organizational solutions such as simplification of the documentation process or teaching users and producers. Furthermore, it will take about 10 years until producers systematically add temporal metadata. We can interpret this fact as a reaction to the development of GIS capable of storing temporal metadata and the implementation of the INSPIRE metadata standard that has been adopted by member states of the European Commission and includes guidelines for temporal metadata (European Commission 2010).

Challenges that have not been mentioned for vector data but for raster data include the high volume of data, lack of band width, and slow storage technologies. While these three problems have a high probability to be overcome in at least 10 years, there are also a few experts who think these problems will never disappear. Similar to vector data archiving the use of proprietary software and nonstandard file formats by the data producer is also a problem for raster data. Again, it is suggested to be a setback which is difficult to overcome, though it would be an impacting measure to decrease the challenge of legacy raster data archiving.

Despite continually decreasing IT prices there seems to be a general need for cheaper storage, related to the very large volumes of data created. It is not surprising that higher bandwidth capacity and faster storage have similar values of impact, as they are both important when transferring or migrating data. IT prices were already mentioned as generally impacting archiving (see Figure 5) especially in the context of decreasing public funding (see Figure 2). In third place for raster and vector data is a documentation issue: there is a need for simpler documentation and metadata. Simpler metadata is given as a measure to improve the technological (!) aspects of preservation. We discuss metadata further in the section title "Discussion."

Access to Legacy Data

Hereafter, we present the results of Delphi Group 1 regarding the questions about access to superseded geodata. The following figure shows factors that increase access to superseded geodata in relation to how easy these challenges are to address. The factors with the biggest positive impact on the increase show on the right side of the figure. Similarly, the factors which are easier to address appear in the upper part of the figure.

On the right-hand side, we can see that the use of GIS-related applications available on many platforms will also increase the access to legacy data. The use of open technologies is positioned on the left-hand side of the chart which stands for a low positive impact on access. Comparing with Figure 5, open technologies show higher relevance for preservation than for access. Open government data policies and central cataloguing have to do with the visibility of the data and were ranked at the same level regarding their positive impact on access. The most interesting starting points when it comes to increasing access are the factors in the upper right corner: central

cataloguing encouraged by initiatives like INSPIRE, APEnet/Europeana, and national catalogues as well as open data policies and transparency that will increase the awareness of the existence of data sets are key here (Geospatial Multistate Archive and Preservation Partnership and Library of Congress 2010). Given the broad attention to which they are currently subject now is an excellent point to address these issues.

DISCUSSION

Based on these results we can say that the main issues that came forth are primarily not of a technological nature. Especially considering that the technological issues mentioned have a higher probability to be overcome in at least 10 years than nontechnological, more effort should go into organizational challenges. Nevertheless, new solutions developed outside the preservation community seem to offer new possibilities such as cloud computing and faster bandwidth. Furthermore, technological progress can lead to financial facilitation as would be the case with better compression rates or cheaper storage media. Symptomatic for the fast changing environment is, on the one hand, the fact that several trends, challenges, or issues are new compared to those reported some years ago in project reports. For example, this is the case of cloud storage. On the other hand, XML-technology, mentioned as a solution in the reports, was not reflected by the Delphi group. One of the reasons could be that this technology is already so broadly in use, for example as XML-based file formats, that it seems not necessary to make it explicit. Other issues, such as version control systems that are not set up for a larger number of features and historic states, have been recognized in the project reports and by the experts but seem likely to be overcome soon. Even though there are many legal issues (see Figure 4), it does not make much sense to address them on an international level. Contrariwise, spatial data preservation institutions are facing the lack of use of standards, and this change must be made by an international effort. This does not mean that standards do not exist, but often there are too many, and they are not necessarily international. Where international standards exist, such as those defined by INSPIRE or the Open Geospatial Consortium, they must be implemented by many institutions; otherwise a standard is of no help for long term preservation. It is clearly worth investing in this point as likely to have the biggest positive impact on archiving (Figure 4). Although standardization is addressed at different levels (metadata about geodata, file formats, interoperability, etc.), it is one of the most resistant subjects, estimated to be reached only far in the future (Figure 5).

The case is more promising for open technology; it can be adopted by a single institution and produce already a positive effect for the memory institution that has to take over this data (Figure 6). In addition, use of open source software is considered a task easy to address (Figure 8).

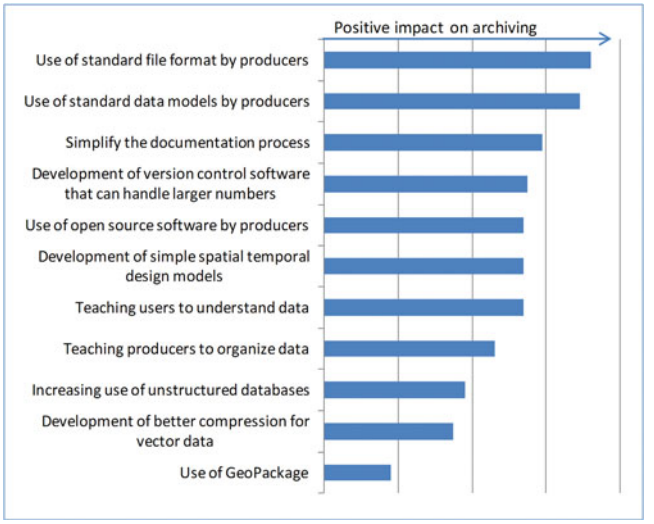

FIGURE 7 Measures that could be taken to overcome the problems of vector data archiving in order of their positive impact.

Regarding the research question about the responsibility of such measures, we see that some issues must be addressed by the community of stakeholders, such as standardization or collective cataloguing, and some can be surmounted by the individual memory institution or producer, such as the proper documentation of data sets and data models. On some challenges, the stakeholders have little power, for instance on IT prices or the GIS knowledge of the user community. Nevertheless, with different service

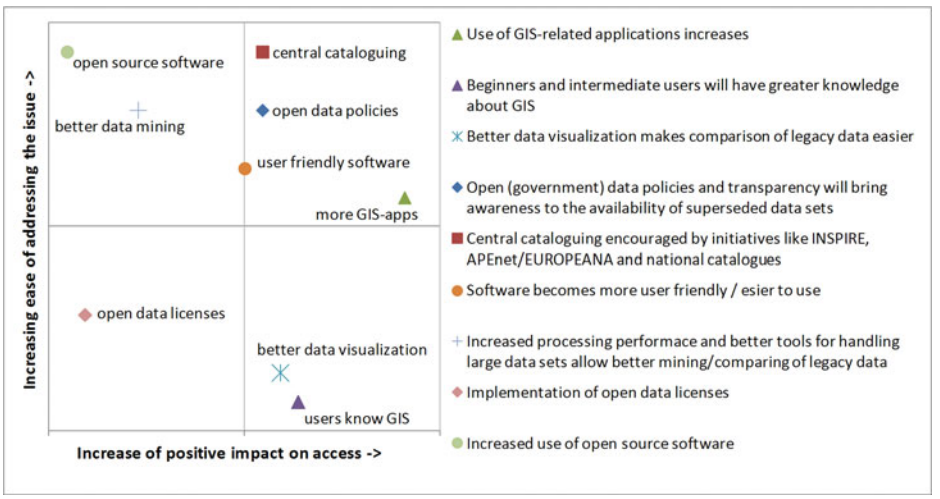

FIGURE 8 Measures that increase access to legacy data in relation to how easy it is to address these issues.

levels they can transcend a lack of user knowledge. Adapted services are common in libraries (Stone Muilenburg 2001), but we find also examples in archives; for example, the service for academia set up by UK Ordnance Service and Edina and NOAAs service levels for different user expertise (Cramer and Kott 2010; Duerr, Weaver, and Kaminski 2010). One service level could be educational tools for the user teaching him or her to understand the data. Another educational approach is suggested for producers, trying to emphasize the importance of organizing data. Such close attention would surely have an impact on archiving, though it is hard to implement in times of restricted budgets.

The study shows that the lack of standardization and the use of proprietary formats is still a big problem. While standardization efforts on a global level are estimated to have the biggest positive effect to preservation, only half of the experts believe that data producers are going to use non-proprietary software and file formats by the end of 2034. Nevertheless, if a producing agency is trusted at the same time with the long term storage for the data, acting on standardization would surely be an effective measure to improve preservation. Archives receiving data from different producers have less influence on the variety of file formats they have to curate than the producers themselves. In order to scale down the cost of preservation, it is important that at the point of transfer from a producer to a long term archive data should be converted into a standardized format. If data was produced already in an open standardized format, this would simplify the transfer to the archive.

Cost of IT and digital storage specifically recur regularly in the comments of experts and literature. As archives usually have no influence on prices, cost reduction is an important measure to improve preservation. As a good starting point to improve archiving, we suggest the controlled disposal of superfluous data as a measure to reduce cost. Disposal was suggested as the easiest problem to address and the fourth in terms of having a positive impact on preservation of geodata if applied. In order to face this issue, many archives and producers develop selection criteria or appraisal policies that assist them in deciding what to keep (Dorfey et al. 2009; Duerr, Weaver, and Kaminski 2010; The National Oceanic and Atmospheric Administration (NOAA) 2008; Tjalsma and Rombouts 2011; Weaver, Meier, and Duerr 2008; Bos et al. 2010).

In Figures 5 and 6, we have seen that simpler documentation is suggested as a measure to improve missing metadata. The following arguments might explain why: when an archive simplifies the metadata standard in the first place, it might reduce cost due to easier implementation of the software solution. Indeed, software for archives are mainly still individual solutions and highly costly. Secondly, metadata and documentation simplification can also increase the quality of the documentation thanks to the effective percentage of data, models, or processes being documented. If the geospatial

archive has influence on the metadata at the production state, this is a good starting point to improve archiving. If not, it is still well advised to control the documentation of the data set on acquisition, the process step of introducing data in a digital archive.

An independent memory institution has less influence on the standardization of the metadata and semantics or the documentation process at the producer site than an institutionally attached archive, but it should support metadata standardization efforts. Standardization of metadata would make it more immediately interoperable. Metadata also helps the comprehension of the intellectual and technical content of the data, which by itself is a condition for re-use in the future.

Figure 4 shows a strong connection between re-use and preservation. Among the measures most easy to address are improving the visibility and re-use of data. This combination makes data re-use an effective measure towards improving its preservation. In order to re-use data, it must be retrieved and correctly interpreted by humans and machines. Re-use can be improved by better finding instruments, quality metadata, and public awareness such as publications citing the data. Additionally, technology has to follow up as well: files should be migrated to formats that can be read by future software, and storage media has to be refreshed. Nevertheless, if data is constantly requested and retrieved, technology updates will be natural and constant.

Finally, increasing awareness of the existence of data sets is considered important. This can be done by central catalogs or interconnected infrastructures. Certainly, current trends for open government data and transparency will also influence the visibility of the data. Data will be re-used only when people are aware of its existence. Data in use has a stronger stakeholder community to defend its survival.

CONCLUSIONS AND FUTURE WORK

This research emanated from a series of unorganized issues and challenges reported by preservation projects for geodata. Its intention was to complete the list of issues by means of expert questioning. The Delphi method allowed then for comparing the impact of trends and solutions on archiving with the ease of addressing them in order to support the decision process of institutions that want to implement or improve a preservation process.

Thanks to the expert opinions, we have found promising starting points for improving topographic data preservation and have come to the following conclusions:

1. The issues we should focus on as a community of stakeholders in the next few years are those that need more time implementing: it is important for preservation to simplify documentation of geodata and to change the way

we think about geodata as layers. Now that legacy geodata starts to be seen as an asset, we need to collaborate with the data producer in its proper management.

2. The measures easier to implement are the disposition of data, fostering re-use, publicity for data and the work archives do, and the protection of data to build trust. All these measures lie in the power and responsibility of individual institutions.
3. All stakeholders have responsibility: For improving preservation, the producer of geodata should think of archiving and include items in the whole production process. The archive should foster re-use when the legal situation allows it. Finally, the community of stakeholders should standardize as much as possible. Concerning access the community of stakeholders tries to influence through the INSPIRE directive about cataloguing data and open data policies. On the technological side, we should ask for easier software and open source technology.
4. The challenges brought up by the experts are basically the same as those mentioned in the literature. Nevertheless, it is striking how clearly re-use of data is standing out as an easy to implement favoring factor to preservation. Some new means to address the challenges of access were presented even though minor in the expected impact. An exception with relatively more impact lies in the possibility of cloud computing and influencing the user knowledge of GIS.

Although some challenges are more urgent for geodata than for other data types such as the correct documentation of the complexity of data models or the dominance of proprietary software and file formats, most can be applied directly to preservation of any data. Therefore, an overview of challenges and measures concerning digital preservation in general would also reveal starting points for spatial data archiving.

The explorative method used had a sufficient amount of participants of different stakeholders to reveal new topics; nevertheless, we cannot disregard the fact that experts suggest barriers and measures based on their personal experience and horizons. Not all challenges or measures must apply to the specific context of a memory institution searching for advice in spatial data archiving. When national differences are unavoidable, such as in legal affairs, we did not get tangible suggestions for improving the problems. A limitation of this study lies in the limited amount of rounds and participants. This could be the reason why some challenges or measures referred to in scientific literature were not picked up by the Delphi groups. Further rounds may have eventually revealed more issues by presenting the first mentioned to the experts and asking for completion. By distributing the questions over two different expert groups, we were able to ask more questions and adjust them to participants' profile. Nevertheless, this meant that we limited the number of responses per question, which was reduced in some cases to four.

This lets the question come up of how to provide sustainable access for data in the intermediate state of archiving up to 70 years old when this data is still expected to be queried in all its previous stages. Can we build database systems capable of registering the history of all features? Directly related is the cost of the storage space for vector databases when continuous updates are archived. Historized databases are a forthcoming option for intermediate archives that needs further exploration. Up to now, for some memory institutions it is financially not possible to bear the cost of such a huge database and manage the updates (Fleet, personal communication 2013).

Figure 4 showed a clear relation between re-use and preservation. Although there are studies that try to evaluate the value of data, among other things the value of re-use, there is no study yet about how and under what circumstances increased use and access benefits preservation. This article can enumerate the measures that were mentioned to improve re-use as discussed in the previous section, but the explorative nature of this study does not allow for deeper analysis. Therefore, we would suggest further work on legacy data re-use and its effects in the light of its benefit for long term preservation of spatial data. Furthermore, the trends and propositions that came up with the experts and do not show in the literature could be scientifically analyzed, for example the effectiveness of user education for the archived data in the light of the Open Archival Information Systems (OAIS) standards requirement to adapt the service to the user knowledge.

REFERENCES

Beruti, Vincenzo, Esther Conway, M. Eugenia Forcada, David Giaretta, and Mirko Albani. 2010. ESA Plans: A pathfinder for long term data preservation. In *iPres 2010*. http://epubs.stfc.ac.uk/bitstream/6403/beruti-76.pdf

Bleakly, Denise R. 2002. Long-term spatial data preservation and archiving: What are the issues? *Contract*. Albuquerque. http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2002/020107.pdf

Blue Ribbon Task Force. 2010. Sustainable economics for a digital planet: Ensuring long-term access to digital information. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Bos, Marguérite, Helen Gollin, Urs Gerber, Jérémie Leuthold, and Urs Meyer. 2010. Archiving of geodata: A joint preliminary study by Swisstopo and the Swiss Federal Archives. http://www.swisstopo.admin.ch/internet/swisstopo/en/home/topics/geodata/geoarchive.parsysrelated1.59693.downloadList.93958.DownloadFile.tmp/preliminarystudyarchivingofgeodata.pdf

Caruso, Giuseppa, Luigi Briguglio, Brian Matthews, and Barbara Polsinelli. 2013. D31.1: Parameters for long term preservation and data sustainability models.

Conway, Esther, Sam Pepler, Wendy Garland, David Hooper, Fulvio Marelli, Luca Liberti, Emanuela Piervitali, Katrin Molch, Helen Glaves, and Lucio Badiali. 2013. Ensuring the long term impact of earth science data through data cu-

ration and preservation. *Information Standards Quarterly 25*(3) (January): 28. doi:10.3789/isqv25no3.2013.05. http://dx.doi.org/10.3789/isqv25no3.2013.05

Cramer, Tom, and Katherine Kott. 2010. Designing and implementing second generation digital preservation services: A scalable model for the Stanford Digital Repository. *D-Lib Magazine 16*(9/10) (September). doi:10.1045/september2010-cramer. http://www.dlib.org/dlib/september10/cramer/09cramer.html

Cushing, Judith B., Nicole E. Kaplan, Christine Laney, Juli Mallett, Ken Ramsey, Kristin Vanderbilt, Lee Zeman, et al. 2008. Integrating ecological data: Notes from the Grasslands ANPP Data Integration Project. In *Environmental Information Management Conference 2008 Albuquerque, New Mexico, September 10–11, 2008*, Gries, Corinna, and Matthew B. Jones (eds.), pp. 52–61.

Dingwall, Glenn, Richard Marciano, Reagan Moore, and Evelyn Peters McLellan. 2005. From data to records: Preserving the geographic information system of the City of Vancouver. *Archivaria 64*(Fall 2007): 181–198. http://journals.sfu.ca/archivar/index.php/archivaria/article/view/13157/14408

Dorfey, Beate, Sabine Graf, Bernhard Grau, Jörg Homberg, Karsten Huth, Christian Keitel, Peter Sandner, and Andrea Wettmann. 2009. Handreichung Zur Archivierung Elektronisch Vorliegender Geodaten. ARK AG ESys. ARK IT-Ausschuss. http://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/ark/handreichung_geodaten_20090928.pdf

Duerr, R. E., R. L. Weaver, and M. Kaminski. 2010. Data acceptance procedures and levels of service at the National Snow and Ice Data Center. In *2010 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2322–2325. *IEEE*. doi:10.1109/IGARSS.2010.5650356. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5650356

Ellipse, Projektteam. 2013. Projekt Ellipse. Konzeption Der Archivierung von Geobasisdaten Des Bundesrechts: Konzeptbericht. Bern. http://www.bar.admin.ch/themen/00876/00939/index.html?lang=de&download=NHzLpZeg7t,lnp6I0NTU042l2Z6ln1acy4Zn4Z2qZpnO2Yuq2Z6gpJCDeIR5gGym162epYbg2c_JjKbNoKSn6A–

Erwin, Tracey, and Julie Sweetkind-Singer. 2009. The National Geospatial Digital Archive: A collaborative project to archive geospatial data. *Journal of Map & Geography Libraries 6*(1) (December 31): 6–25. doi:10.1080/15420350903432440. http://www.tandfonline.com/doi/abs/10.1080/15420350903432440

Erwin, Tracey, Julie Sweetkind-Singer, and Mary Lynette Larsgaard. 2009. The National Geospatial Digital Archives—Collection development: Lessons learned. *Library Trends 57*(3): 490–515. doi:10.1353/lib.0.0049. http://muse.jhu.edu/content/crossref/journals/library_trends/v057/57.3.erwin.html

European Commission. 2010. INSPIRE metadata implementing rules: Technical guidelines based on EN ISO 19115 and EN ISO 19119. http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf

European Commission. High Level Expert Group on Scientific Data. 2010. Riding the wave: How Europe can gain from the rising tide of scientific data. Final report.

Fildes, Robert, and P. Geoffrey Allen (eds.). 2011. *Forecasting. Vol. III: Judgemental methods and forecasting competitions*. Los Angeles: SAGE.

Fleet, Christopher. 1999. Ordnance survey digital data in U.K. legal deposit libraries. *LIBER Quarterly 9*(2): 235–243.

Fleet, Christopher. 2003. The legal deposit of digital spatial data in the United Kingdom. *LIBER Quarterly 13*(1). http://liber.library.uu.nl/index.php/lq/article/view/7708/7746

GeoConnections Policy Advisory Node. Working Group on Archiving and Preserving Geospatial Data. 2005. Archiving, management and preservation of geospatial data. ftp://ftp2.cits.rncan.gc.ca/pub/geott/ess_pubs/292/292109/cgdi_ip_23e.pdf

Geospatial Multistate Archive and Preservation Partnership (GeoMAPP). 2011. Best practices for archival processing for geospatial datasets. http://www.geomapp.net/docs/GIS_Archival_Processing_Process_v1.0_final_20111102.pdf

Geospatial Multistate Archive and Preservation Partnership, and Library of Congress. 2010. GeoMAPP Interim Report: 2007–2009. http://www.geomapp.com/docs/GeoMAPP_InterimReport_Final.pdf

Harris, R. 2001. Earth observation data archiving in the USA and Europe. *Space Policy 17*(1) (February): 35–48. doi:10.1016/S0265-9646(00)00052-7. http://linkinghub.elsevier.com/retrieve/pii/S0265964600000527

Janée, Greg. 2008. Preserving geospatial data: The National Geospatial Digital Archive's approach. In *Archiving*. pp. 25–29.

Kalb, Hendrik, Edward Pinsent, and Matthias Trier. 2013. Interoperability of Web archives and digital libraries: A Delphi study. In *iPRES*.

Kennedy, Adam M., Suzanne M. Remillard, Donald L. Henshaw, Lawrence A. Duncan, and Barbara J. Bond. 2008. Converting data to information: Coupling lab-level database functionality with primary LTER data archiving systems. In *Environmental Information Management Conference 2008 Albuquerque, New Mexico, September 10-11, 2008*, Gries, Corinna, and Matthew B. Jones (eds.), pp. 77–82.

Lazorchak, Butch, Zsolt Nagy, Dennis Goreham, Steven Morris, Alec Bethune, and Matt Peters. 2008. Creating a business plan for the archival preservation of geospatial data. In *Archiving*, p. 79.

Leyk, Stefan, Ruedi Boesch, and Robert Weibel. 2006. Saliency and semantic processing: Extracting forest cover from historical topographic maps. *Pattern Recognition 39*(5) (May): 953–968. doi:10.1016/j.patcog.2005.10.018. http://linkinghub.elsevier.com/retrieve/pii/S0031320305004139

Mcgarva, Guy, Steven P. Morris, and Greg Janée. 2009. Technology watch report: Preserving geospatial data. *Architecture*. www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee

Moran, Mary Susan, Barbara S. Hutchinson, Stuart E. Marsh, Mitchel P. McClaran, and Aaryn D. Olsson. 2009. Archiving and distributing three long-term interconnected geospatial data sets. *IEEE Transactions on Geoscience and Remote Sensing 47*(1) (January): 59–71. doi:10.1109/TGRS.2008.2002815. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4689344

Morris, Steven P., James Tuttle, and Jefferson Essic. 2009. A partnership framework for geospatial data preservation in North Carolina. *Library Trends 57*(3): 516–540. doi:10.1353/lib.0.0050. http://muse.jhu.edu/content/crossref/journals/library_trends/v057/57.3.morris.html

National Academy of Sciences. National Research Council, and Committee on the Preservation of Geoscience Data and Collections. 2002. *Geoscience data and collections: National resources in peril*. Edited by The National Academy Press. http://www.nap.edu/catalog.php?record_id=10348

North Carolina Center for Geographic Information and Analysis. 2009. North Carolina Geospatial Data Archiving Project: 2008 local government geoarchives survey.

North Carolina Geographic Information Coordinating Council. 2008. GICC archival and long term access ad hoc committee: Final report. *Perspective*. http://www.ncgicc.com/Portals/3/documents/Archival_LongTermAccess_FINAL11_08_GICC.pdf

Okoli, Chitu, and Suzanne D. Pawlowski. 2004. The Delphi method as a research tool: An example, design considerations and applications. *Information & Management 42*(1) (December): 15–29. doi:10.1016/j.im.2003.11.002. http://linkinghub.elsevier.com/retrieve/pii/S0378720603001794

Przyborski, Aglaja, and Monika Wohlrab-Sahr. 2014. *Qualitative Sozialforschung: Ein Arbeitsbuch. 4*. erw. Au. München: Oldenbourg.

Shaon, Arif, Carsten Rönsdorf, Urs Gerber, Kai Naumann, Paul Mason, Andrew Woolf, Michael Kirstein, Marguérite Bos, and Göran Samuelsson. 2012. Long-term sustainability of spatial data infrastructures: A metadata framework and principles of geo-archiving. In *iPres*. https://ijpis.net/Global/Forskning/Center och institut/CEDIF/Geodata/GeoPres_IPRES_CR.pdf

Shaon, Arif, and Andrew Woolf. 2011. Long-term preservation for spatial data infrastructures: A metadata framework and geo-portal implementation . *D-Lib Magazine 17*(9/10) (September). doi:10.1045/september2011-shaon. http://www.dlib.org/dlib/september11/shaon/09shaon.html

Stone Muilenburg, Jennifer. 2001. The changing role of GIS in the map room. In *The map library in the new millennium*. Parry, Robert B., and C. R. Perkins (eds.), pp. 56–71. London: LA Publishing.

Sweetkind, Julie, Mary Lynette Larsgaard, and Tracey Erwin. 2006. Digital preservation of geospatial data. *Library Trends 55*(2): 304–314. doi:10.1353/lib.2006.0065. http://muse.jhu.edu/content/crossref/journals/library_trends/v055/55.2sweetkind-singer.html

The National Oceanic and Atmospheric Administration (NOAA). 2008. NOAA Procedure for scientific records appraisal and archive approval: Guide for data managers. https://www.ngdc.noaa.gov/wiki/images/0/0b/NOAA_Procedure_document_final.pdf

Tjalsma, Heiko, and Jeroen Rombouts. 2011. Selection of research data: Guidelines for appraising and selecting data. http://www.dans.knaw.nl/sites/default/files/file/publicaties/Selection_of_research_data_DEF.pdf

Weaver, Ronald L. S., Walter M. Meier, and Ruth M. Duerr. 2008. Maintaining data records: Practical decisions required for data set prioritization, preservation, and access. In *2008 IEEE International Geoscience and Remote Sensing Symposium*, III–617–III–619. *IEEE*. doi:10.1109/IGARSS.2008.4779423. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4779423

Zaslavsky, Ilya. 2001. Archiving spatial data: Research issues. *San Diego Supercomputer Center: Technical Report*.