

**Publishing in the Digital Age:**  
*A survey of digitization and long-term  
preservation in the publishing world*

**By Stacey von Winckelmann**  
**Intern at the Koninklijke Bibliotheek**  
**University of Leiden, Book and Digital Media Studies**  
**24 March 2011**

# **Table of Contents**

<b>Background.....</b>	<b>3</b>
<b>Demographics.....</b>	<b>4</b>
<b>Digital Materials.....</b>	<b>5</b>
<b>Long-term Preservation.....</b>	<b>8</b>
<b>Publishers and Libraries.....</b>	<b>10</b>
<b>Conclusions.....</b>	<b>12</b>
<b>Recommendations.....</b>	<b>14</b>

## **BACKGROUND**

With the increase in both born digital and digitized material available, it is important that research is focused on the direction publishing organizations are taking in regard to the issues of long-term accessibility and long-term preservation. For libraries, especially national libraries, building an effective, sustaining archive that can provide long-term accessibility and preservation can be considered part of the established library mantra of ‘Free access of information for all!’ Revealing and understanding the direction publishing organizations are moving toward will assist libraries in understanding how to better serve their patron populations.

This survey has been created to uncover the extent of publisher’s knowledge, commitment and concern, with respect to digitization and long-term preservation. Furthermore, some the expectations publishers have of libraries within the context of long-term preservation, archiving, and the potential for future partnerships was also explored. Lastly, guiding questions used to focus the direction of the survey were:

1. How do publishing organizations view libraries?
2. What are common file formats currently being used by publishing organizations?
3. What type of metadata formats are publishing organizations sending to libraries?
4. How do publishing organizations see the future of publishing in terms of digital libraries and long-term preservation?
5. What if any are their expectations of libraries?

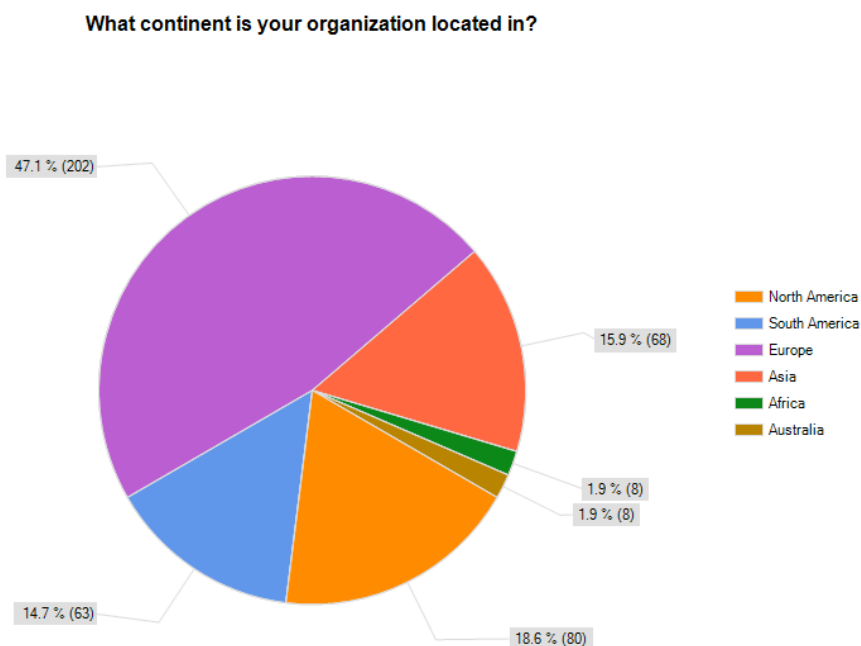
The survey was constructed with four sections, demographics, digital materials, long-term preservation, and publishers and libraries. Keeping the needs of publishers in mind, this last section of the survey was meant as a needs assessment and/or satisfaction check regarding the archiving agreements and/or services provided by libraries or outside organizations, such as Portico.

The goals of the survey were met through the large number of responses and extensive comments that were left by the participating publishers. The respondents had an opportunity to provide comments on various questions within the survey. These particular comments helped focus the perspective that each publisher was responding from.

This survey was made possible thanks to the many interviews, meetings, and feedback provided by KB staff and department heads. Additionally, weeks of independent research contributed to the design of the survey as well.

## **DEMOGRAPHICS**

The survey was constructed using SurveyMonkey and was distributed through email to the Federation of European Publishers, the European Publishers Council, DOAJ, STM, Springer, Brill, and Elsevier. Personal emails requesting participation from certain members of these organizations were sent to approximately 200 publishing organizations. It is estimated that the survey had the potential to reach just over 5000 publishing organizations, with a return rate in SurveyMonkey of 770 responses. Of those 770 responses, 583 were viable. While the survey respondents came from 6 different continents (*figure 1*), the majority of results came from European publishers (202). Of the size 583 publishing organizations varied, with majority of respondents, 68 percent (396), coming from very small organizations that publish less than 50 journals a year. Scholars and students were the main audience for the majority of publishers, while 83 percent of those used open access as their business model. Just over 400 publishers requested a copy of this report be sent to them.

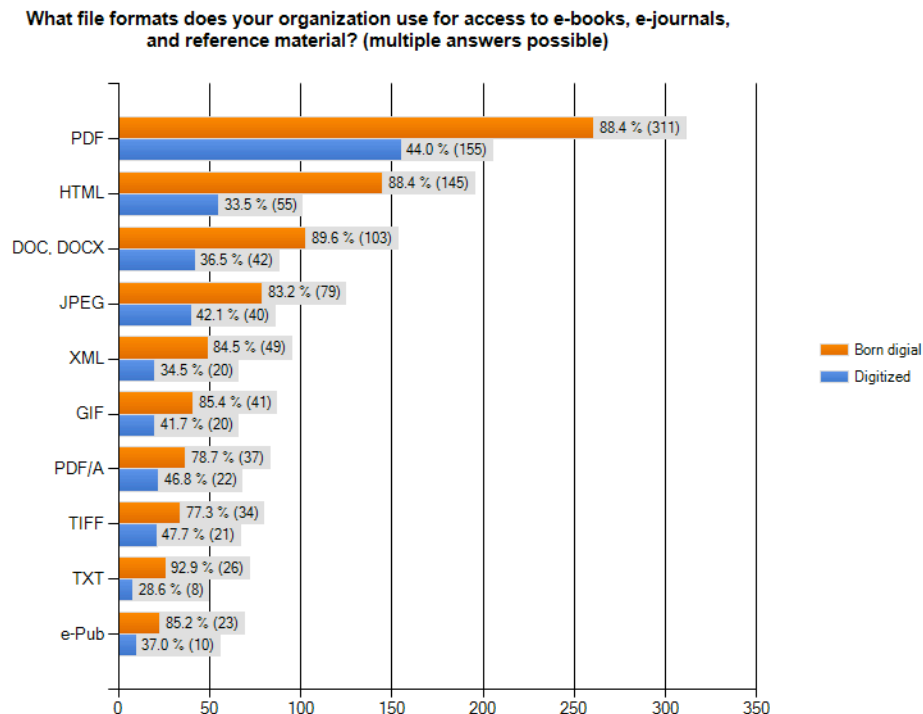


**Figure 1:** *Publisher locations*

## **DIGITAL MATERIALS**

To digitize or not to digitize, that is the question. The first section of the survey focused on how publishing organizations process their born digital and digitized materials. For the purpose of this survey, born digital was defined as materials that originate in a digital form. It was not important within the context of the survey whether the object eventually receives a physical counterpart.

When asked what types of publications are offered by their organization, 51 percent said born digital. A good number of publishers also offer both born digital and digitized material. Most carry out their digitizing in house and have been doing so for 3 to 5 years. Publishers have also been offering born digital material for the same length of time. A large number of publishers, 72 percent, said that they are 100 percent digital. On the topic of file formats used for born digital and digitized material (*figure 2*), publishers use PDF and HTML as the most common file formats for e-books, e-journals, and reference material. JPEG and DOC/DOCX were also popular choices. Other respondents commented that they were using TEX, DjVu, PNG, Scribd, CIF, InDesign, and Latex.



**Figure 2:** File formats

When asked what criteria publishers look at for the selection of file formats (figure 3), the majority said ‘Popularity’, ‘Simplicity’, and ‘Openness’ were *Very important* for their consideration, while ‘Robustness’ and ‘Self-documentation’ were regarded as *Important*. The metadata standard used by publishers for their born digital and digitized files was based on the material type. Dublin Core was the most common standard for e-books, e-journals, and reference works. SCORM and MARC21 were also used for e-books, MARC21 and METS for e-journals, and EAD and SCORM for reference works.

A many publishers, 45 percent, were unaware if they worked with aggregators, while 20 percent of the publishers indicated that they do work with aggregators. The publishers working with aggregators are sending them PDF and Dublin Core for both born digital and digitized material. A slightly smaller group is sending XML and MARC21 to the aggregators they work with.

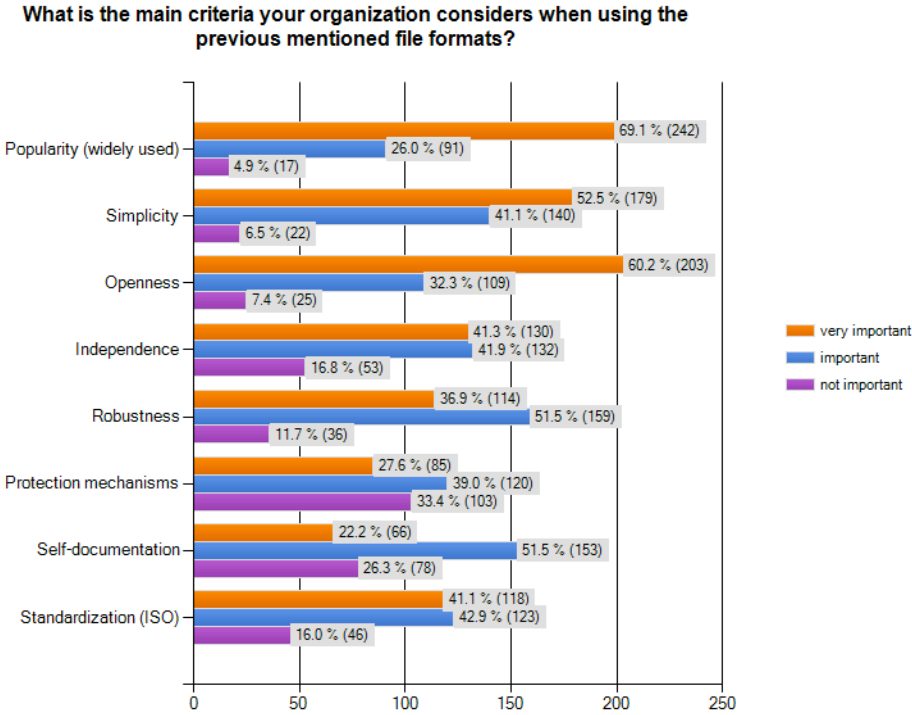


Figure 3: Criteria for file formats

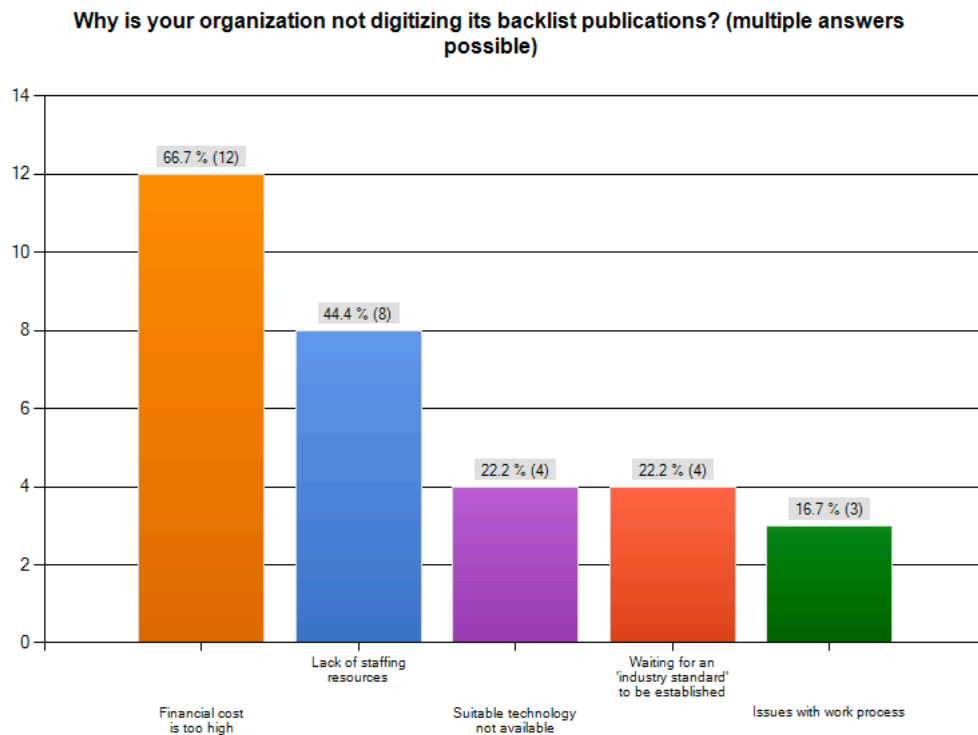
**Non-digitizing Publishers**

Only 14 of 576 publishers said that they were not digitizing (figure 4). When asked what their reasons and future plans were in this area the respondents said:

- 67 percent say the financial cost is too high and 44 percent lack of staffing resources

- 45 percent plan to begin digitizing within the next 5 years
- 47 percent have no plans to begin digitizing

All 14 of these publishers are very small organizations that publish less than 50 journals a year. One publisher commented that all of their journals have been digitized. However, the financial cost to bring the material online is holding them back.

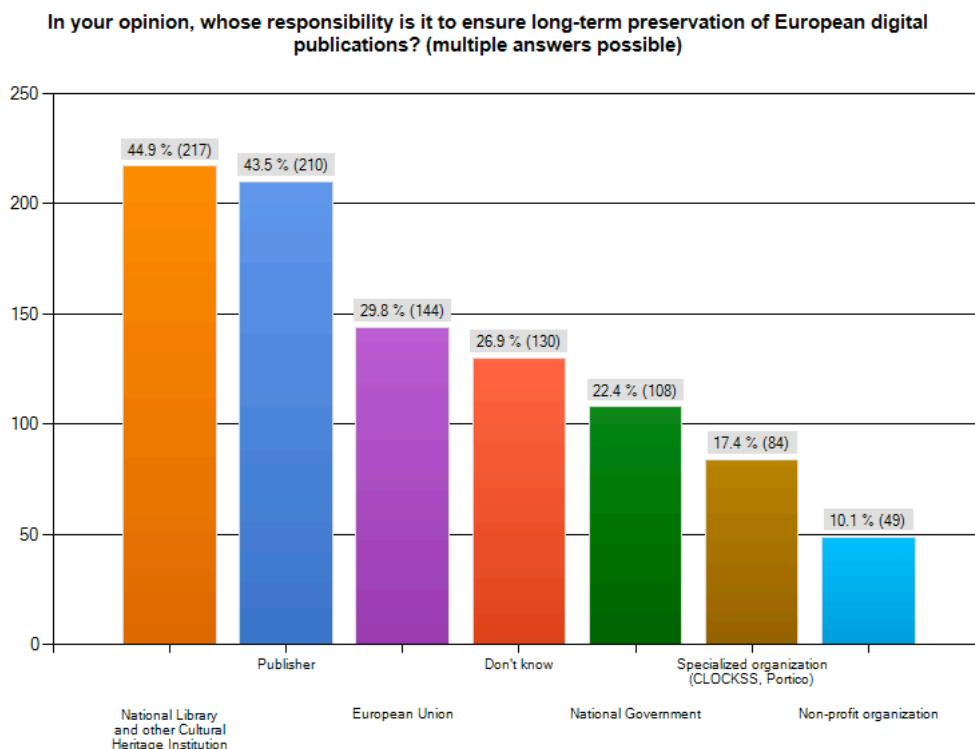


**Figure 4:** *Why not digitizing*

## LONG-TERM PRESERVATION

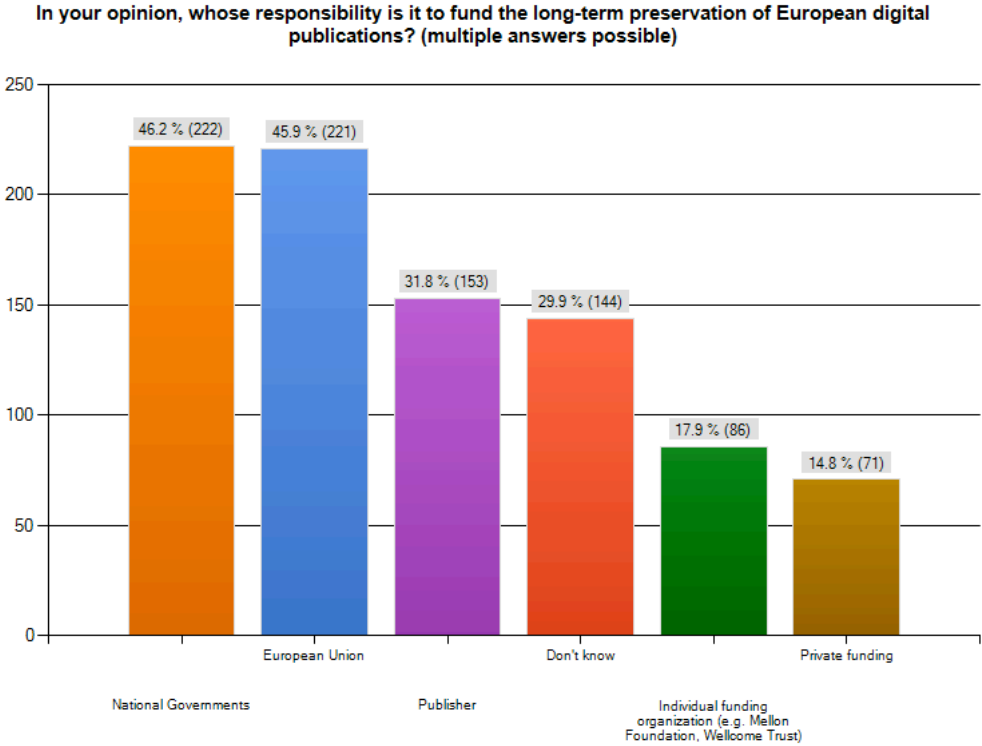
For the purpose of this survey, long-term preservation was defined as all measures that have to be taken to maintain digital objects as accessible and usable for an indefinite period. Of the publishers surveyed, 79 percent do not participate in a long-term preservation program. For the publishers that do participate in a long-term preservation program, LOCKSS was the most frequently used, but the KB, CLOCKSS, and Portico were also being mentioned. Less cited preservation programs were DOAJ, The German National Library, arXiv, and OJS.

When asked which aspects of a long-term digital archive are important ‘Long-term accessibility’ was considered *Very important*. In addition, ‘Generating metadata’ and ‘Developing a strategic plan’ were considered *Important* services. Forty-five percent of publishers believe that a national library or cultural heritage institution should be responsible for ensuring long-term preservation (*figure 5*). A near equal amount of publishers believe they should be responsible. Small publishers commented they would like to see larger publishers take the lead with ensuring long-term accessibility, as it would help all publishers.



**Figure 5:** Ensuring long-term preservation

National governments and the European Union were the favorites to fund long-term preservation in Europe, both receiving 46 percent of the votes (*figure 6*). Numerous publishers commented on how any funding for this venture would be welcome. They also believed collaboration among large publishers and other large institutions to fund preservation would be feasible.



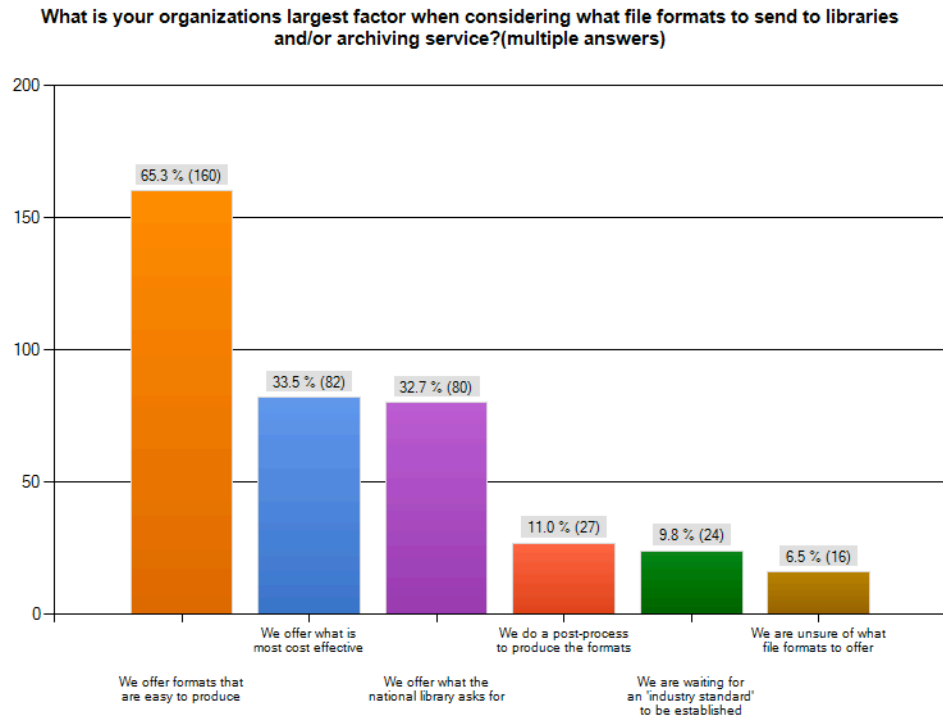
**Figure 6:** *Funding long-term preservation*

## **PUBLISHERS AND LIBRARIES**

The goal of this section was to serve as a needs assessment and satisfaction check for publishers. This is in respect to the long-term preservation services that libraries and archiving services provide.

Thirty-eight percent of publishers indicated that they are not using any type archiving service, legal deposit, or archiving agreement with a library. The 89 percent of publishers that are engaged in archiving agreements, indicated that they are satisfied with the process and that their expectations are met with their current agreements. A few publishers commented that they were unsure why they did not have an archiving agreement. One publisher mentioned that without the proper assistance it is too difficult to understand an agreement of this type.

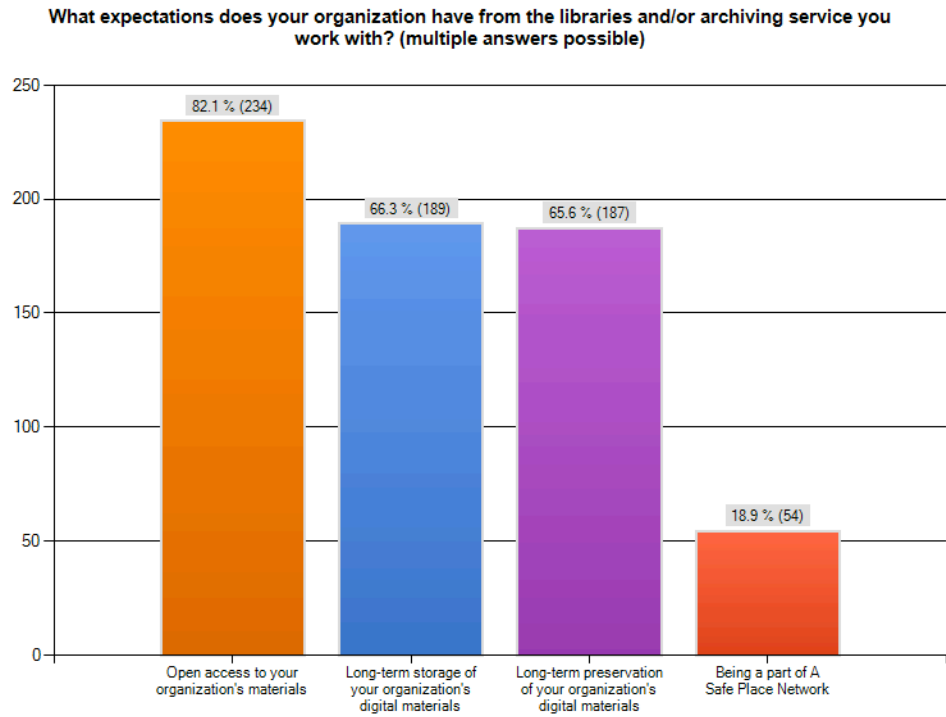
The 51 percent of publishers that are using archiving agreements believe their archiving agreement clearly states the file formats and metadata standards they are required to send, and 63 percent of publishers only send what is required of them. When asked what their largest factor was when considering what file formats to send to libraries or archiving services (*figure 7*), 65 percent explained that they offer formats that are easiest to produce. Offering the format the national library requests and also what is most cost effective were common with 33 percent of the responses.



**Figure 7:** *File formats sent to libraries or archiving services*

Publishers gave a similar response when describing their relationship with the libraries or archiving services they work with. Publisher's perceive their relationship with these organizations as both mutually beneficial and cooperative. Only a small number viewed the relationship between themselves and libraries or archiving services as one of competition. All seven of these responses came from European publishers.

Open access is the single major expectation publishers have from libraries and/or archiving services (*figure 8*). For the most part, these respondents came were very small publishers that publish less than 50 journals a year and use an open access business model.



**Figure 8: *Publisher expectations***

## CONCLUSIONS

The results of the survey were analyzed as a whole, by continent, size of the organization, and the business model employed by each. Comparatively speaking, each group responded similarly. This result could be due to the globalization occurring in all industries around the world. The differences that do exist are small, and may not need any further investigation than to be mentioned here. One difference between the continents is that both North and South America have entirely digital collections. The two continents offer both born digital and digitized material, whereas, the publishers not currently digitizing are mostly located in Europe, with a few in Asia, Africa, and Australia. Another significant difference is the global location of publishers who view libraries as competitors. These particular publishers are located across the European continent.

Even though, the sheer number of publishers that responded to the survey is overwhelming, 67 percent came from very small open access publishers all around the world. This could be in part, because DOAJ was an active supporter of this survey. Their publishers responded in the greatest number and contributed to the international influence of the survey. These are an active group of publishers who want their voices to be heard. Many have a clear idea of what direction publishing is moving in and others are looking at larger publishers and institutions for answers. Nonetheless, they are working to stay relevant and sustainable.

Sound familiar? Libraries and publishers are in the same position, working hard to remain relevant and sustainable. A large amount of research has been dedicated to digitization and long-term preservation in the library and publishing worlds. The same conclusions and questions have been drawn repeatedly over the last few years. Yet, no action is taken, only more research. When will the time come for someone (publisher or library) to begin the process of testing the theories and results from these studies?

In a concurrent survey conducted by a fellow intern entitled, 'How libraries process their digitalized material,' national, state, and university libraries were asked about digitization and long-term preservation. Libraries and publishers were asked similar questions about digitization, file formats, metadata standards, and their audience. Below is a brief comparison of the results.

- Libraries are in the same position as publishers. The financial cost and lack of staffing resources are the main reasons for not digitizing.

- Libraries are also using MARC21, Dublin Core, EAD, or METS, and so are many publishers. This should make the process of implementing archiving agreements a straightforward task for both groups of publishers and libraries.

Libraries use PDF and JPEG for access and PDF and TIFF for long-term storage.

PDF is the most common file format for libraries and publishers.

- Libraries and publishers value the same characteristics when selecting file formats, ‘Openness’, ‘Simplicity’, and ‘Self-documentation’.
- Libraries and publishers serve the same populations but in very different ways. The main audiences for both organizations are scholars and students. This shared audience could be why some European publishers view libraries as competitors.

## **RECOMMENDATIONS**

The recommendations based on the survey results are offered as a guide for both libraries and publishers. This is a glimpse of the current process and the future direction of publishing organizations around the world.

### **Digital Material**

Libraries and publishers are using similar file formats and metadata standards. This is a good sign for creating the possibility of libraries and publishers working more closely together in the future. There is a concern that very small publishing organizations are not digitizing because of the cost and staffing issues. An even greater concern, are the 47 percent of publishers who say they have no plans to digitize. As more and more journals are born digital and/or become digitized, competition among the publishers is increasing. Non-digitizing publishing organizations are in danger of becoming extinct if they are not capable of keeping up with their competition.

### **Long-Term Preservation**

According to this survey, most publishers believe that it is the responsibility of national libraries and cultural heritage institutions to ensure long-term preservation to European digital publications. The publishers also believe that they themselves are equally responsible. Does this mean that a door is open between national libraries and publishers to work together? An almost equal result labeling both parties responsible could signify that similar ideas of what needs to be accomplished in the future is on everyone's mind. The issue of who should fund the long-term preservation in Europe, either the national governments and/or the European Union, revealed little. These results were in line with answers given in other surveys on this topic.

### **Libraries and Publishers – A Needs Assessment**

Most of the publishers surveyed indicated their expectations were being met with their current Archiving Agreement and they know what metadata and file formats are being requested by these particular institutions. The majority of publishers said they only send the file formats and metadata standards that the library is requesting. However, later in the survey the majority indicated the publishers offer libraries and archiving services the format that is easiest to produce. There seems to be a discrepancy between what publishers say they are sending and what they actually do send. It is interesting to note that offering the file formats and metadata standards that the archiving agreement requests were not the most popular choice. Lastly, the majority of respondents said 'Open access' to their organization's materials is the expectation of both the libraries and archiving services they work with.

Unfortunately, 38 percent of publishers do not use any form of archiving service. Based on the following comments left by publishers when asked if they send copies of their born digital or digital publications to an archiving service or a legal deposit, it appears that many do not know what an archiving agreement/service is or why these groups are important:

- We insert them (*copies of their publications*) in our website
- No, we do not sell them; we provide them for open access.
- Google Book Search
- I am not sure what you mean here
- I guess I should be talking to BLL/Library of Congress

Libraries could begin educating publishers at appropriate seminars and conferences on the benefits of having an archiving agreement with a national library or archiving service, especially if it is apart of a Safe Place network.

Twenty-one percent of publishers claim their needs are not being met with their archiving agreement. A few publishers shared their reasons why:

- We never seem to hear back from LOCKSS (either UK or US) about our application to become a participant
- CLOCKSS, LOCKSS, PORTICO, etc. are still more promising than actual industrial-grade solutions - CLOCKSS and LOCKSS in particular suffer from being under the heavy hand of Stanford
- Very complicated to implement, understand, and/or use (*This publisher has an agreement with a library*)
- No automatic solution for long-term archiving
- No institutional repository in place

The discrepancies in file formats and metadata requested by a library or archiving service compared to what the publishers are actually sending could be another possible reason for issues. Following up with publishers who have signed an archiving agreement is good customer service. This service gives both parties a chance to evaluate satisfaction levels and places significant value on this working relationship. This process should also ensure that the library or archiving service receives the file formats and metadata standards it has requested.

I would like to thank the Koninklijke Bibliotheek, The National Library of the Netherlands for the opportunity to create and distribute this survey, and all of the wonderful publishers who gave their time to participate in this survey and leave their invaluable comments.