

## Born Digital Legal Deposit Policies and Practices

**Frederick Zarndt**

Digital Divide Data, Coronado CA 92118 USA.

[frederick@frederickzarndt.com](mailto:frederick@frederickzarndt.com)

**Dorothy Carner**

University of Missouri-Columbia, Columbia MO 65211, USA.

[carnerd@missouri.edu](mailto:carnerd@missouri.edu)

**Edward McCain**

Donald W. Reynolds Journalism Institute, University of Missouri-Columbia, Columbia MO 65211, USA. [mccaine@rjionline.org](mailto:mccaine@rjionline.org)<sup>1</sup>



Copyright © 2017 by Zarndt, McCain, Carner. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

---

### Abstract:

*In 2014, the authors surveyed the born digital content legal deposit policies and practices in 17 different countries and presented the results of the survey at the 2015 International News Media Conference hosted by the National Library of Sweden in Stockholm, Sweden, April 15-16, 2015.<sup>2</sup> Three years later, the authors expanded their team and updated the survey in order to assess progress in creating or improving national policies and in implementing practices for preserving born digital content. The 2017 survey reach has been broadened to include countries that did not participate in the 2014 survey.*

*To optimise survey design, and allow for comparability of results with previous surveys, the authors briefly review 17 efforts over the last 12 years to understand the state of digital legal deposit and broader digital preservation policies (a deeper analysis will be provided in a future paper), and then set out the logic behind the current survey.*

Keywords: e-legal deposit, survey, web archiving, digital preservation

---

<sup>1</sup> The authors would like to acknowledge the help of Olga Holownia, International Internet Preservation Consortium, and Stephen Wyber, IFLA, in preparing this paper, as well as the help of Abbie Grotke, Library of Congress, Nicola Bingham, British Library, Helena Byrne, British Library, and Wan Wong, National Library of Australia for their help in beta-testing the survey.

<sup>2</sup> Zarndt, Frederick; Carner, Dorothy & McCain, Edward, 2015, "An International Survey of Born Digital Legal Deposit Policies and Practices," [http://www.kb.se/dokument/utbildning/IFLA-KB-2015/13-2015\\_international\\_survey\\_of\\_born\\_digital\\_legal\\_deposit\\_policies\\_and\\_practices.pdf](http://www.kb.se/dokument/utbildning/IFLA-KB-2015/13-2015_international_survey_of_born_digital_legal_deposit_policies_and_practices.pdf) (paper) and Carner, Dorothy; McCain, Edward & Zarndt, Frederick. "An International Survey of Born Digital Legal Deposit Policies and Practices for News," 2014. Available at <https://www.slideshare.net/cowboyMontana/an-international-survey-of-born-digital-legal-deposit-policies-and-practices> (the accompanying PowerPoint slides).

---

In 2015, the *International survey of born digital legal deposit policies and practices* offered an overview of the current situation regarding e-Legal Deposit around the world. A paper was duly presented at the International Federation of Library Associations (IFLA) News Media section satellite conference in Stockholm, Sweden. Two years later, the authors wished to conduct another survey to examine the state of electronic legal deposit legislation worldwide, and ascertain what, if any, changes had taken place.

The resulting survey is a collaboration among four organizations with interest and expertise in questions around e-Legal Deposit and/or with broad membership networks including institutions working in the field, namely the International Federation of Library Associations and Institutions (IFLA), the International Internet Preservation Consortium (IIPC), the Donald W. Reynolds Journalism Institute, and the University of Missouri Libraries.

Bringing together a core group of people involved in each of these bodies, a steering group worked together to research existing surveys on e-Legal Deposit and broader digital preservation, establish a new survey to obtain the latest information on policies and practices, and carry out analysis of the results.

Each of the organisations involved has its own perspectives and priorities which needed to be taken into account. However, after a brief discussion amongst the collaborators, it was clear that a single survey focused broadly on national born digital legal deposit policies and practices followed by additional surveys focused on specific content types (news, audio-visual content, archived websites) or on particular technical approaches would provide the most comprehensive and useful answers, and result in the broadest participation. It is the broad survey on digital legal deposit policies and practices that is presented in this paper.

The survey questions are found at the below link and in Appendix 3. It is important to note that the survey is structured only to show some questions when there are positive answers to previous questions. The survey opened in mid-July 2017 and will remain open until the end of August 2017; the “live” survey is at <https://www.surveymzmo.com/s3/3651847/2017-digital-e-legal-deposit-survey>.

## **SUMMARY OF E-LEGAL DEPOSIT SURVEY 2014**

The 2014 survey was sent to specific individuals known to one of the authors (not randomly) at approximately 20 national libraries around the world. Replies to the survey were returned during the period from May 2014 to March 2015.

Here is a list of the libraries that responded to the survey:

Australia: National Library of Australia  
Croatia: Nacionalna i sveučilišna knjižnica u Zagrebu  
Denmark: Statsbiblioteket (Aarhus)  
Estonia: Eesti Rahvusraamatukogu  
Finland: Kansalliskirjasto  
France: Bibliothèque nationale de France  
Germany: Deutsche Nationalbibliothek

Latvia: Latvijas Nacionālā bibliotēka  
Luxembourg: Bibliothèque nationale de Luxembourg  
The Netherlands: Koninklijke Bibliotheek  
New Zealand: National Library of New Zealand  
Norway: Nasjonalbiblioteket  
Poland: Biblioteka Narodowa  
Singapore: National Library Board  
Sweden: Kungliga biblioteket - Sveriges nationalbibliotek  
Switzerland: Schweizerische Nationalbibliothek / Bibliothèque nationale suisse  
United States: Library of Congress

The survey consisted of two parts: Policies and Practices. The Policies section asked three questions about born-digital legal deposit laws or policies. The Practices section included six questions about implementation of those laws and policies. The complete survey can be found in the Appendix. In the 2015 survey report, individual responses from respondents in each country are also included in order to provide answers with greater depth.

In the 2015 report in Stockholm, the authors conclude that legal deposit laws vary widely from country to country. Nordic countries have been leaders in the capture of digital content, while many others still make no legal provision for collecting digital content. Overall, of the 16 countries surveyed, only seven had policies that addressed the deposit of born-digital content.

## **OTHER SURVEYS FROM THE PAST 12 YEARS**

To optimize the quality of responses to the questions posed in this inquiry, we identified 17 previous survey instruments that were used to assess digital preservation practices and policies for different types of content [and practices] over the previous 12 years. For the purpose of clarifying the query formats utilized and the information collected from these surveys, the authors place them into six categories: Audiovisual Preservation, Electronic Legal Deposit, Web Archiving, Digital Preservation of News, Preservation Standards and Best Practices, and National/Federal Policies. Within this structure, surveys are cited chronologically. A timeline is available below.

### **Survey Categories**

#### ***Audiovisual Preservation***

In 2007, the newly created International Federation of Library Associations (IFLA) Audiovisual Multimedia Section (AVMS) created a survey to identify which countries had policies for preserving audiovisual materials. The authors encountered significant challenges linked to the complexity of the issues surrounding AV preservation<sup>3</sup>. In 2010, IFLA AVMS regrouped and conducted another survey with the purpose of refining the 2007 effort<sup>4</sup>. The International Association of Sound and Audiovisual Archives (IASA) joined forces with

---

<sup>3</sup> Besser, Howard & van Malssen, Kara. August 12, 2010., Preliminary 2008-2009 Results for “AVMS Legal Deposit Survey.” <http://besser.tsoa.nyu.edu/howard/Talks/legal-deposit.pdf>.

<sup>4</sup> “AVMS IFLA Audiovisual & Multimedia Legal Deposit Survey” redux, 2010. <https://www.surveymonkey.com/r/7MQ89B7?sm=SYCHJCUfA2y91weXb8ZuTQ%3d%3d#q1> . “IFLA Audiovisual & Multimedia Legal Deposit Survey” <https://www.ifla.org/files/assets/avms/documents/legal-deposit-survey.pdf>.

IFLA AVMS to conduct another survey in 2016<sup>5</sup>. The jointly deployed survey's goal was "to create a new global register for legal deposit for audiovisual materials country by country." The register, published on the IASA website, is currently being crowdsourced.

Another audiovisual survey was conducted in 2008 by the Training for Audiovisual Preservation in Europe (TAPE) group. Funded by the Culture 2000 Programme of the European Union, the survey was focused on European collections, which were primarily in analog format with some content stored on disks or tape<sup>6</sup>. The responses indicated that the current preservation system was being overwhelmed by the exponential growth in the amount of content produced, as well as the lack of facilities and skilled professionals needed to manage the workflow.

### ***Electronic Legal Deposit***

In 2009 the British Library surveyed all members of the Conference of European National Libraries (CENL) to examine the status of electronic legal deposit legislation in those countries<sup>7</sup>. They repeated the survey effort in 2011<sup>8</sup>. The results indicated that electronic legal deposit laws were lagging behind those for print publications.

### ***Web Archiving***

In 2005 the International Internet Preservation Consortium (IIPC) conducted a survey to "identify and classify many of the conditions found on websites that influence the harvesting of content and the quality of an archival crawl."<sup>9</sup>

The National Library of the Netherlands conducted a web archiving survey in 2007. Assuming a user-centered approach, the focus was on access, with the central question: "What should the contents and search options of the web archive look like?"<sup>10</sup>

The IIPC conducted a Member Profiles Survey in 2008, with 35 of the 39 member institutions responding. The Questions were divided into two sections: "Part 1: About You and Your Web Archiving Activities" and "Part 2: About Your IIPC Participation: Your Contributions and Expectations." Three questions examined "Legal Issues and Policies," with the results indicating that 15.6% of respondents have legal authority related to web archiving."<sup>11</sup>

In 2013, the IIPC Preservation Working Group (PWG) surveyed the IIPC membership once again to better understand current web archiving practices<sup>12</sup>.

---

<sup>5</sup> Balberg, Trond & Ranft, Richard, "IASA-IFLA Legal Deposit Survey, 2016. <http://www.ifla-av-legal-deposit-form.iasa-web.org>.

<sup>6</sup> Edwin, Klijn & de Lusenet, Yola. "Tracking the Reel World. A Survey of Audiovisual Collections in Europe," 2008. [http://www.ica.org/sites/default/files/WG\\_2008\\_PAAG-tracking\\_the\\_reel\\_world\\_EN.pdf](http://www.ica.org/sites/default/files/WG_2008_PAAG-tracking_the_reel_world_EN.pdf)

<sup>7</sup> British Library, "International Survey to CENL on Legal Deposit," presented to CDNL, 2010. [http://www.cdnl.info/images/PDFs/CDNL\\_2010/CDNL\\_2010\\_BL\\_international\\_survey\\_on\\_e-Legal\\_Deposit.pdf](http://www.cdnl.info/images/PDFs/CDNL_2010/CDNL_2010_BL_international_survey_on_e-Legal_Deposit.pdf)

<sup>8</sup> British Library, "International Survey to CENL on Legal Deposit," presented to CDNL, 2011. [http://www.cdnl.info/images/PDFs/CDNL\\_2011/legaldeposit\\_20survey\\_20CDNL\\_20Slides\\_20Aug.pdf](http://www.cdnl.info/images/PDFs/CDNL_2011/legaldeposit_20survey_20CDNL_20Slides_20Aug.pdf)

<sup>9</sup> <https://web.archive.org/web/20170317153421/http://netpreserve.org/resources/web-harvesting-survey>

<sup>10</sup> "National Library of Netherlands Web Archiving Survey," 2007. [https://www.kb.nl/sites/default/files/KB\\_UserSurvey\\_Webarchive\\_EN.pdf](https://www.kb.nl/sites/default/files/KB_UserSurvey_Webarchive_EN.pdf)

<sup>11</sup> Grotke, Abigail. "International Internet Preservation Consortium 2008 Member Profile Survey Results", 2008. <https://web.archive.org/web/20160310155956/http://netpreserve.org/sites/default/files/resources/Membersurvey.pdf>

<sup>12</sup> Steinke, Tobias & Jones, Gina., "2013 International Internet Preservation Consortium (IIPC) Preservation Working Group (PWG) Survey on Web Archiving Practices." Results discussed in: Goethals, Andrea; Oury, Clément; Pearson, David;

The National Digital Stewardship Alliance (NDSA) Content Working Group (CWG) initiated a national (U.S.) survey in 2011 “to better understand the landscape of web archiving activities in the United States, including identifying the organizations or individuals involved, the types of web content being preserved, the tools and services being used, and the types of access being provided.”<sup>13</sup>

The NDSA CWG web archiving survey of 2013 sought to pose those questions once again, but additionally asked about overall policies related to archiving programs<sup>14</sup>.

By 2016 the NDSA web archiving survey’s additional goals were “to enable historic comparisons with the 2011 and 2013 surveys and inquire about program details not previously included” such as new archiving tools and on/off-site storage<sup>15</sup>.

### ***Digital Preservation of News***

In 2014, the Donald W. Reynolds Journalism Institute at the University of Missouri conducted a national (U.S.) telephone survey of news organizations with the purpose of looking at the kinds of born-digital content being created and the practises surrounding preservation of such content<sup>16</sup>.

Also in 2014, Zarndt, Carner & McCain deployed an email survey to cultural heritage organizations around the world, asking them to share their respective national born-digital legal deposit policies and practices for news content. The results were presented at the IFLA News Media section’s satellite meeting in Stockholm, Sweden the following year<sup>17</sup>.

### ***Preservation Standards and Best Practices***

In 2011, NDSA deployed another national (U.S) survey directed at organizations that were either engaged in or planning to archive content from the web. The goal was to get a snapshot of storage practices within the membership of NDSA<sup>18</sup>.

The IFLA Preservation Guidelines/Standards/Best Practices survey was conducted in 2016 with the goal to discover currently used preservation standards, guidelines and best practices for material in any format<sup>19</sup>.

---

Sierman, Barbara & Steinke, Tobias. “Facing the Challenge of Web Archives Preservation Collaboratively: The Role and Work of the IIPC Preservation Working Group”. D-Lib Magazine, May/June 2015. Vol. 21, Nr. 5/6  
<http://www.dlib.org/dlib/may15/goethals/05goethals.html>

<sup>13</sup> “National Digital Stewardship Alliance Web Archiving Survey Report,” Produced by the NDSA Content Working Group, June 2012. [http://www.digitalpreservation.gov/documents/ndsa\\_web\\_archiving\\_survey\\_report\\_2012.pdf](http://www.digitalpreservation.gov/documents/ndsa_web_archiving_survey_report_2012.pdf)

<sup>14</sup> Bailey, Grotke, Hanna, Hartman, McCain, Moffatt, Taylor. “Web Archiving in the United States: A 2013 Survey.” An NDSA Report, September 2014.  
[http://ndsa.org/documents/NDSA\\_USWebArchivingSurvey\\_2013.pdf](http://ndsa.org/documents/NDSA_USWebArchivingSurvey_2013.pdf)

<sup>15</sup> Bailey, Grotke, McCain, Moffatt, Taylor. “Web Archiving in the United States: A 2016 Survey” An NDSA Report, February 2017.  
[http://ndsa.org/documents/WebArchivingintheUnitedStates\\_A2016Survey.pdf](http://ndsa.org/documents/WebArchivingintheUnitedStates_A2016Survey.pdf)

<sup>16</sup> Carner, McCain & Zarndt. August, 2014., “Missing Links: The digital News Preservation Discontinuity,” [https://www.ifla.org/files/assets/newspapers/Geneva\\_2014/s6-carner-en.pdf](https://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-carner-en.pdf)

<sup>17</sup> Zarndt, Frederick; Carner, Dorothy & McCain, Edward, 2015, “An International Survey of Born Digital Legal Deposit Policies and Practices,” [http://www.kb.se/dokument/utbildning/IFLA-KB-2015/13-2015\\_international\\_survey\\_of\\_born\\_digital\\_legal\\_deposit\\_policies\\_and\\_practices.pdf](http://www.kb.se/dokument/utbildning/IFLA-KB-2015/13-2015_international_survey_of_born_digital_legal_deposit_policies_and_practices.pdf)

<sup>18</sup> Altman, Micah; Bailey, Jefferson; Cariani, Karen; Gallinger, Michelle; Ows, Trevor, 2012, “Data for NDSA Storage Report.” <https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/19768>

## National / Federal Policies

In 2016, the United Nations Educational, Scientific and Cultural Organization (UNESCO) Platform to Enhance and Reinforce the Sustainability of the Information Society Trans-globally (PERSIST) conducted a survey with multiple goals<sup>20</sup>, including:

- Global overview of current policies and/or strategies in UNESCO member states
- Assess the role and involvement of governments in long-term digital preservation
- Give insight into the implementation of those strategies and policies
- Give a short description of some selected examples

### Timeline

	Audiovisual Preservation	E-Legal Deposit	Web Archiving	Digital News Preservation	Preservation Standards & Best Practices	National Policies & Strategies
2005			<a href="#">IIPC Web Harvesting</a>			
2006						
2007	<a href="#">IFLA AVMS</a>		<a href="#">NL Netherlands</a>			
2008	<a href="#">TAPE - EU</a>		<a href="#">IIPC Member Profile</a>			
2009		<a href="#">BL</a>				
2010	<a href="#">IFLA AVMS</a>					
2011		<a href="#">BL</a>	<a href="#">NDSA</a>		<a href="#">NDSA</a>	
2012						
2013			<a href="#">IIPC-PWG NDSA</a>			
2014				<a href="#">RJI Zarndt, Carner McCain</a>		
2015						
2016	<a href="#">IFLA AVMS &amp; IASA</a>		<a href="#">NDSA</a>		<a href="#">IFLA</a>	<a href="#">UNESCO PERSIST</a>

<sup>19</sup> "IFLA Survey on Preservation and Conservation Guidelines/Standards/Best Practices," 2016. <https://f.hypotheses.org/wp-content/blogs.dir/2696/files/2016/09/IFLAPreservationSurvey.pdf>

<sup>20</sup> Brungs ,Julia; Marz, Vera & de Niet, Marco. "Workshop of the UNESCO PERSIST Content and Best Practices Working Group". Frankfurt am Main, 23-24 February 2017. <https://unescopersist.files.wordpress.com/2017/04/persist-cbp-frankfurt-workshop-report.pdf>

## **What have we learned?**

For the purposes of this paper, the authors examine only the issues surrounding the execution of the survey itself, not the results of the previous surveys. We intend to delve into a deeper analysis of the 17 previous surveys in a future report.

Based on the structure and execution of the previous surveys, several themes emerged:

*Keep it simple!* We chose to streamline the queries and potential responses. That meant asking simple “yes” or “no” questions as often as possible and using multiple choice formats for most of the remaining situations. We also provided free text fields for many questions, which allowed for a variety of responses and valuable feedback suggesting potential changes to future surveys. We noticed that authors of previous efforts commented that posing complicated questions to respondents sometimes made it difficult for survey takers to answer and (according to their analysis) resulted in lower participation rates.

*Use professional tools:* Rather than utilizing email for delivering the survey, we used an easily accessible online survey instrument, SurveyGizmo, that allowed for the use of question skip logic, letting respondents skip irrelevant questions, based on previous answers.

*Use multiple angles to approach respondents:* In order to improve response rates, we chose to solicit responses through multiple channels. Initially, personal emails to individuals asking for people inside the organization to pass the survey along to the appropriate individuals, were used. Future options include soliciting responses through targeted email lists and blogs, posting on websites and newsletters may also be deployed if the deadline is extended. This paper and the associated sessions at the IFLA News Media Satellite meeting in Dresden, as well as at the World Library and Information Congress 2017, 19-25 August, in Wroclaw, Poland, will also help identify and mobilise responses.

*Use a language people understand:* We realize that there is a significant shortcoming in our process in that we are only collecting responses in English, which, we assume, reduces our sample. Resources permitting, it would be desirable to provide versions of the survey in several of the most-used languages, either the United Nations languages (English, French, Chinese, Russian, Arabic, Spanish) or the IFLA languages (all the UN languages plus German).

## **CONCLUSION**

With digital media becoming ever more dominant in cultural and scientific production, the need to adapt approaches to preservation becomes more pressing. The particular features of digital production – speed, cross-border collaboration, volume – require innovative and effective responses. In order to avoid gaps in the historical record, creating digital equivalents of practices such as legal deposit is essential.

This is a fast-moving issue. Active collection and preservation of born digital content, including audiovisual, news media and websites, is still relatively new to many countries around the world. Policies, practices and tools continue to evolve, as countries and institutions seek to find better means of preserving digital materials. What is collected, and how, varies strongly between countries, although there is doubtless progress towards more effective solutions.

As this paper has underlined, there has also been considerable investment in trying to understand the landscape of digital legal deposit and related policies. The 17 surveys reviewed in this paper indicate that people and organizations are concerned about this issue, and a number of institutions are pioneering new and interesting approaches from which others could learn.. These show that it is possible to bring together the necessary resources and expertise to make a difference. They also show that memory institutions such as libraries are ready to act, and serve to mobilise similar efforts from content producers and governments.

In order to support this continued exchange of ideas and practices, the survey introduced in this paper will not only update our current understanding of digital legal deposit policies and practices, but could be repeated on a regular basis in order to track progress. It can also be adapted to explore practices for specific types of content or other areas of digital preservation activity.

Clearly coordinating future survey would help eliminate unnecessary duplication of effort as this effort moves forward. There would also be merit in both bringing together available results in a single place (not least the full results of this survey, once published), as a first step towards supporting research and the sharing of experience and practice. Such an initiative could also favour greater comparability between survey results over time, allowing for the identification of trends. IFLA and the IIPC may want to consider creating such a single home for results in order to support this goal.



## Appendix 1: 2017 e-Legal Deposit Survey Preliminary Results

As of 11 August 2017, 19 organizations have completed the survey, each organization from a different state or country. Thanks to each of the respondents!

Australia: National Library of Australia  
Austria: Österreichische Nationalbibliothek  
Croatia: Nacionalna i sveučilišna knjižnica u Zagrebu  
Denmark: Det Kgl. Bibliotek  
Estonia: Eesti Rahvusraamatukogu  
Finland: Kansalliskirjasto  
France: Bibliothèque nationale de France  
Germany: Deutsche Nationalbibliothek  
Germany: State Parliament of Hamburg, Information Service  
Iceland: National and University Library of Iceland  
Latvia: National Library of Latvia  
New Zealand: National Library of New Zealand  
Norway: National Library of Norway  
Portugal: Fundação para a Ciência e Tecnologia Portugal  
Singapore: National Library Board Singapore  
Slovenia: Narodna in univerzitetna knjižnica  
Switzerland: Schweizerische Nationalbibliothek / Bibliothèque nationale suisse  
The Netherlands: Koninklijke Bibliotheek  
United States: Library of Congress

The “raw” results of the survey are summarized below. In a subsequent paper, the authors will analyse the “raw” results in detail. A complete list of the survey questions is found in an Appendix.

### **Part 1/2: Policies for e-Legal deposit of digital content**

1. Does your country / state have a legal deposit law?
- Yes
  - No, but my organization collects digital publications anyway
  - No
  - I don't know

17 of 19 respondents answers *Yes*, 2 answered *No, but my organization collects digital publications anyway*.

#### **Sample Comments**

The legal deposit law is only for physical items right now. We are reviewing to extend it to include digital content. We do collect digital deposits on a voluntary basis from publishers currently and encourage publishers to do so even though it is not required by law.

The situation in our country has not really changed since the 2014 survey. The law does not give a publisher an obligation to deposit any digital works on its own initiative, but if the National Library makes a request to the publisher, the publisher shall be under an obligation to comply.

We have individual contracts with publishers in order to collect their digital publications. For web archiving we use some sort of Fair Use approach.

2. Does the legal deposit law cover digital works?

- Yes
- No
- I don't know

14 of 19 respondents answered *Yes*, 3 respondents answered *No*

3. Do the laws of your country / state **require** publishers to legally deposit digital works? In this case we mean that publishers **MUST** send digital works to one or more legal deposit authorities.

- Yes
- No
- Sometimes
- I don't know

6 of 19 respondents answered *Yes*, 3 answered *No*, and 4 answered *Sometimes*.

### Sample Comments

Digital works published on the internet (public electronic network) must be made accessible to the library for download (even if behind a paywall). Publishers do not need to 'send' them in. Digital publishing on a physical media (e.g. DVD) is subject to deposit.

In 2016 our library welcomed long-anticipated changes to the copyright law. For the first time in its history, the Library could at last collect electronic publications under the legal deposit provisions of the law. Legal deposit provisions were extended to cover the online publishing landscape. This includes all national print and electronic books, journals, magazines, newsletters, reports, sheet music, maps, websites and public social media.

Our legal deposit act distinguishes between electronic publications (published on a physical carrier - those include sound recordings on vinyl, cassettes, compact discs and mini discs; video recordings on video cassettes, video discs and interactive compact discs; and software on floppy discs, discs and compact discs) and online publications, created in our country and containing textual, visual and audiovisual information (including of limited accessibility), which essentially means web pages. Electronic publications have to be deposited at the National Library, online publications are to be harvested by the National Library (and publishers must provide an access for harvesting to the publications of limited accessibility).

Our law permits us to take a copy, and, if we need it, to require the publisher's assistance in doing this. This contrasts with the situation for physical format items, where the obligation is on the publisher to deposit.

The legal deposit law from the 80s in general does NOT cover digital works. However, there is decree-law from 2006, that extends the legal deposit to also include MsC and PhD theses in digital format. Our organization also manages the network of repositories that preserve the theses.

4. Do the laws of your country / state require cultural heritage institutions (libraries) to harvest websites and webpages that are publicly available (not behind a subscription paywall)?

- Yes
- No
- Only for some websites and webpages
- I don't know

14 of 19 respondents answered *Yes*, 5 answered *No*.

### **Sample Comments**

"Require" isn't quite the right word - we have the right to copy, but the intention of the legislation was to be selective rather than comprehensive in digital collecting.

Our law says that we next to printed information we also shall collect, describe, disseminate and archive information on other carriers than paper. This includes digital information. But websites are not mentioned literally.

Under the legal deposit provisions in the Act, the National Library requests the delivery of online material through the process known as web crawling or web harvesting. This process uses harvesting robots to initiate requests to the web servers delivering online content using the HTTP protocol 'Get' request process.

5. Do the laws of your country / state require cultural heritage institutions (libraries) and publishers of websites and webpages to cooperate in order to preserve digital works when these works are behind a subscription paywall?

- Yes
- No
- Only some publishers
- I don't know

11 of 19 respondents answered *Yes*, 7 answered *No*, and 1 answered *I don't know*.

### **Sample Comments**

If it is not possible to make a copy of the web publication upon web archiving from the web, the National Library shall submit a request to the publisher to submit the copy and the publisher is required to enable making a copy.

Publishers are required to make this material accessible. There are no penalties outlined for non-compliance. We have not actively pursued this as most material of interest is not behind a paywall.

Recent changes allow this to happen, but the National Library has been focusing online books and serials and harvesting open access websites. Work to explore how to harvest material behind a paywall will commence in 2018.

The Law says that the National Library can harvesting and preserving websites, and are not required to notify the website owners in advance. We can ignore robots.txt files to make sure we get everything, but prefer and initiate a Cooperation between Publishers and the National Library.

### **Part 2/2: Practices for e-Legal deposit of digital content**

6. Does your library receive digital works from publishers? For this question by "receive" we mean that publishers initiate the transmission of digital works to the legal deposit authority (library). In tech speak, the publisher "pushes" the works to the authority (library).

- Yes
- No
- I don't know

14 of 19 respondents answered *Yes*, 5 answered *No*.

7. If publishers "push" digital works to libraries, how do you receive them?

- FTP
- RSS
- email
- Content delivered on physical storage device (hard drive, thumb drive, etc)

- Shared folder in the Internet cloud
- Other

10 of 19 respondents answered *FTP*, 1 answered *RSS*, 4 answered *email*, 7 answered *Content delivered on physical storage device*, 3 answered *Shared folder*, and 10 answered *Other*.

	Percent	Count
FTP	71.4%	10
RSS	7.1%	1
email	28.6%	4
Content delivered on physical storage device (hard drive, thumb drive, etc.)	50.0%	7
Shared folder in the Internet cloud	21.4%	3
Other	71.4%	10

#### **Sample Comments on *Other***

OAI-PMH (metadata with direct links to files, so actually more pull than push); web form with upload

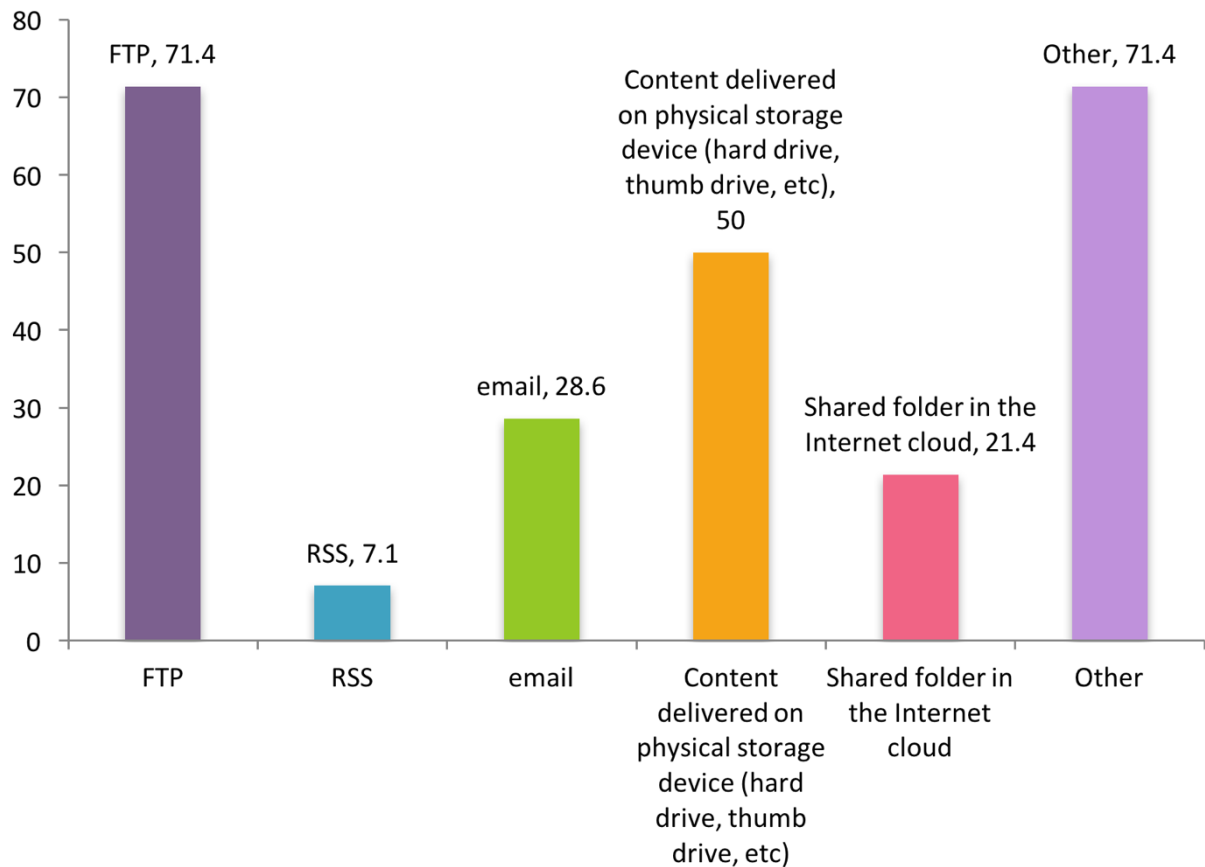
Publishers upload the file in preferred formats via our deposit website.

The Library has developed an e-deposit portal that enables publishers to deposit digital publications with the Library. The Library is also developing secure FTP sites for publishers to provide ONIX metadata, digital object and cover art. This material is then ingested into National Library systems and made accessible.

We have a deposit webpage enabling upload over HTTP(S). We've also done some custom arrangement when needed.

We have produced a special deposit interface for e-publishers

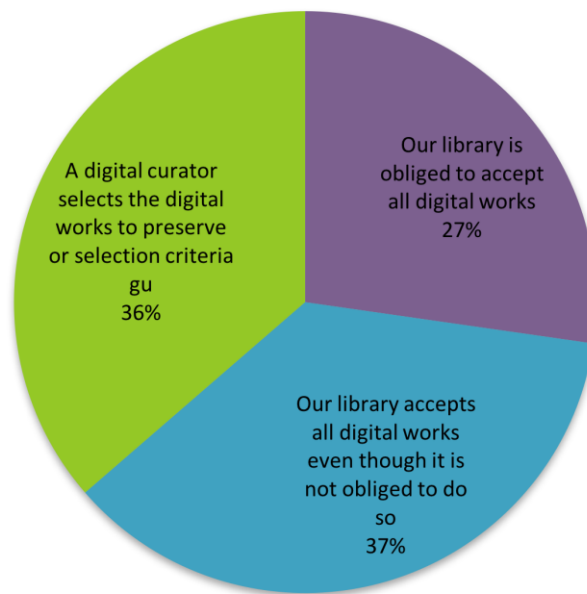
Web application called Publishers Portal



8. If publishers “push” digital works to libraries, how does your library decide which publishers?

- Our library is obliged to accept all digital works
- Our library accepts all digital works even though it is not obliged to do so
- A digital curator selects the digital works to preserve or selection criteria guide which digital works to preserve.

3 of 19 respondents answered *Our library is obliged to accept all digital works*, 4 answered *Our library accepts all digital works even though it is not obliged to do so*, and 4 answered *A digital curator selects the digital works*.



Several respondents did not choose one of these options but instead answered this questions with the following comments.

### Sample Comments

As digital works which are neither webpages, nor published on a physical carrier (e-books, for example) fall in a somewhat grey area of the law (it can be argued that they are online publications, but in this case we should be harvesting them, not accepting via file transfer), we are accepting them on case to case bases, mostly via formal agreements (for example, to receive print files of newspapers) or informal agreements (for example, to receive e-books and other digital publications) with publishers. Therefore we accept any kind transfer method, and most types of formats.

If it's a publisher we have a contract with he can use FTP to send publications - no need to decide anymore if we collect the content or not. If a publisher uses the web platform to upload a digital book a curator is checking if the content belongs into our collection.

Legal deposit applies to online and offline publications. Our e-deposit service is available for published works. A work is published if it is made available to the public for sale or for free. This includes websites, books, journals, sheet music, maps, magazines and newspapers. Work that is deposited that is out of scope can be rejected. Bulk deposit methods are available upon agreement/negotiation with publishers. The Library is an approved channel within CoreSource and this provides one method of bulk deposit. We are currently developing others.

We've reached out to certain publishers holding material that we are especially keen on getting in a digital format. Additional deposits made by individual publishers are also accepted, pending curator approval.

9. In what format(\*s) does your library accept digital works?

- EPUB
- PDF (any type)
- MOBI
- TIFF
- JPEG
- Open Doc
- Other

	Percent	Count
EPUB	85.7%	12
PDF (any type)	100.0%	14
MOBI	21.4%	3
TIFF	64.3%	9
JPEG	57.1%	8
Open Doc	14.3%	2
Other	64.3%	9

Several respondents choose *Other* and listed the following formats.

<b>Sample of Other Responses</b>
MP3, MP4
PDF 2000 goes for the e legal deposit newspapers.
PNG, and any other.



The e-deposit service accepts EPUB, PDF or mobi files for books, journals, magazines, newsletters and music scores. Our preference is epub. The service accepts PDF, GeoPDF, TIFF, or GeoTIFF files for maps. We do not accept Word documents. For cover art publishers can upload JPG, JPEG, TIF or TIFF cover images with an RGB colour profile. Files must be under 250 MB. While the edeposit system will accept images with a CMYK colour profile, we cannot currently display them online.

We are reasonably format neutral, and will accept all formats.

Websites contain all sort of formats. We store in ARC at the moment.

Word; (as annex files we accept also other formats like video, sound and picture files)

XML (e.g. JATS/NLM-DTD), HTML

ZIP for HTML and packages with attachments (research data)

10. Does your library offer a batch or bulk legal deposit service to publishers?

- Yes
- No
- I don't know

9 of 19 respondents answered *Yes*, 5 answered *No*. In addition, several respondents provided the following details.

### Sample Responses

Bulk deposit methods are available upon agreement/negotiation with publishers. The Library is an approved channel within CoreSource and this one method of bulk deposit. We are currently developing others.

But only upon request. Usually the library is the one requesting.

E.g. If we receive older vintage newspapers we offer them the digitised files in return when they are done. If we regularly receive New Newspapers they are made available to them on our website, this also goes for the local Libraries around the country.

Publishers can deliver in bulk to us if they find it inconvenient to dispatch items frequently.

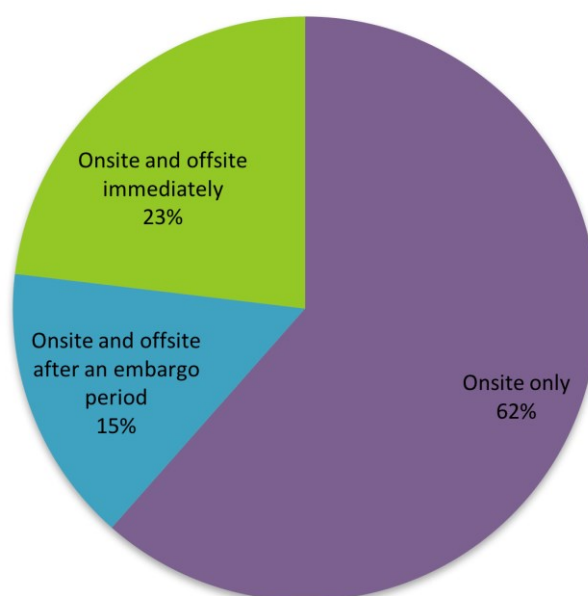
This option should become available in a near future when we will introduce a new web interface for submitting digital copies.

We already announced the possibility to serve our publishers as digital archive but they seem not to be interested or have another solutions.

11. What type of access do you provide to e-legal deposit digital content? For this question, **onsite** means within the library premises or on networks controlled by the library. **Offsite** means outside of the library premises and on networks not controlled by the library. **Embargo** means the period of time, usually specified by the publisher, for which access to the content is either limited or denied.

- Onsite only
- Onsite and offsite after an embargo period
- Onsite and offsite immediately
- Content can be freely downloaded

	Percent	Count
Onsite only	61.5%	8
Onsite and offsite after an embargo period	15.4%	2
Onsite and offsite immediately	23.1%	3



Several responded *None of the above* and made the following comments:

### Sample Responses

According to the legal deposit it is onsite. But the right holders can grant us the right to give offsite access and for these publications we offer offsite access.

Access depends on negotiations with publishers since we don't have a legal deposit. To some digital content we can give free access incl. downloads, some content can only be looked at onsite and some are under an embargo.

By the law publisher has right to assign the type of access. All above types are possible.

Everything is available onsite, and some selected e legal deposit content is available offsite. The off site material is made available based on agreements signed between the National Library .

If a publisher makes the content freely available to the public, without restrictions on its use or access by members of the public, the Library can do the same. For other material we can make up to three copies available in our Reading Room, on computers from which the content can't be printed, downloaded, emailed, etc.

Onsite-only access is currently limited to 2 dedicated PCs, with no download/upload capability. Print only.

The web archive is freely accessible onsite and offsite immediately. We have an agreement with the national research agency to provide access to all publicly co-financed publications. If there is an embargo or copyright limit, the publications should be available in the library premises only.

12. Does your library harvest websites and webpages?

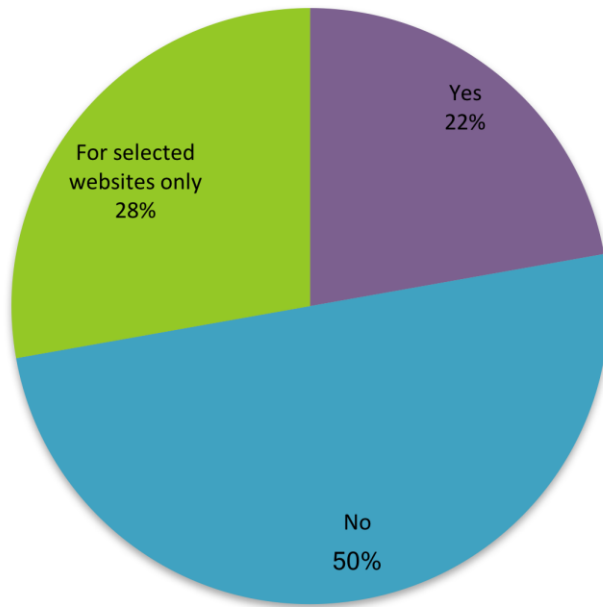
- Yes
- No
- I don't know

18 of 19 respondents answered *Yes*, only 1 respondent answered *No*.

13. If your library harvests websites and webpages, does this included those behind a paywall?

- Yes
- No
- For selected websites only

4 of 19 respondents answered *Yes*, 9 answered *No*, and 5 answered *For selected websites only*.



Respondents gave further detail as follows:

**Sample Responses**

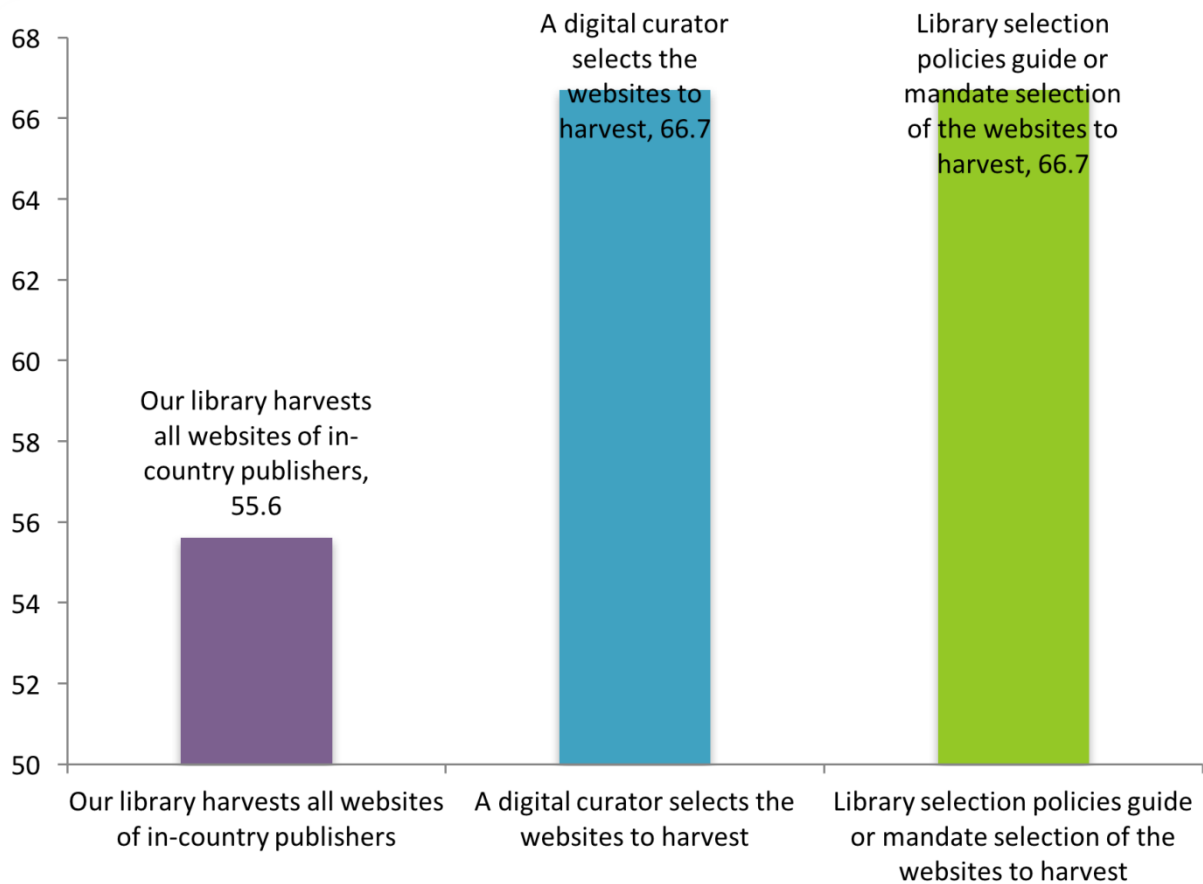
This is not a part of the harvesting going on as of today, but is being included in our next solution. It will be used on national or local newspapers web sites.

Where it is deemed to be of sufficient value to pursue the matter.

14. If your library harvests websites and webpages, what criteria are used to decide if born digital works from a particular published should be preserved?

- Our library harvests all websites of in-country publishers
- A digital curator selects the websites to harvest
- Library selection policies guide or mandate selection of the websites to harvest

	Percent	Count
Our library harvests all websites of in-country publishers	55.6%	10
A digital curator selects the websites to harvest	66.7%	12
Library selection policies guide or mandate selection of the websites to harvest	66.7%	12



In addition, the following comments and explanations were given:

**Sample Responses**

The main criteria is national author, national language or published nationally. For the thematic collection we have about 1375 websites that we harvest on a regular basis. For the domain based harvesting the number of seed URLs is 117,000.

Although we are entitled by law to harvest websites behind paywall, in reality we so far haven't requested access to any protected website.

We accept user suggestions.

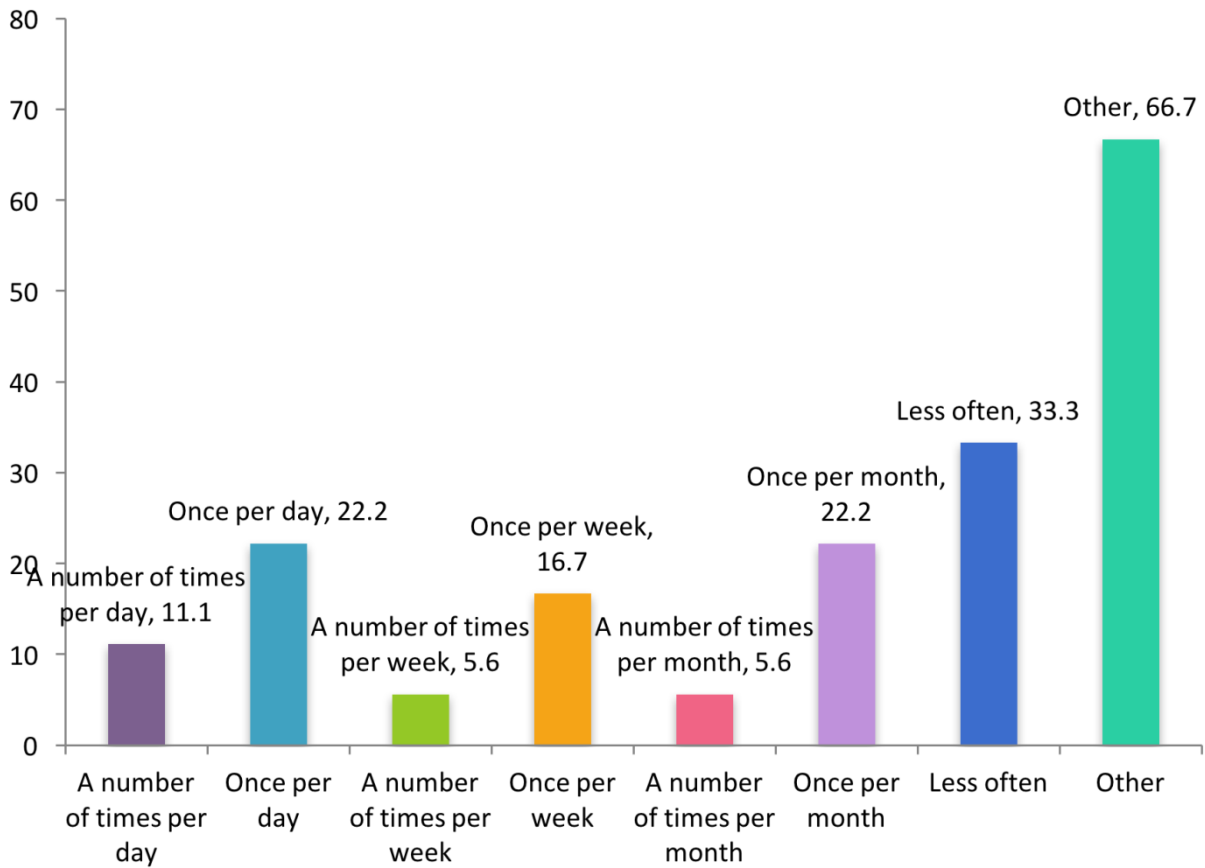
We did one national domain crawl and do selective crawls related to topics and events.

We harvest "everything" 4 times a year, and selected pages "all the time". On top of that we have curated harvestings of events - ie. elections.

15. If your library harvests websites and webpages (excluding digital news), how frequently does it harvest?

- A number of times per day
- Once per day
- A number of times per week
- Once per week
- A number of times per month
- Once per month
- Less often
- Other

	Percent	Count
A number of times per day	11.1%	2
Once per day	22.2%	4
A number of times per week	5.6%	1
Once per week	16.7%	3
A number of times per month	5.6%	1
Once per month	22.2%	4
Less often	33.3%	6
Other	66.7%	12



In addition, the following comments and explanations were given:

### Sample Comments

Default is twice a year, but this can be different for some websites. Event crawls are always individually configured.

Different materials have different timelines, in addition there are campaigns, e.g. elections etc.

It depends on the website, for example during our general election some sites will be harvested daily.

Once in four months for most websites and ad hoc for selected websites

Standard frequency is one per year, but the following options can be chosen as well: Twice a year, all 2 years, all 4 years, once only.

We do bulk harvesting of all websites under our national TLD and national language content under other TLDs once a year (we limit the size of data collected per site). Then we have a list of most valuable websites (meeting the specific selection criteria) that we harvest fully once a year. Thirdly, we harvest Twitter accounts of national politicians and government institutions a number of times per week/month depending on their tweeting-activity.

We harvest all websites 3x a year. Select websites are harvested more frequently

Selective crawls daily, domain crawl every 2 years

16. If your library harvests digital news websites and webpages, how frequently does it harvest?

- Library does not harvest digital news websites or webpages
- A number of times per day
- Once per day
- A number of times per week
- Once per week
- A number of times per month
- Once per month
- Less often
- Other

	Percent	Count
Library does not harvest digital news websites or pages	16.7%	3
A number of times per day	22.2%	4
Once per day	33.3%	6
A number of times per week	11.1%	2
Once per week	16.7%	3
A number of times per month	5.6%	1
Once per month	22.2%	4
Less often	11.1%	2
Other	50.0%	9



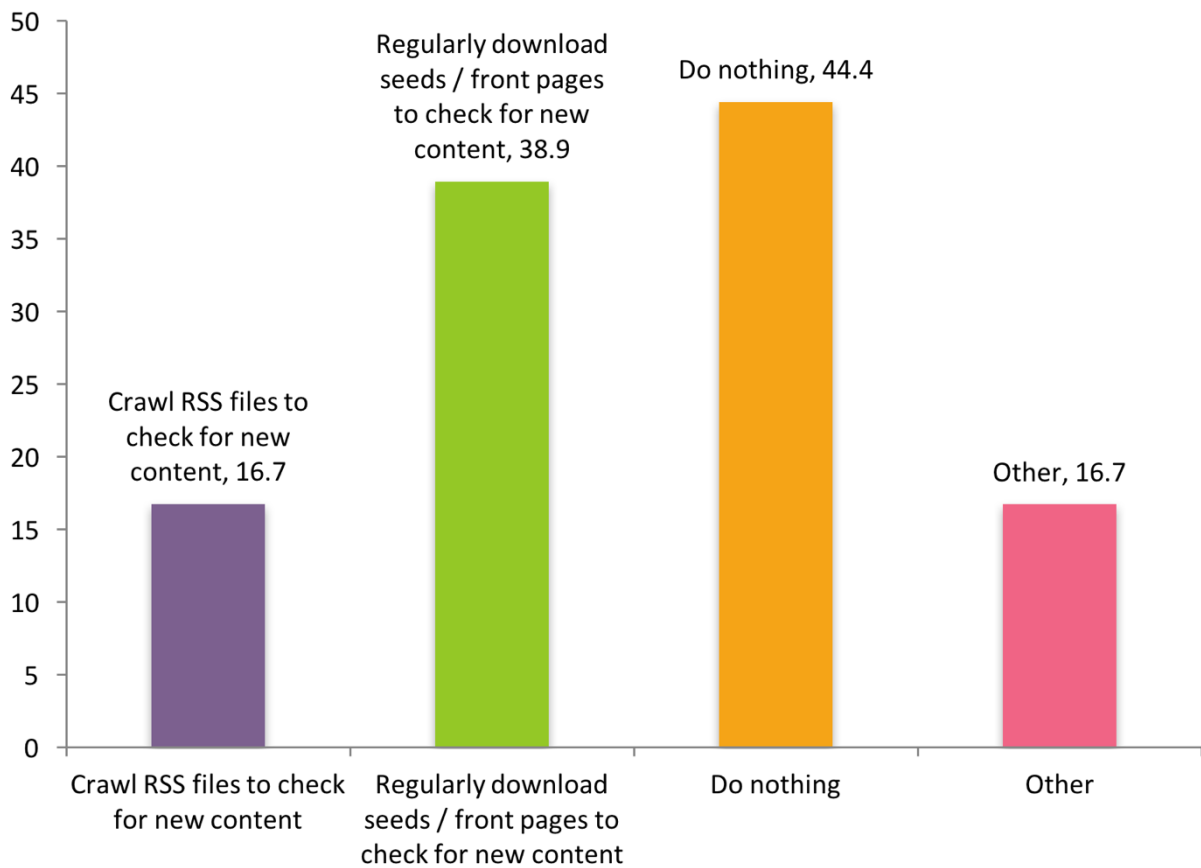
In addition, the following comments were made for *Other*:

<b>Sample Comments</b>
Ad hoc basis for selected content
Depending on the material
It depends on the complexity of the site whether and how often we harvest.
News are part of the selected pages “all the time”
The harvesting period depends on the type of serial publication. There are some titles that we have to collect several times a day, others are published daily, weekly or monthly.
We are still experimenting with news pages.
We harvest all websites 3x a year. Select websites are harvested more frequently. News websites are frequently chosen for more regular harvests.

17. Depending on the publisher, born digital content published on the web may be updated several times in an hour, day, or week. What methods does your library use to capture updated pages?

- Crawl RSS files to check for new content
- Crawl sitemaps to check for new content
- Regularly download seeds / front pages to check for new content
- Do nothing
- Other

	Percent	Count
Crawl RSS files to check for new content	16.7%	3
Regularly download seeds / front pages to check for new content	38.9%	7
Do nothing	44.4%	8
Other	16.7%	3



**Sample Responses**

Current harvesting policy focuses mainly of webpages of government agencies, cultural institutions and events and other socially or culturally important resources. News resources are not being harvested mainly because of their size, but also because they are deemed not important by the harvesting policy. For this reason only selected articles from news resources are preserved (mainly opinion articles). Most of the webpages are harvested once a year.

Regularly (once a day, several times a week, once a week, several times a month, once a month, etc.) harvests born digital content, especially news portals and websites.

Two-prong strategy collecting RSS feeds twice daily and home page and other content less frequently (monthly and/or quarterly).

We currently mostly do this manually, but we do use some of the methods above on a very few sites.

18. Does your library require preservation of its digital content?

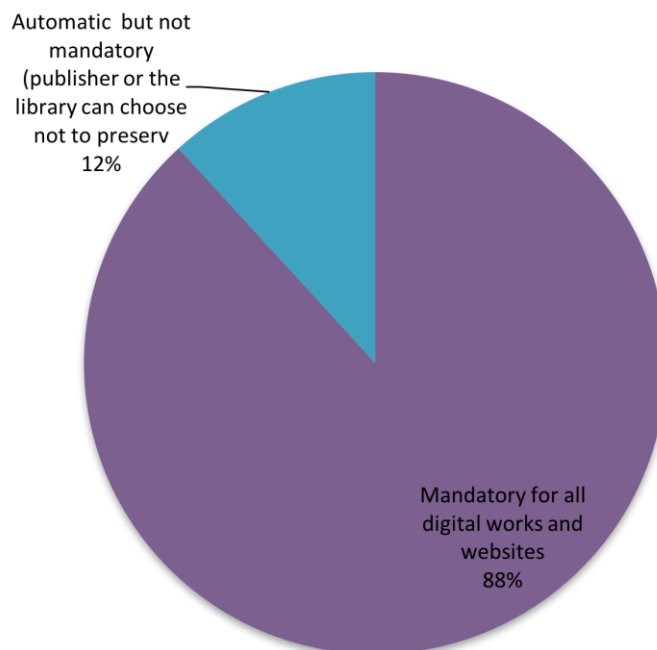
- Yes
- No
- I don't know

17 of 19 respondents answered *Yes* and 1 answered *No*.

19. At your library is digital preservation ....

- Mandatory for all digital works and websites
- Automatic but not mandatory (publisher or the library can choose not to preserve certain content)
- Optional

	Percent	Count
Mandatory for all digital works and websites	88.2%	15
Automatic but not mandatory (publisher or the library can choose not to preserve certain content)	11.8%	2



## Appendix 2: Earlier survey on e-legal deposit, digital preservation, and web archiving

[2007: IFLA AVMS](#)  
[2008: TAPE-EU](#)  
[2010: IFLA AVMS](#)  
[2016: IFLA AVMS & IASA](#)  
[2009: British Library](#)  
[2011: British Library](#)  
[2005: IIPC Web Harvesting](#)  
[2007: National Library of the Netherlands](#)  
[2011: NDSA](#)  
[2013: NDSA](#)  
[2013: IIPC\\_PWG](#)  
[2016: NDSA](#)  
[2014: Reynolds Journalism Institute](#)  
[2014: Zarndt, Carner & McCain](#)  
[2011: NDSA Infrastructure](#)  
[2016: IFLA Preservation Guidelines/Standards/Best Practices](#)  
[2016: UNESCO PERSIST National / Federal Policies and Strategies for Preservation of Digital Heritage](#)

## Appendix 3: 2017 e-Legal Deposit Survey

### **Part 1/2: Policies for e-Legal deposit of digital content**

1. Does your country / state have a legal deposit law?
  - Yes
  - No, but my organization collects digital publications anyway
  - No
  - I don't know
  
2. Does the legal deposit law cover digital works?
  - Yes
  - No
  - I don't know
  
3. Do the laws of your country / state **require** publishers to legally deposit digital works? In this case we mean that publishers **MUST** send digital works to one or more legal deposit authorities.
  - Yes
  - No
  - Sometimes
  - I don't know
  
4. Do the laws of your country / state require cultural heritage institutions (libraries) to harvest websites and webpages that are publicly available (not behind a subscription paywall)?
  - Yes
  - No
  - Only for some websites and webpages

- I don't know

5. Do the laws of your country / state require cultural heritage institutions (libraries) and publishers of websites and webpages to cooperate in order to preserve digital works when these works are behind a subscription paywall?

- Yes
- No
- Only some publishers
- I don't know

## **Part 2/2: Practices for e-Legal deposit of digital content**

6. Does your library receive digital works from publishers? For this question by "receive" we mean that publishers initiate the transmission of digital works to the legal deposit authority (library). In tech speak, the publisher "pushes" the works to the authority (library).

- Yes
- No
- I don't know

7. If publishers "push" digital works to libraries, how do you receive them?

- FTP
- RSS
- email
- Content delivered on physical storage device (hard drive, thumb drive, etc)
- Shared folder in the Internet cloud
- Other

8. If publishers "push" digital works to libraries, how does your library decide which publishers?

- Our library is obliged to accept all digital works
- Our library accepts all digital works even though it is not obliged to do so
- A digital curator selects the digital works to preserve or selection criteria guide which digital works to preserve.

9. In what format(\*s) does your library accept digital works?

- EPUB
- PDF (any type)
- MOBI
- TIFF
- JPEG
- Open Doc
- Other

10. Does your library offer a batch or bulk legal deposit service to publishers?

- Yes
- No
- I don't know

11. What type of access do you provide to e-legal deposit digital content? For this question, **onsite** means within the library premises or on networks controlled by the library. **Offsite** means outside of the library premises and on networks not controlled by the library. **Embargo** means the period of time, usually specified by the publisher, for which access to the content is either limited or denied.

- Onsite only
- Onsite and offsite after an embargo period
- Onsite and offsite immediately
- Content can be freely downloaded

12. Does your library harvest websites and webpages?

- Yes
- No
- I don't know

13. If your library harvests websites and webpages, does this included those behind a paywall?

- Yes
- No
- For selected websites only

14. If your library harvests websites and webpages, what criteria are used to decide if born digital works from a particular published should be preserved?

- Our library harvests all websites of in-country publishers
- A digital curator selects the websites to harvest
- Library selection policies guide or mandate selection of the websites to harvest

15. If your library harvests websites and webpages (excluding digital news), how frequently does it harvest?

- A number of times per day
- Once per day
- A number of times per week
- Once per week
- A number of times per month
- Once per month
- Less often
- Other

16. If your library harvests digital news websites and webpages, how frequently does it harvest?

- Library does not harvest digital news websites or webpages
- A number of times per day
- Once per day
- A number of times per week
- Once per week
- A number of times per month
- Once per month
- Less often
- Other

17. Depending on the publisher, born digital content published on the web may be updated several times in an hour, day, or week. What methods does your library use to capture updated pages?

- Crawl RSS files to check for new content
- Crawl sitemaps to check for new content
- Regularly download seeds / front pages to check for new content
- Do nothing
- Other

18. Does your library require preservation of its digital content?

- Yes
- No
- I don't know

19. At your library is digital preservation ....

- Mandatory for all digital works and websites
- Automatic but not mandatory (publisher or the library can choose not to preserve certain content)
- Optional

20. Where do you work (official organizational name)?

21. In which country do you work?

#### **Appendix 4: 2014 e-Legal Deposit Survey**

The survey questions from the 2014 survey were:

##### **Policies**

1. Do the laws of your country require publishers to legally deposit born digital news? In this case we mean that publishers **MUST** send born digital news to one or more legal deposit authorities.

2. Do the laws of your country require cultural heritage institutions (libraries) to harvest news organization websites that are publicly available (not behind a subscription paywall)?

3. Do the laws of your country require cultural heritage institutions (libraries) and publishers to cooperate in order to preserve born digital news when this news is behind a subscription paywall?

##### **Practices**

1. Does your library receive born digital news from publishers by FTP or similar means? For this question by "receive" we mean that publishers initiate the transmission of born digital news to the legal deposit authority (library). In tech speak, the publisher "pushes" the news to the authority (library).

2. If publishers "push" news to your library, how does your library decide which publishers? What criteria are used to decide if born digital news from a particular publisher should be preserved?

3. Does your library harvest news websites? If your library does harvest news websites, how frequently does it harvest? Once a day? Once a week? Multiple times per day or week or month?

4. Depending on the publisher, news stories published on the web may be updated several times in an hour, day, or week. Do your library's harvest practices take any action if a news story is updated (new version)?

5. Depending on the frequency of your library's web harvest, the harvest of a news website may miss new versions of a story or may miss entire stories if the publisher updates its website with a higher frequency than it is harvested. If this is the case for your library's harvest schedule, please estimate the number of stories or versions of stories that your library's new harvest misses. ("I don't know" is an acceptable answer.)

6. If your library harvest news websites, how does your library decide which websites? In other words, what criteria are used to decide if born digital news from a particular publisher should be preserved? What criteria are used to determine harvest frequency?