

An international survey of born digital legal deposit policies and practices

Frederick Zarndt

Global Connexions, Coronado CA 92118 USA.

frederick@frederickzarndt.com

Dorothy Carner

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

carnerd@missouri.edu

Edward McCain

Donald W. Reynolds Journalism Institute, University of Missouri-Columbia, Columbia MO 65211, USA.

mccaine@rjionline.org

Abstract

That news publication has changed dramatically since the advent of the Internet and the Web is no news to anyone. There are many examples of established news organizations that have either stopped printing newspapers or shifted to publishing news on websites or through social media such as Facebook and Twitter. There are even more examples of new news organizations that have never printed news on paper and are digital only.

To the authors' knowledge, every country has one or more legal deposit organizations tasked with preserving news for future generations. Legal deposit laws in some countries have been amended to include news that may never be instantiated on paper (*born digital* news). However, legal deposit laws are by no means universally amended and, even when such amendments have been made, their embodiment in practice varies widely.

As a follow-on to the paper *Missing links: The digital news preservation discontinuity* (<http://www.ifla.org/node/8933>) presented in August 2014 at IFLA News Media section satellite conference at the ITU Library in Geneva, Switzerland, the authors have surveyed cultural heritage organizations (libraries) around the world about their respective national born digital legal deposit policies and practices. We share the survey results and consider the ramifications of inadequate born digital news preservation policies and practice to future generations.

Overview

Legal deposit is defined as “a statutory obligation which requires that any organization, commercial or public, and any individual producing any type of documentation in multiple copies, be obliged to deposit one or more copies with a recognized national institution.” [<http://www.ifla.org/publications/ifla-statement-on-legal-deposit>] Although IFLA's statement on legal deposit doesn't explicitly differentiate printed documents from digital documents, in fact most national legal deposit laws address only the former. It is only recently that legal deposit

laws in some countries have been amended to include digital¹ documents as well as printed, but implementation of digital legal deposit is in its infancy (at best). And according to UNESCO “the main purposes of legal deposit are to create a comprehensive collection of national publications and to compile an authoritative national bibliographic record, in order to ensure their preservation and provide easy access to them.”

[<http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-legal-deposit-legislation/>]. Current legal deposit practices for web archives are summarized by the International Internet Preservation Consortium at <http://netpreserve.org/legal-deposit>; note that these practices are not specific to news.

The *Columbia Missourian*, a daily newspaper published by world’s first journalism school, the Missouri School of Journalism (established September 1908), became the first newspaper to publish both a print and an online edition (October 27, 1992). Also claiming to be the first online newspaper (1993), *The Tech*, is a student-managed and written newspaper for the Massachusetts Institute of Technology (MIT). Regardless of which newspaper claims the honor, it is at least 2 years before Brewster Kahle started the *Internet Archive* (<http://www.archive.org>) in 1995. At that time the *Internet Archive* would have been the only service that could have collected and preserved news published by either newspaper². And obviously this was many, many years before legal deposit organizations began web harvests, either informally (not mandated by national policy) or formally (mandated by national policy).

Although the authors have not counted or estimated the number of online news publications begun since 1993, undoubtedly there are many. What percentage of the news published on the Internet has been collected and preserved? To the best of our knowledge no one has estimated the percentage, but it is likely very, very low.

Why are we interested in born digital news? Because, like news published in newspapers, magazines, radio, and television, we, like journalist Alan Barth, believe that “news is only the first rough draft of history³”. Without collecting and preserving the first rough draft of history, more and more of which is published online in digital form, future generations shall find it difficult to understand how the past shapes their present.

In this paper we are interested in text-based news created and published digitally (electronically). And although we are concerned primarily with text-based born digital news, such news may of course include digital photographs, digital videos, or other digital media. The born digital news of interest may exist in digital form only or may also exist in print. Here we are concerned only with born digital news from organizations or individuals whose business is to primarily to publish news; we are not concerned with casual posts to Facebook, Twitter, SnapChat, Google+, Wordpress, or the like.

¹ This paper uses “born digital”, “electronic”, and “digital” to refer to documents (books, news, newspapers, magazines) which exist in digital form and which may or may not exist as print on paper.

² The 1st *Internet Archive* capture of the *Columbia Missourian* website is from 4-Apr-2001. The 1st *Internet Archive* capture of *The Tech* website is from 2-Mar-2000.

³ Barth, Alan. Review of *The Autobiography of a Curmudgeon* by Harold L. Ickes. *The New Republic*, Volume 108, p. 677. 1943.

For the survey and this paper born digital news includes

- Stories published on a news organization's or journalist's website (these stories may or may not be published in print). these stories are primarily text based but may include photos and illustrations as well as audio or video.
- PDF files of printed newspapers.

One can argue that this definition is too narrow and ought to include blogs, email newsletters, RSS feeds, tweets, etc, however we want to consider only news published in not too unfamiliar forms. The reason is simply to narrow our discussion and prevent brain cramps which might result from trying to define news more broadly but not too broadly.

Survey

The authors wrote a survey consisting of questions about national born digital legal deposit policies and practices. The survey was sent to colleagues at approximately 20 national libraries around the world; these colleagues were personally known to 1 of the authors (we say this to show that the survey was not sent to randomly selected libraries). Replies to the survey were returned from May 2014 to March 2015 (yes, more than nine months — a certain amount of badgering, hectoring, and pestering was necessary in some cases). The following libraries responded to the survey

Australia National Library of Australia

Croatia Nacionalna i sveučilišna knjižnica u Zagrebu

Denmark Statsbiblioteket (Aarhus)

Estonia Eesti Rahvusraamatukogu

Finland Kansallis Kirjasto

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 has 2 parts. The first part (Policies) asks 3 questions about national born digital legal deposit laws or policies, the second part (Practices) asks 6 questions about implementation of these laws or policies (cf. Appendix 1). Here we review the questions, the reasons for asking the questions, and, by way of illustration of the question's intent, give some of the responses we received. All responses are summarized in charts, table, and conclusion below. Each of the responses below is unaltered except for spelling or minor grammar errors where such errors might affect understanding of the response.

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.

With this question we wanted to discover if national policies require legal deposit of born digital news and if cooperation of news publishers with legal deposit authorities is mandated by policy. We also wanted to discover if lawmakers in the surveyed countries had even considered born digital legal deposit.

Australia Not currently, however, we expect legislation to give the National Library of Australia legal deposit rights for electronic material to be introduced to our Federal Parliament some time this year. State and Territory governments also have legal deposit laws some of which include digital material (<https://www.nla.gov.au/legal-deposit/requirements-australia-wide>).

Finland As far as online materials are concerned, a publisher never has an obligation to deposit any works of any kind on its own initiative. Either the National Library collects works without any measures required from the publisher; or the Library has to make a specific request to the publisher to take necessary action. However, if the Library makes such a request, then, yes, the publisher shall be under an obligation to comply.

Latvia No, Legal Deposit Law doesn't require publishers to send digital data to NLL. It is our duty (as a National Library) to harvest required websites and publishers are required to provide free access if necessary.

the Netherlands No, we do not have a legal deposit.

New Zealand The *National Library of New Zealand (Te Puna Mātuaranga o Aotearoa) Act 2003* and the *National Library Requirement (Electronic Documents) Notice 2006 (Legislation)* authorise the National Librarian to *copy* (my emphasis) any Internet document which meets the Act's definition of a public document. The **Legislation** also requires publishers to assist the National Librarian to store and use an identical copy of an electronic document, if, at any time, the National Librarian makes a written request for assistance.

The **Legislation** also requires publishers of off-line documents (an electronic document that is not an Internet document) to give 1 or 2 copies (usually 2) of such documents to the National Librarian, within 20 working days after the document is first published.

So, for digital newspapers which are published on the Internet (whether or not there is any restriction on access to the document) the responsibility rests with us to harvest these, rather than publishers sending copies, as they are required to with hard copy. Publishers must provide the National Library of New Zealand with downloadable access to paywalled files if requested, or to deposit copies if they are unable to provide such access.

Singapore The laws as at 4th July 2014 does not require the mandatory legal deposit of born digital library materials. However, there are legislation amendment initiatives currently in the pipeline. The relevant legislation will be promulgated by 2016 but this may be subject to change. At present, the National Library Board, Singapore (NLB) writes to the various entities and individuals in the Republic of Singapore and gets digital legal deposit on a consent basis.

United States No. The legal deposit regime for the Library of Congress is connected with the U.S. copyright laws and regulations. Print materials published or distributed in the U.S. are required to deposit 2 copies with the Library of Congress. However, works that are published only electronically and that have no physical counterparts are exempted from the deposit requirements until the Copyright Office issues a demand for deposit. In 2012, the US Copyright Office issued an interim regulation modifying that exemption to require deposit, when requested by the Library of Congress, of online-only serials, including newspapers. At this time, only e-journals are being demanded (see <http://www.copyright.gov/circs/circ07d.pdf>).

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)?

Again with this question we wanted to discover if national policies require legal deposit of born digital news by web harvest but may not require publishers to cooperate, for example, by providing access to news protected by a paywall. And again we wanted to discover if lawmakers in the surveyed countries had even considered born digital legal deposit.

France Yes, it is mandatory for the national library to crawl news organization websites that are publicly available.

Luxembourg Yes. The BnL is obliged to harvest news websites as part of the larger web publications harvesting obligation.

Sweden No, there is no law requiring any institution in Sweden to harvest the web sites of news organization, although the National Library has been doing so since September 1996 and since 2002 is allowed to store the harvested material and make it available on non-online computers at the library (SFS 2002:287 - http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Forordning-2002287-om-behan_sfs-2002-287/).

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?

Publishers of printed news may be obligated to send one or more issues of the newspaper to the national legal deposit organization, or, more rarely, the legal deposit organization may acquire the issues on the open market. With this question we want to discover if the burden of collecting born digital news is entirely on the legal deposit organization or is on both the born digital news publisher and the legal deposit organization.

Denmark Yes, cf. the law: “§ 10. The person under the legal deposit obligation must upon demand inform the legal deposit institution about access codes and provide other information etc. necessary for gaining access to the material, produce copies of the material and make the material available to the general public. The person under a legal deposit obligation is entitled to demand that passwords etc. not be made available to any third party.”

Finland Yes, with qualifications....online news is not a special type of content but is treated as any other type of material published online. The [Cultural Materials Depositing and Preservation Act (1433/2007)] recognizes three cases: (1) The National Library harvests online materials independently without notifying the publisher or requiring action from it. (2) If web harvesting cannot happen without action from the publisher, the National Library may require from it such assistance as is needed to enable web harvesting. (3) Should web harvesting not be possible at all, the National Library may request a deposit of online materials.

In practice, case number 2 is very rare. We either harvest freely accessible materials; or request a deposit. – In strict interpretation, the Library should request depositing only when harvesting without or with the assistance of the publisher. In reality, other factors add in. For example, possibility of getting metadata as a part of a deposit may move the Library to ask for one in case of harvestable material, too.

Luxembourg Yes. Publishers are required to give the library the technical means to harvest publications behind a paywall, respectively oblige publishers to submit this data on physical carriers (i.e. hard disks).

Sweden Yes, the new legislation requires publishers to send or make material available to the National Library whether it is behind a paywall (i.e. requiring a login) or not, as long as it has been accessible to the general public with or without payment.

United States No. As noted for question 1, we are now using legal copyright demand to acquire e-journals from publishers; these may be behind publisher firewalls. To do this, publishers work cooperatively with the Library of Congress. It is anticipated that eventually we would acquire born digital news that we “demand” for deposit either through cooperation that allows harvesting through a paywall or by some deposit mechanism.

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).

The intent of this question is clear: If the national born digital legal deposit law requires publishers to “push” born digital news to the legal deposit authority. A “push” policy is analogous to laws in some countries which require a publisher to “deposit” one or more printed copies of a newspaper with one or more depositories.

Australia Not currently. However, plans are underway to trial this with a small newspaper publisher in the near future. At least one Australian state/ territory library, however, is receiving material in this way – Northern Territory Library – NewspapersNT (<http://www.territorystories.nt.gov.au/handle/10070/190886>) pdfs are received via ftp from the publisher.

Croatia The current legal deposit regulation (Library law from 1997) states that the National and University Library in Zagreb (NUL) has the right to collect online publications. The NUL's Croatian Web Archive (HAW) (<http://haw.nsk.hr/en>) harvests all born digital material.

Estonia Yes. The publisher of the newspaper is the one who sends the pre-print files to the FTP-server of NLE voluntarily. But as we don't have a law, the first step is usually done by the library. We have a long history of negotiations with publishers behind, to explain the value of digital archiving. Today we are in the situation where also the Estonian Newspaper Association (EALL) supports us. We archive the electronic pre-print files for preservation purposes and will provide publishers with their files back whenever the publishers need them again. Part of the agreement is that publishers use NLE's server as an intermediate station, where media monitoring companies, who have negotiated licences from EALL, can download the pre-print files for their commercial use from the NLE's ftp server, every day, early in the morning. Access to the publications, archived in NLE's digital archive DIGAR (digar.ee), is organised following the access restrictions and embargo time which are defined by the publisher.

Latvia Right now we don't harvest news websites. But we have a few cooperative partners as “Dienas Mediji” (one of Latvian biggest newspaper publishers) who sends periodicals (not born digital) and “Latvijas Vēstnesis” (articles are available only in digital format).

Norway The publishers use ftp, we have a script which receives all the new editions at special time every day. We are controlling the input to the library.

Sweden The new legislation actually (and absurdly) prescribes deliveries on a physical carrier (USB/harddrive etc). The reason is that this is considered to be the lowest common denominator and we will have some publishers (although perhaps not of news) that will want to use this. However, most

publisher will want to use one of the three online methods currently used, namely 1) web upload, 2) FTP or 3) RSS-feed (from which the library then fetches individual articles, images etc.).

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?

Not all countries require every printed newspaper to be legally deposited, for example, most “free” newspapers⁴ in the United States do not need to be deposited at the Library of Congress. Even countries such as Australia and Singapore, which do not yet have formal born digital legal deposit laws, collect and preserve some born digital news. The intent of this question is to discover the criteria used for choosing which born digital news to collect and preserve. The criteria may be mandated by the born digital legal deposit policy for countries with such policies, or they may be adopted by the depository in countries without formal born digital legal deposit policies but which do nevertheless collect and preserve born digital news.

Croatia All content published on the Internet is considered to be published, and everybody who publishes is a publisher. These can be publishers in the traditional sense of the word, as well as authors of personal web pages. Reputation and reliability of the publisher are important selection criteria. Our criteria for selection can be found at <http://haw.nsk.hr/en/selection-criteria>.

Estonia Speaking of pre-print-files, we preserve in our digital archive the objects, which have a record in our OPAC. If there isn't a record in the catalog for some item (electronic pre-print files arrive before there is a record done in OPAC), we send the notice to the acquisition department to decide.

Luxembourg [Bibliothèque nationale de Luxembourg] will try to get all publishers with [its] “pull” model.

Norway We have cooperation with different publishers. Most of the born digital news we receive are newspaper and e-books.

Singapore As the NLB act for digital content has not been approved, we are not collecting any born digital news. There are only two major newspaper publishers in Singapore, namely Singapore Press Holdings (SPH) and Mediacorp, and both have entered an agreement with NLB to deposit the PDFs of their print version under the NewspaperSG initiative. However, this is not done for their online versions.

Switzerland We have projects with commercial publishers beforehand and agree on content they have to send. There is a new project under way: platform for self publishers – publishers upload e-books together with metadata directly to the library. Selection criteria will be defined. Planned for 2015.

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?

Similar in intent to the previous question. With this question we want to discover if libraries are harvesting news websites and how often.

Australia Online newspapers Currently we only regularly harvest one major online newspaper, the Sydney Morning Herald (SMH). We have been archiving the online SMH since June 2009 and we harvest

⁴ “free” means that readers must not pay for it. Even “free” newspapers are not (cost-)free to produce and publish.

it daily. In fact it is done twice daily although normally only one of the copies is kept. We have an automated scheduled harvest that runs in the middle of the night and we do another manually initiated harvest around 10am. The reason for this is to ensure a scheduled capture (the automated schedule) which is particularly important over weekends and holidays; but we also aim to get the site at the mid-morning high interest news time, hence the 10am harvest. The harvest is scoped to capture the front page and all the content linked from that page. So it is not the entire SMH site but it is more than just the front page – essentially it captures the front page and the content you would expect to be able to continue to read from the front page but which would require a click through to get to. We do also harvest, separately, another part of the SMH, what used to be called the ‘National Times’ but is now just headed ‘Federal Politics’. This is basically the opinion section of the SMH.

We also do harvests of special sections such as those relating to elections; for example the SMH ‘Australia Decides’ section from the last Australian federal election.

Sometimes we have been able to obtain permissions from other newspapers for harvesting and archiving some limited content. For example we were able to harvest the Guardian Australia’s election coverage pages (but not the whole site), daily.

We do a number of other newspapers, mostly ethnic newspaper sites which are harvested at varying intervals but certainly not daily (although for a time we did do a Chinese language online newspaper daily).

News sites

We do not regularly harvest online news sites. However, like the online newspaper sites we do time limited collection around events such as elections, disasters, wars etc. Frequency for collection varies (e.g. daily, fortnightly)

News commentary

We tend to do more in the way of collecting news commentary. For example we harvest The Conversation weekly.

We also do annual whole .au domain harvests – and have done since 2005 – and this will capture all online news sites (if they are on the .au domain) albeit only once a year though that will of course pick up any content that is retained on those sites. However, there is no quality checking done of this harvesting so what we collect is only the freely available content that the harvester is technically able to collect. These collections are not publicly available.

Estonia Yes, but not on regular basis. Once a year, there is whole national domain harvesting which covers also news sites. Occasionally there are some event harvesting - elections, olympics and there is the online media coverage important target. In 2015 a daily harvest of news websites will be established. In some cases even twice a day and in some cases a weekly interval would be enough.

the Netherlands The web archive of the Koninklijke Bibliotheek does not include news sites. There is one exception, however. As a test, we do harvest nu.nl (free of charge news website) on a daily basis at noon, with the publisher’s explicit consent.

Norway [The Nasjonalbiblioteket] harvests about 1500 different url, every day. The front page is harvested every hour. Three levels down every 6th hour. About 200 websites are Norwegian newspaper on the Internet.

United States The Library of Congress has harvested some news sites for short periods of time as part of event-driven collections , weekly or monthly. However, for those collections, we have limited selection of news sites because of limited response rates and explicit permissions requirements. In the past two years, LC has begun weekly harvesting of some general news sites with particular political points of view as part of an ongoing “public policy” collection. The success of these harvesting efforts has led to expansion of our selection of news sites. We have recently initiated “General news on the Internet,” a

web archive collection that will include additional sites presenting U.S. general news, such as Huffington Post or BuzzFeed. These sites serve a national (rather than regional or local) audience and are not associated with presentation of the same (or similar) news via another platform or format (such as NYTimes.com or CNN.com). The collection does not include news portals but rather sites that significantly represent reporting original to that web site with journalism that has some voice (rather than machine-generated).

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)?

If one visits popular (high traffic) online news publishers such as *ProPublica*, *Huffington Post*, *BuzzFeed*, etc, one quickly notices that their home pages and news stories change frequently, sometimes several times in a day. The changes mostly do not follow a fixed schedule. With this question we want to discover if libraries adjust their born digital news collection practices to capture these changes.

Germany Not yet. Usually we have longer update times. We are looking for cooperations and ways to do it, I expect for the beginning exploration projects. So we are very open for cooperation.

Norway [The Nasjonalbiblioteket] just [has] automatically harvesting. [See Norway's answer for practice 3]

Sweden Yes, the library uses RSS feeds from the publishers and typically collects new material and new versions of older material once an hour.

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.)

Referring again to the fact the online news publishers frequently change their websites or publish new versions of online stories (see comments for question 4 above), the authors suspected that even if born digital legal deposit is mandated, the collection practices would take no notice of the change or new versions. In fact Sweden and Norway have adapted their collection practices to online news publishing practices. For other countries which are collecting online news, no one has estimated how much news might be missed because the frequency of collection does not match the frequency of online news updates.

Latvia Our future intentions in news website archiving involve answering questions about how deep need to be harvesting - all levels deep or only first or second. We hope that in future we will have a system which will find topical keywords so important websites can be harvested more deeply.

Norway I don't know, but I think the harvesting frequency we use will cover a lot of articles.

Sweden Not very many I should think and mainly updates, due to the frequent collecting, but no estimate has been made so far.

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?

Libraries have curation criteria for printed (hard copy) materials. With this question we want to discover if they also have curation criteria for their born digital news.

Australia Limitations on the selection of material to harvest are dictated by legal constraints (permissions), technical constraints and resource constraints. For these reasons a lot of our focus is on collecting news sites around events. Within the Sydney Morning Herald, the only major online newspaper we regularly harvest, sections may be selected according to special events – e.g. an election. We have been able to negotiate permission from other newspapers to harvest and archive limited content – e.g. the Guardian Australia’s election coverage pages (but not the whole site). We have also selected other newspapers, mostly ethnic newspaper sites which are harvested at varying intervals. We do time limited collection of online news sites around events such as elections, disasters, wars etc.

With the Sydney Morning Herald online we have experience in doing a daily harvest that we believe captures the essential content. This harvesting is supplemented by less frequent harvesting of other news content organised by type (e.g. political commentary) or events.

Criteria for determining harvest frequency for news websites are generally the same as for all our web archiving. Such criteria includes: the objective to capture all the substantive content but not capture every change to a website; as infrequent a harvest schedule that will adequately achieve this objective; a schedule that permits capturing the website in a size that can be managed given the existing technical infrastructure and available human resources. That latter could mean we do a site more frequently but to a shallower depth of harvesting so as to make the collecting processes more efficient, especially if there is a high turnover of content as is likely to be the case with news and commentary sites.

Croatia The decision about which websites, publisher or harvesting frequency is made according to the HAW's selection criteria at <http://haw.nsk.hr/en/selection-criteria>.

Finland Newspapers’ websites are harvested on the basis of a list obtained from the Finnish Newspapers Association; it contains all Finnish newspapers. (The Association defines newspaper as a *printed* newspaper that is subject to charge; issued at least once a week; publishing contents that are diverse and topical; and that may include an online version as well as other online news and advertisement services.)

The list of newspapers is complemented by an in-house list of circa 1,000 sites that is harvested monthly. The list focuses on news sources in public administration, academia, and NGOs.

Latvia We have developed guidance for web archiving. Since we don’t archive news websites, there aren’t specific criteria for born digital news. If necessary, common criteria for all harvesting websites will be complemented.

Norway In Norway we are waiting for a new Act of Legal Deposit. We hope that the new act will allow us to harvest everything in Norwegian. Now we have to contact every institution/publishers to get permission to harvest. First of all we contacted all the newspaper publishers, because we thought this was the most important websites.

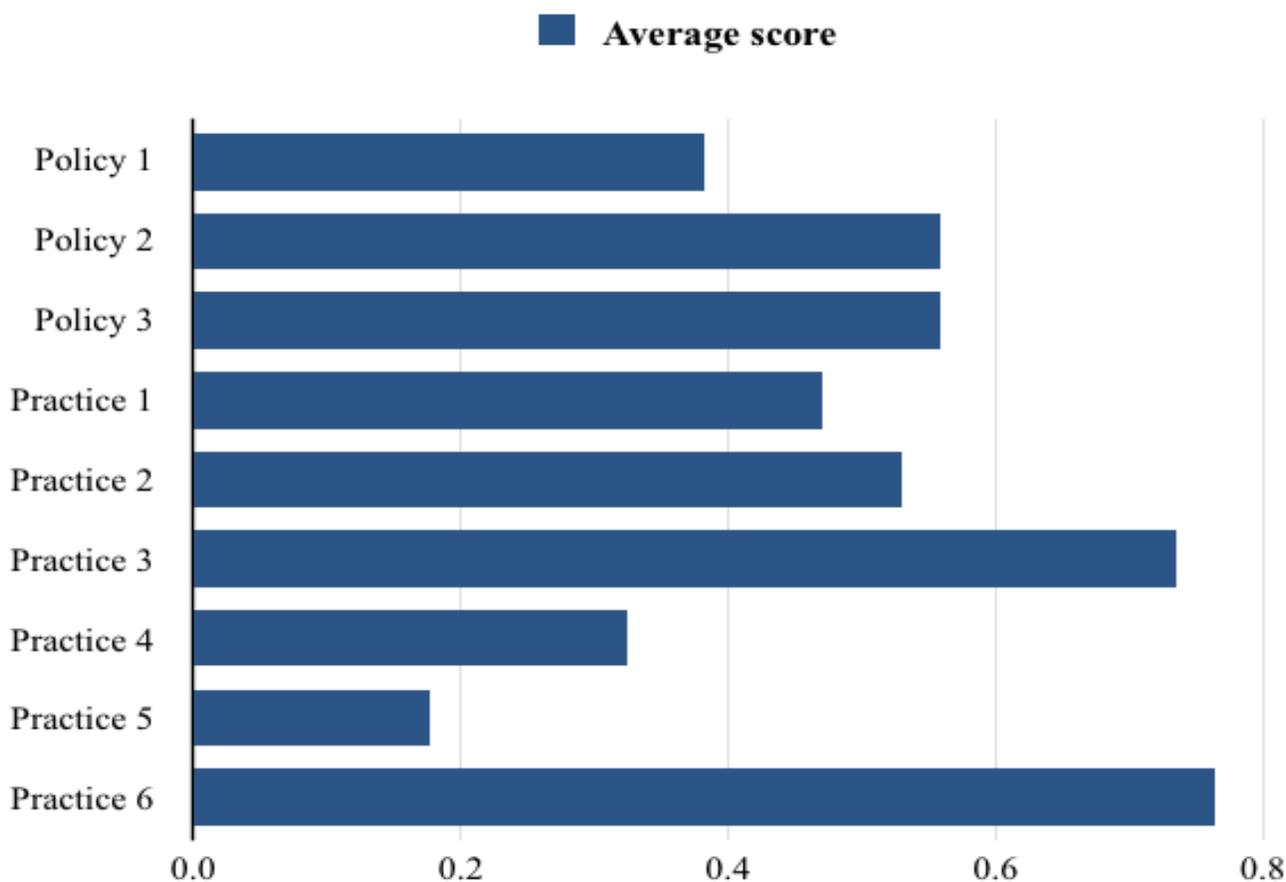
In Norway we have 19 different counties. It’s also important to cover websites from these counties and all the places connected to them. We have also focused on theme harvesting/special happening (ex. Terror, financial crisis, sports.....)

Singapore NLB does not harvest digital news. NLB selects online material based on themes or events of local interest. We are also required to seek permissions from web owners/publishers.

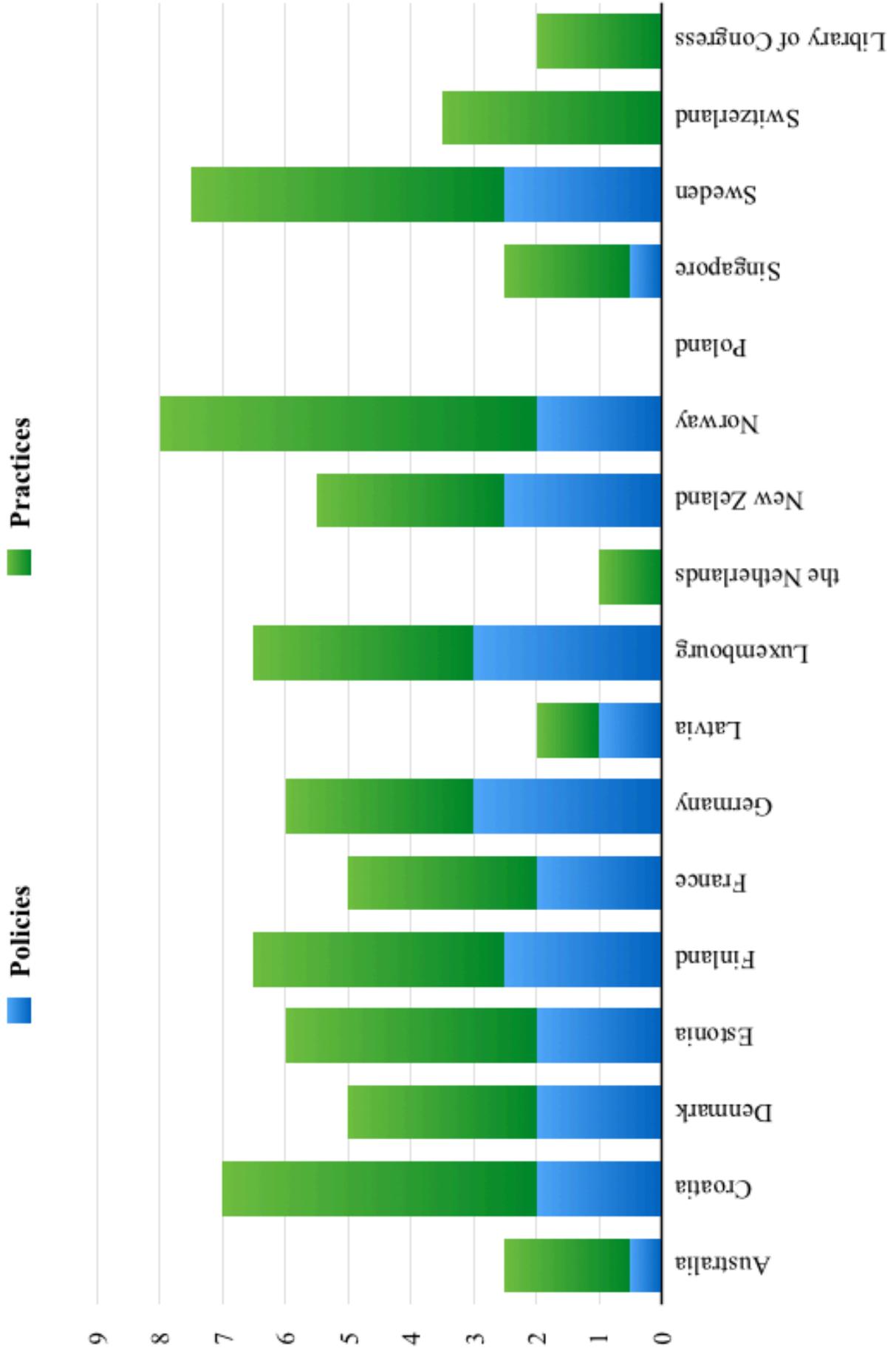
Simple visual representation of the survey results

Responses to individual survey questions were given a score depending on whether the response was negative — not favorable to the collection and preservation of born digital news — or positive — favorable to the collection of born digital news. The scores were 0 and 1 for negative and positive responses respectively. If in the judgment of the authors, the question was answered neither completely affirmatively nor completely negatively but something in between, it was arbitrarily given a score⁵ of 0.5. The scores do not reflect any moral or legal judgment about the answers: The scores are merely a means to a simple graphical representation of the survey results.

A word of explanation about the *Average score* and the *Policies and Practices* charts: For the *Average score* chart, if **every** surveyed country's national policies and practices were perfectly conducive to born digital news collection and preservation, the average score for each policy and each practice would be 1. Similarly if **each** country's national policies and practices were perfectly conducive to born digital news collection and preservation, then the score for **that** country would be 9 (1 point for each of the 3 policy questions and 1 point for each of the 6 practices questions).



⁵ Some answers to survey questions were ambiguous, that is, the answer did not or could not address the question specifically. If this was the case and in the arbitrary judgement of the authors, the question was scored 0.5.



Born Digital Legal Deposit Policies and Practices

	Policy 1	Policy 2	Policy 3	Practice 1	Practice 2	Practice 3	Practice 4	Practice 5	Practice 6	Policies	Practices	Total
Australia	0.0	0.5	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.5	2.0	2.5
Croatia	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	2.0	5.0	7.0
Denmark	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	2.0	3.0	5.0
Estonia	0.5	0.5	1.0	1.0	1.0	1.0	0.0	0.0	1.0	2.0	4.0	6.0
Finland	0.5	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	2.5	4.0	6.5
France	0.0	1.0	1.0	0.0	0.0	1.0	0.5	0.5	1.0	2.0	3.0	5.0
Germany	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	3.0	3.0	6.0
Latvia	0.0	0.5	0.5	0.5	0.0	0.0	0.0	0.0	0.5	1.0	1.0	2.0
Luxembourg	1.0	1.0	1.0	0.0	1.0	1.0	0.5	0.0	1.0	3.0	3.5	6.5
the Netherlands	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.5	0.0	0.0	1.0	1.0
New Zealand	1.0	0.5	1.0	0.5	1.0	0.5	0.5	0.0	0.5	2.5	3.0	5.5
Norway	0.5	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	6.0	8.0
Poland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Singapore	0.0	0.5	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.5	2.0	2.5
Sweden	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.0	1.0	2.5	5.0	7.5
Switzerland	0.0	0.0	0.0	1.0	1.0	0.5	0.0	0.0	1.0	0.0	3.5	3.5
Library of Congress	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	2.0	2.0
Total	6.5	9.5	9.5	8.0	9.0	12.5	5.5	3.0	13.0			
Average score	0.4	0.6	0.6	0.5	0.5	0.7	0.3	0.2	0.8	1.5	3.0	

Conclusion

Legal deposit laws vary widely from country to country. While some, like Sweden and Denmark have been capturing digital content for several years, many countries have no digital component in their legal deposit laws. Of the 16 countries responding to the surveys, seven have policies that allow them to capture digital content while nine do not.

Australia Currently, Australia’s national library only regularly harvests one major online newspaper: the Sydney Morning Herald. It is harvested twice daily and only one of the copies kept. Selection criteria are dictated by legal, technical and resource constraints. They focus on collecting news sites around events. Harvest frequency is based on a schedule that will adequately capture the entire substantive content but not every change to the website. Under consideration by Parliament in the Civil Law and Justice legislation Amendment Bill 2014, Australia is actively working on new digital deposit legislation.

Croatia In Croatia, the National and University Library in Zagreb (NUL) has the right to collect online publications. The NUL’s Web Archive (HAW) harvests all born digital material. Consequently, the NUL has the right to gather news because of the provisions in the law where it is stated that NUL collects online publications. News, is not treated differently than any other online publication, collecting is based on HAW guidelines which include reputation and reliability of the publisher. The NUL collects, describes and archives news websites once a day, once a week, once a month, and several times a year depending on their mode of issuance.

Denmark Denmark's cultural heritage institutions have been collecting news sites as part of their online harvesting program since 2005. Denmark's legal deposit law requires publishers to cooperate to preserve born digital news even if it is behind a paywall. Although there is no law "requiring" any institution in Sweden to harvest the websites of news organizations, the National Library has been doing so since September 1996. Proposals from an advisory board dictate the criteria. Denmark "pulls" harvested content rather than having it pushed to them. They use "broad crawls" of the Danish Internet to provide snapshots taken four times a year; they "selectively crawl" or continuously harvest 80-100 sites from once a month to six times a day to capture frequent updated sites; and complete "event crawls" two to three times per year.

Estonia Estonia's Legal Deposit Act covers web publications since June 1, 2006, but it leaves the National Library of Estonia (NLE) the right to decide what to collect. Through an agreement with the Estonian Newspaper Association archive preprint newspaper files (.pdf) in NLE's digital archive Digar, but there is no mandate to archive digital news at this time. Criteria for collection was developed by a working group of Estonian research and memory institutions in 2011, which suggests that the most used news sites are import to capture, in addition to some examples of citizen journalism and non-mainstream news sites. Original content is favored. Domain harvesting, which includes websites, currently is scheduled once a year. Daily harvesting is to be established in 2015.

Finland Under Finnish legal deposit law publishers never have an obligation to deposit any works of any kind *on its own initiative*. Either the National Library collects works without any measures required from the publisher; or the Library has to make a specific request to the publisher to take necessary action. However, *if* the Library makes such a request, then, yes, the publisher shall be under an obligation to comply.

France France's legal deposit law requires the national library to crawl news organization websites that are publicly available. They estimate about 100 news websites are harvested per day, with other press websites harvested weekly or annual. Harvest frequency depends on many factors: rate of publication, continuity of collections, audience, originality of content, etc. When the library requires publishers to provide content, they must comply, but the system is a "passive" one that requires only that publishes not prevent web harvesting. Criteria for capture are based on the dictates of the legal deposit and the press service of law economy and politics departments.

Germany Since 2006, Germany's laws have required cultural heritage institutions to harvest born digital content, if it is in the "public interest." Even though they concentrate on harvesting free material, paywalls offer no restriction for the need to acquire digital content. Harvesting news sites is only done for testing or under special conditions.

Latvia Under current legal deposit law the National Library of Latvia has the right but not the obligation to collect born digital news. Publishers are obliged to cooperate if the Library decides to collect online news. Presently collecting online born digital news published in Latvia is not a priority.

Luxembourg In Luxembourg, legal deposit laws mandate publishers to make born digital news available in a format that the National Library of Luxembourg (BnL) can access and copy. The BnL is obliged to harvest news websites as part of the larger web publications harvesting obligation. Publishers are required to give the library the technical means to harvest publications behind a paywall. Criteria for harvesting are based on the importance of the content.

Luxembourg has been working toward implementing electronic legal deposit since 2005. They are currently working to get software systems and people in place to fully implement it in 2015.

New Zealand If a digital newspaper is published on the Internet, whether or not there is any restriction on access to the document, the responsibility resides with the National Library of New Zealand to harvest them. The National Library is allowed to make copies of publically available websites, but not required to do so. Harvesting is done as part of dedicated projects, for a limited period of time. Most online content is aggregated on the www.stuff.co.nz site, so collecting this site provides the primary source of web news.

Norway Norway's Legal Deposit Act of 1990 was created before the Internet was a publishing platform. However, they have the cooperation of various publishers and must contact each institution or publisher to harvest. Of the 1500 urls captured every day, about 200 of them are Norwegian newspapers on the Internet. Front pages are harvested every hour and three levels down every 6th hour, with additional harvests for special events and themes.

Singapore Singapore laws do not require mandatory legal deposit of born digital news at this time. However, legislation is in the pipeline. At present, the National Library Board (NLB) of Singapore writes to various entities to request digital deposits on a consent basis. Although the NLB act for digital deposit has yet to be approved, they have agreements with the two major newspaper publishers: Singapore Press Holdings and Mediacorp to deposit pdfs of their print version. Currently, NLB is required to get permission to capture online material based on themes or events of local interest.

Sweden Sweden's new law on legal deposit of electronic material states that publishers of news must send or provide access to born digital news, to the National Library of Sweden, provided that (1) they also publish news in print or on TV/radio, (2) the born digital material differs from the print/broadcast material (is unique to the web), (3) the material is "of a permanent character", i.e. not e.g. a wiki that can be edited at will. The Library receives born digital content by USB, Web upload, FTP or RSS feeds. Since 2002, harvesting occurs once a day. From 1996-2001, harvesting occurred a few times per year. However, if an online news organization updates its content frequently, new versions are collected once an hour.

the Netherlands, Poland, Switzerland, United States the Netherlands, Poland, Switzerland and the United States, currently, have no digital legal deposit laws. In the United States, the Library of Congress has harvested some news sites for short periods of time as part of event-driven collections. In the past two years, LC has begun weekly harvesting of some general news sites with particular political points of view as part of an ongoing "public policy" collection. LC has also recently initiated "General news on the Internet," a web archive collection that will include additional sites presenting U.S. general news, such as Huffington Post or BuzzFeed. The harvest process captures a snapshot weekly and is based on popularity and large readership.

Based on the disparity of legal deposit laws for born digital content, it is evident that there are large gaps in the way countries preserve born digital news. It is probably fair to say that large countries with multiple governmental subdivisions may have a more difficult time enacting mandatory legal deposit legislation covering all digital news publishers. Commerce regulation and practices as well as the cost of human and technological resources needed to administer such programs in each country make it difficult to mandate legal deposit. However, if the world hopes to avoid experiencing another period similar to the "Dark Ages" when the cultural record of the twenty-first century is lost to bit rot and inaccessibility, cultural institutions, governments and public and private enterprises must find a way to preserve born digital news content. Legal

deposit laws can help provide a first step in preserving our digital heritage. Modeling and enhancing successful legal deposit systems in some of the countries mentioned in this paper might provide a viable path forward.

Appendix 1: Survey emailed to libraries

Born digital news includes for the purposes of this survey

- *Stories published on a news organization's website (these stories may or may not be published in print). these stories are primarily text based but may include photos, illustrations, or videos*
- *PDF files of printed newspapers*

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 publishers 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?*

Most of these questions are not simple "yes" or "no". The answers may take considerable time. For this we are very grateful. And we would not ask if we did not think that these are important issues.

References

1. Carner, Dorothy, McCain, Edward & Zarndt, Frederick. *Missing links: The digital news preservation discontinuity*. Paper delivered at the International Federation of Library Association Satellite Conference: Geneva, Switzerland, August 13, 2014. Accessed March 2015 @ <http://www.ifla.org/node/8933>
2. UNESCO. Guidelines for legal deposit legislation. Accessed March 2015 @ <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-legal-deposit-legislation/>
3. Larivière, Jules. *Guidelines for legal deposit legislation*. 2000. Accessed March 2015 @ <http://unesdoc.unesco.org/images/0012/001214/121413eo.pdf>
4. International Internet Preservation Consortium. *Legal deposit*. Accessed March 2015 @ <http://netpreserve.org/legal-deposit>
5. Barth, Alan. Review of *The Autobiography of a Curmudgeon* by Harold L. Ickes. *The New Republic*, Volume 108, p. 677. 1943.