



*Last updated 18 January 2023*

This guide is meant to be a first step toward exploring and understanding the kinds of digital resources available to information seekers through news publications. This is not intended to serve as a complete list of information sources available to researchers from journalistic materials, but rather as an overview, a place to help orient the information seeker and to provide some useful starting places.

### Collecting Digital News Formats

The links below are intended to provide a brief introduction to collecting digital news resources in various formats. They provide general information and examples regarding preservation, software, tools, realtime distribution, and aggregator services that can support building digital news collections.

#### Eprints/PDF

An “e-print” or “e-paper” is a born-digital PDF usually generated for the print edition of a newspaper. May be distributed online or through other targeted channels.

- o [Texas Newspaper PDF Preservation: A Low-Cost Solution with Tremendous Value](#) (2014)

#### Web Archives

Web archiving is a process for preserving the World Wide Web, usually in the form of browser pages, sometimes using an RSS or ATOM feed, in a manner that attempts to replicate the original user experience. The web archiving process typically employs web crawlers and automated processes to capture the large amounts of linked content available online.

- o [Tools & software - IIPC](#)
- o [Jones, Gina M. and Neubert, Michael \(2017\) "Using RSS to Improve Web Harvest Results for News Web Sites," \*Journal of Western Archives: Vol. 8: Iss. 2, Article 3.\*](#)

#### Article-level text

Article-level text for news content relies on one or more of the various text formats that can be exported from a Content Management System (CMS) as a complete story or article, including searchable metadata and sometimes also the images, sound, videos etc that were published together with the text. The articles can be distributed using a number of different methods such as RSS, FTP, OAI-PMH or manual uploading.

- o Example of a specification for RSS used for legal deposit:  
[Legal deposit of electronic documents by means of RSS feeds](#) (National Library of Sweden).

There are also tools which can be used to scrape news websites for article-level text and metadata (for example, to curate a dataset for research projects).

- o [Newspaper3k: Article scraping & curation](#)

## Subscription Resources

Sometimes referred to as news aggregators, these private enterprises generally follow a model where they collect and license content from news publications, then identify market segments and deliver content to new customers. Revenue from such agreements can be important to news organizations as they strive to maintain income streams. Aggregators also perform a variety of services for publishers to help them leverage the value of their content, even making content more discoverable to the news organizations themselves.

- o [NewsBank](#)
- o [Newspapers.com](#)
- o [ProQuest](#)

## Digitizing for Preservation and Access (Analog to Digital)

Also called digital reformatting, this process involves using scanners or other devices with digital sensors to capture analog print or electronic (such as magnetic tape) content. Newspapers are frequently digitized page by page and then OCR software is employed to produce searchable text of the digital images.

- Library of Congress: [Preservation Digital Reformatting Program](#)
- Educopia: [Guidelines for Digital Newspaper Preservation Readiness](#) (2014)
- Educopia: [Chronicles Interoperability Tools](#) (2014)
- [A Collaborative Approach to Newspaper Preservation](#) (2020)

## National Legal Deposit Rules for Born-Digital Content (examples)

Legal deposit laws require news organizations to submit copies or allow digital capture of published content for preservation in a repository, often in a national library. The legislation regulating legal deposit may or may not be part of copyright law.

- International Internet Preservation Consortium [Member Legal Deposit Descriptions](#)
- [Newspapers: Legal Deposit and the Digital Era](#) (2011)
- [Born Digital Legal Deposit Policies and Practices](#) (2017)
- [Swedish e-legal deposit law](#) (2021)

## International Networking Resource Communities for Library Digital News Collecting

These links represent international organizations providing forums for collaboration and discussion of issues relevant to digital news collecting - web archiving and preservation of born-digital information.

- [International Internet Preservation Consortium](#)
- [Digital Preservation Coalition](#)

## News Landscape Reports since 2020

These reports focus mainly on the current state of the news industry itself. The annual Reuters Institute report is now in its tenth year, with a more global perspective. The RJI/MU Libraries report provides an in-depth examination of working newsrooms and the multiple challenges and opportunities presented by born-digital news preservation.

- Reuters Institute Annual Report: [2021 Digital News Report](#)
- Reynolds Journalism Institute and MU Libraries: [The State of Digital News Preservation](#) (2021)

### International Online News Projects

International interest in preserving born-digital news can take a variety of forms. The projects below are representative of projects that cross national borders and take different approaches to preserving news media.

- International Internet Preservation Consortium [Online News Around the World](#) Snapshot Archive (2018)
- News websites are an important part of collaborative rapid response webarchive collections created in response to major global events such as the [COVID-19 pandemic](#) or the [war in Ukraine](#).
- [International Coalition on Newspapers](#) (ICON) database of newspaper digitization projects
- Donald W. Reynolds Journalism Institute [Journalism Digital News Archive](#)

### Openly Available National Digitized Newsprint Collections

There are many digital newsprint resources openly available online. Below is a representative sampling of digital newsprint collections available listed by geographic location. Many National Libraries also have openly available digitized newsprint titles that can be accessed through their online catalog. Finally, for reference Wikipedia maintains a [List of Online Newspapers Archives](#)

#### Africa

- [Electronic Newspapers by Country](#), maintained by Columbia University Libraries

#### Asia

- [NewspaperSG](#): National Library Board of Singapore

#### Australia and New Zealand

- Australia –
  - [Newspapers & Gazettes](#): Trove, National Library of Australia
- New Zealand -
  - [Papers Past](#), National Library of New Zealand

#### Europe – [Europeana Newspapers](#)

- Germany – [Deutsches Zeitungsportal/German Newspaper Portal](#) (Deutsche Digital Bibliothek (DDB))
- France – [RetroNews](#), the newspaper website of the Bibliothèque nationale de France (BnF)
- Netherlands – [Delpher](#) is the free, online repository of digitized printed material from the Netherlands (Dutch only). It was created and is maintained by the Koninklijke Bibliotheek (KB)

- Norway - <https://www.nb.no/search?mediatype=aviser> (Norwegian only)
- Sweden - <http://tidningar.kb.se/> (Swedish only). Material that is copyrighted can only be read in its entirety at the National Library and a few other Swedish libraries.
- Finland - [https://digi.kansalliskirjasto.fi/search?formats=NEWSPAPER&set\\_language=en](https://digi.kansalliskirjasto.fi/search?formats=NEWSPAPER&set_language=en)
- Denmark - <https://www2.statsbiblioteket.dk/mediestream/avis>
- United Kingdom
  - [The British Newspaper Archive](#)
  - [Resources Online: The National Library of Scotland](#)
  - [Newspaper Database: The National Library of Ireland](#)
  - Welsh Newspapers (National Library of Wales)

#### Central America

[Online Catalog](#): Center for Research Libraries Global Resources Network

#### North America

- Mexico -
  - [Hemeroteca Nacional Digital de México](#), Digital National Newspaper Library of Mexico (Spanish only)
  - [Historic Mexican & Mexican American Press collection](#), University of Arizona Libraries (Spanish only)
- United States –
  - [Chronicling America](#) (Library of Congress/National Digital Newspaper Program)

#### South America

[Online Catalog](#): Center for Research Libraries Global Resources Network

#### Computational Research In Digital News Collections

Online access to born digital and digitized news sources has not only made such collections much more accessible to researchers, but has also opened up opportunities for new forms of research using computational methods. Web archives and digitized newspapers are being explored by researchers in the digital humanities using techniques including emerging AI and machine learning technologies.

- [GDELT Project](#) (processes many streams of global news media to create a computable record that can be explored, analyzed and visualized)
- [Newspaper Navigator](#) (an LC Labs project exploring the visual and textual content within the Chronicling America collection using machine learning)
- [Oceanic Exchanges](#) (uses computational methods across multiple digitized newspaper corpora to examine patterns of information flow across national and linguistic boundaries)
- [British Library News Datasets](#)

#### News Media Access and Preservation Publications

For more information on published reports with a focus on digital news media, see the following:

- [IFLA News Media Section Publication Repository](#)