

## Digital preservation: Why is this important to me?

Jody L. DeRidder | July 12, 2016 | Reynolds Journalism Institute

Journalists are dependent upon access to back files for research and context, but those back files may no longer be there. Almost all news content created in the U.S. today is digital, but digital content is even more fragile than print and might be scattered over a variety of media and storage systems.

### [How long is digital content accessible and usable after creation?](#)

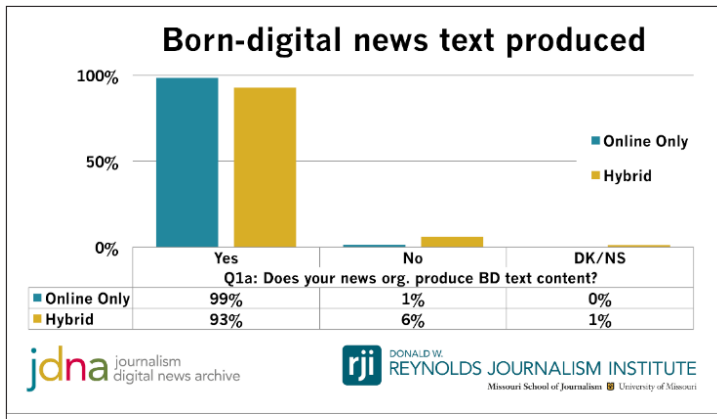
Device	25RH 10°C	30RH 15°C	40RH 20°C	50RH 25°C	50RH 28°C
<b>D3 magnetic tape</b>	50 years	25 years	15 years	3 years	1 year
<b>DLT magnetic tape cartridge</b>	75 years	40 years	15 years	3 years	1 year
<b>CD/DVD</b>	75 years	40 years	20 years	10 years	2 years
<b>CD-ROM</b>	30 years	15 years	3 years	9 months	3 months

The need for digital preservation is clear. Ninety-nine percent of online only news organizations and 93 percent of hybrid enterprises create born-digital text content. Survey results from 2014.

*All media types eventually fail. Two of the most significant factors are temperature and humidity. For example, at a relative humidity (RH) of 50 and a temperature of 28 degrees Centigrade, a CD-ROM can be expected to last 3 months. <http://www.dpconline.org/advice/preservationhandbook/media-and-formats/media>*

How long digital content is accessible and usable after creation varies widely depending upon what formats were used, what software and hardware are needed for access, and how the content was stored. Was it captured or stored using proprietary software? If so, access may be lost in a few short months, [as software and hardware continually are updated, and vendors go out of business](#). Was it stored on media that is virtually inaccessible now, such as floppy discs, zip discs, or old forms of hard drives? Is it on the web, but not linked to or maintained anymore? Even CDs and DVDs may begin to lose data [in as little as three months, depending upon the quality of the media, storage and handling](#).

News agency staffing and resources are stretched very thin. Many can no longer afford microfilm backups, and microfilm won't capture social media, videos, audio content and other forms of news.



In a [2014 phone survey by the Reynolds Journalism Institute](#), 93 percent of respondents from hybrid news publishers said they create content that is only in digital format; of responding online news publishers that percentage jumps to 99 percent.

This born-digital news content is in a dizzying array of formats, including multiple types of HTML, MS Word, XML, and others. Videos are generated by approximately 75 percent of the news agencies polled. When limited to image content only, 98 percent to 100 percent of the images are in born digital form only.

Not all content is even backed up, much less managed for long-term access. When asked whether all their born digital content from the past 25 years was backed up, only 57 percent of those from online-only news agencies could say yes; only 12 percent of those from hybrid news agencies could assent. Twenty percent of those from online news agencies said none of their content from the past 25 years was backed up. Seventy percent from this group said their agencies do not even have written policies for managing born digital content.

Yet over 70 percent of respondents said that these archives are very valuable for producing historic content and for quality journalism.

### Backup does not equal preservation

Even if news content is backed up, it's not safe. Media fails, [bit loss creeps in](#), and corrupted files may overwrite good backups. [Server crashes](#), [local or regional disasters](#), human error or intentional malfeasance can wipe out decades of content in an instant. [News agencies fold](#), and valuable historical content is lost forever. In a 2012 [Educopia](#) survey of more than 60 newspaper producers, fewer than half retain their website content or born-digital files for five years or more.



Content may be scattered across multiple systems and media. How can journalists find the information they need? How is the news agency ensuring users can open and use files no matter how old or where they've been stored?

News is important to the community. When newspapers fail, [citizens are less likely](#) to contact public officials, boycott products or services, or participate in civic or community organizations. Genealogists depend upon birth, death and marriage notices; companies researching areas depend upon market information, legal notices and real estate transactions; lawyers need access to past event coverage; historians and scholars need all of the above! Older news is not only useful to journalists to enable them to develop relevant, contextualized stories but also it's also extremely valuable as a historical record for society at large.

Until recent years, publishers sent copies of their old newspapers to libraries, archives and other cultural history institutions for storage and community use. From 1982 to 2011, the [United States Newspaper Program](#) worked with federal and state agencies to locate, catalog and preserve newsprint on microfilm. While largely successful, microfilm is not user-friendly; it is not easily searched and cannot be accessed online. In the past few years, the [National Digital Newspaper Program](#) has made great strides in digitally capturing old microfilmed newspapers by building a [centralized portal](#) through which content can be easily searched and accessed.



*The Chronicling America portal provides access to newspapers digitized from microfilm through the Library of Congress' National Digital Newspaper Program. However, this program does not preserve born-digital content.*

However, this portal only includes pre-1923 content due to copyright concerns. And it does not include any born-digital news content of any kind. How can we collect, collate and preserve news that has developed more recently? Without a process in place, our cultural history will have a tremendous black hole for the current and recently past decades. Not only will journalists be unable to research their stories; there will be no verifiable record of events for use by scholars, researchers and society as a whole.

### Digital content is fragile and short-lived

Digital content fails daily. Computer systems and software update constantly, which often leaves older files inaccessible and unusable. Storage media is also changing rapidly, yet who has the time and resources to monitor all content and migrate it to newer formats, systems and media? Over the past several years, major research institutions and cultural heritage organizations [have been building the infrastructure to manage digital content for long-term access](#). Standards have been developed for [Trusted Digital Repositories](#) to ensure success. But how is more recent news content to be included? How can publishers benefit from the efforts of cultural history institutions if the news content remains stored in back rooms on degrading media and in formats that are rapidly becoming obsolete?

Today's news is tomorrow's history. Now is the time to save new and recent news content, for long-term access, before it is beyond help. If we care about the historical record; if we care about our cultural heritage; if we want journalists to be able to research their stories to develop effective, contextualized news — then digital preservation matters.