



## Early electronic journals: A preservation survey

Nick Szydlowski\*, Rosa Salangsang, Daniella Kate Corpuz

San José State University, United States of America



### ARTICLE INFO

**Keywords:**

Early electronic journals  
Digital preservation  
Preservation survey  
History of technology  
Internet history  
Web archiving

### ABSTRACT

In 1994, the Association of Research Libraries (ARL) published a print directory containing information on every electronic journal that could be identified, anywhere in the world. Thirty years later, this study surveys the current availability and preservation status of those 443 journals. While a significant number of these journals are no longer available, the results indicate that independent preservation efforts by individuals and small groups were a major factor in preserving many of the remaining publications. This study uses the preservation survey method to assess the preservation environment for these journals, and also to build an online directory of the journals to assist researchers in locating these often-overlooked innovators of electronic scholarly communication.

### Introduction

Electronic journals dominate the publishing marketplace, consuming the lion's share of academic library collection budgets and providing the conditions for the exponential growth of global research output, modern metrics of research impact, and many of the other practices and rituals that shape the contemporary academic environment. And yet, the electronic journal, in any iteration, has only been with us for about forty years. In its current form as a web-based publication, the electronic journal turns 30 years old in 2024.

In 1991, three years before the World Wide Web became widely available, the Association of Research Libraries began publishing the *Directory of Electronic Journals, Newsletters, and Academic Discussion Lists*, a print reference work that provides a guide to the early history of online publishing. This study investigates the current status of the earliest electronic journals. In investigating those journals, it also provides a snapshot of a pivotal moment in the history of scholarly publishing - a moment of experimentation and uncertainty which has been largely forgotten as electronic journals have taken their place at the center of scholarly communication.

The research team performed a preservation survey on the 443 publications that ARL identified in the 1994 edition of the Directory, investigating the current availability of the journals and identifying the widely distributed efforts that led to the preservation of some of these journals. The survey found that 34 % of the journals listed in the directory are no longer publicly available in any format. 37 % are available in full, and the remaining 29 % are available in part or only

available through the Internet Archive's Wayback Machine.

The results of the survey reveal the important role that community digital archives have played in the preservation of these journals. 15 % of the journals were preserved either through independent preservation efforts (12.4 %) or small coordinated efforts (2.5 %). For context, university efforts of any kind, including university libraries, preserved 15.3 % of the journals. The critical role of these independent and coordinated efforts in preserving the earliest electronic journals is a key finding of this survey.

In addition to insights into the preservation process, the survey provides a snapshot of the state of electronic publishing 30 years ago, just at the moment when the public gained access to the World Wide Web (WWW), and when most electronic journals hadn't yet built websites. The geographic distribution of the journals reveal activity concentrated in a few countries and urban centers. The journals shared their content through a dizzying array of network protocols, and some were even available through mail, fax, and microfiche. In contrast to today's environment where many journals share publication platforms and workflows, each of these early journals operated a little different from its peers. This biodiversity enriched the publishing environment, even as it continues to present preservation challenges for libraries and access challenges for readers. This study aims to assess the state of these journals and to bring attention to their unique contribution to the history of publishing.

\* Corresponding author.

E-mail address: [nick.szydlowski@sjsu.edu](mailto:nick.szydlowski@sjsu.edu) (N. Szydlowski).

## Literature review

This study builds on research that documents the history of early electronic journals, most notably the *Directory* itself. The primary method of this study - the preservation survey - is rooted in the practice of cultural heritage preservation and conservation. The practice of surveying the preservation status and needs of a particular collection or set of resources complements approaches that focus on web archiving at scale, or on web archiving as a computational or engineering problem.

### Early electronic journals

While the *Directory* is the definitive reference work on the earliest electronic journals, a number of articles published in the ensuing decade provide valuable context as well. [Mogge \(1999\)](#) documents the creation of the directory and provides insight into the rapid evolution of ARL's process. [Roes \(1994\)](#) performs a statistical analysis of refereed scholarly electronic journals, identified through the use of the *Directory* along with two other contemporary indexes: the World Wide Web Virtual Library, which seems to have been hosted by the CERN laboratory, and the CICnet electronic journals project, created by the Committee on Institutional Cooperation, the academic organization that serves members of the Big Ten Athletic Conference. The *Directory* continued to be a recommended resource as electronic journals began to be integrated into library collections and services ([Roes, 1996](#)).

These early electronic journals were an active topic of discussion, particularly in academic libraries, as reflected in a 1995 special issue of *Library Trends* on the topic of "Networked Scholarly Publishing" ([Lancaster, 1995](#)). The issue demonstrates that issues of preservation ([Brichford & Maher, 1995](#)) and discovery ([Weibel, 1995](#)) were apparent to librarians at the very earliest stages of electronic publishing.

The impact that electronic journals would or should have on the broader infrastructure of scholarly communications was a particularly active area of conversation, and that conversation points to the significance of these early journals as cultural artifacts. This is the era in which the Open Access Movement first began to take shape, and the earliest electronic journals were frequently discussed as models for a more open future for scholarly publishing. While detailed description of those conversations is beyond the scope of this article, one valuable resource for understanding this period is another ARL publication from the time *Scholarly Journals at the Crossroads: A Subversive Publication for Electronic Publishing* ([Okerson & O'Donnell, 1995](#)), which captures a series of conversations, and provides a window into the intellectual currents that were present at the time. As one prominent advocate of open access electronic journals wrote, "It is time to stop making apocalyptic predictions about the coming of the electropublication era and to start providing concrete strategies for hastening the day." ([Harnad, 1995](#)). The journals listed in the 1994 *Directory* were both participating in and inspiring a reimagining of the publishing process.

### Condition assessment survey as professional practice and research method

This study takes as its methodological model the condition assessment survey. This method derives from preservation practice within libraries and archives, and it is most commonly performed by surveying the condition and preservation risks of individual items. Items may be sampled or represent the entire population of collection items, and results are typically interpreted collectively, often with the intention of assessing the scope of intervention that may be suggested for specific collections ([Durant, 2019](#)). Notable early surveys at research libraries include studies on the deterioration of books at Stanford University ([Buchanan & Coleman, 1979](#)) and Yale University ([Walker et al., 1985](#)). Even more ambitious surveys have attempted to assess the scope of preservation needs at the national level ([Eden et al., 1998](#)).

In addition to describing the condition of collections, preservation assessments can be used to assess risk to collections and to identify parts

of the collection that are at particularly high risk. [Segaetsho \(2014\)](#) demonstrates an approach that identifies a large number of potential risk factors, along with their causes. [Walker \(2003\)](#) discusses adapting the survey methods created in libraries to archive and museum settings, and discusses varying sources and models of risk for different institutional contexts.

Though not all surveys are conducted with publication in mind, the condition assessment survey continues to be a core professional method of preservation librarians, with recent published examples by [Rinio \(2016\)](#), [Spunaugle \(2022\)](#), and [Brown et al. \(2020\)](#). Particularly relevant to this study is [VandeCreek's \(2022\)](#) study of the preservation status of early digital humanities projects published to the web between 1996 and 2003. That survey focused on the institutional context in which projects were created, but did not find a strong correlation between the institutional resources available and the preservation of digital humanities projects.

Considered as a research method specifically for the study of digital preservation, the condition assessment survey allows direct access to actual outcomes of past practice, in the form of information about the current status of relevant items. It also has the potential to identify risks associated with specific areas of digital content. While the literature about digital preservation at scale is critical to the field - for exemplary recent examples see [Frank and Rothfritz \(2022\)](#) and [Ahmad and Rafiq \(2022\)](#) - focusing on the preservation challenges of a particular collection or set of items can help to direct some attention away from abstract models and towards the practical realities of aging digital content. Surveys allow us to assess the technical, social, and organizational conditions that may lead to lost or damaged digital content and to identify conditions and approaches that have led to positive preservation outcomes.

### Assessing electronic journal archiving and web archiving

Finally, this study exists at the intersection of two distinct areas of digital archiving and preservation work: on the one hand, the archiving of electronic journals, and on the other, the archiving of websites, or web archiving. While there is a broad literature on archiving practices and approaches for both areas, and especially for web archiving, there have been fewer studies that attempt to assess the status of items archived through these methods.

The condition assessment survey has been applied to electronic journals in order to assess the effectiveness of the large-scale journal preservation systems CLOCKSS, LOCKSS, and Portico. In 2010 and 2011, the libraries at Columbia and Cornell Universities collaborated on a report that showed that, at that time, only 26 % of Cornell's journal holdings were archived through either LOCKSS or Portico, with similar results at Columbia ([Kurth et al., 2011](#)), ([Wolven, 2015](#)). A similar study at the MIT Libraries reported similar results ([Willer & Szydłowski, 2012](#)). In this context, [Kenney and Wesley \(2016\)](#) call for collaboration and coordination between all stakeholders in order to improve outcomes for journal archiving.

Moving away from e-journals to web archiving more broadly, the literature has a distinct focus on automated modes of assessment. These include specific metrics or tools for testing the capabilities of web archiving tools ([Kelly et al., 2014](#)), [Brunelle et al., 2015](#)), or for incorporating automated assessment into web archiving workflows ([Kiesel et al., 2018](#)). [Aturban \(2020\)](#) proposes a framework for verifying the fixity of web archives. This research focusing on assessment is part of a broader agenda to develop and improve automated processes for web archiving. This focus on automated assessment reflects the clear advantages of automation when working with large but discrete collections of archived works. Survey methods, while less common in assessing web archiving, have their own advantages, which will be explored and demonstrated throughout this study. While automated assessment is very effective in assessing the accuracy of captured content, survey methods allow assessment of steps in the preservation

process that occur after the content is captured.

## Methodology

The goal of this study is to better understand both the characteristics of the earliest online journals, and their current status, including availability, completeness, current hosting environment, and preservation. As a starting point, the research team selected the 1994 *Directory of Electronic Journals, Newsletters, and Academic Discussion Lists* (King et al., 1994), compiled by the Association of Research Libraries. This directory was published regularly from 1991 to 1998, and it aspired to catalog all of the electronic journals available throughout the world. For each of the 443 journals listed in the 1994 *Directory*, team members manually compiled data from sources including a scanned copy of the directory obtained from Hathi Trust; a series of directed searches using Google, the Wayback Machine, and WorldCat; and more general internet research to find archived copies of the journals. As an example of the types of strategies employed in general internet searching, research team members would often locate dead links to archived copies of the journal, and then use those dead URLs to locate copies of the journal in the Wayback Machine.

This manually compiled data forms the initial dataset used in our analysis. The collection of this data was funded by a grant from San José State University's Office of Research, which funded temporary positions which allowed two of the authors to participate in the study as student employees.

While it would be possible to perform a similar study focusing on any edition of the *Directory*, the 1994 edition captures the history of online publishing at a particularly interesting point. CERN first made the software that underlies the World Wide Web available on April 30, 1993 ([A short history of the Web, 2024](#)). The 1994 directory reflects that timing, with the majority of journals still published by email, Gopher, or FTP. The journals that published via HTTP in 1994 were true early adopters, which makes capturing this particular snapshot uniquely interesting. On a more practical note, the relatively small size of the 1994 data set made it feasible for the research team to perform in-depth searches for individual journals while staying within the parameters of

our funding. The 1994 *Directory* was also the last edition of the directory to attempt to be comprehensive in scope, as the 1995 edition narrowed its focus to more scholarly publications (Mogge, 1999).

After the initial data collection, the research team performed additional normalization and clean-up as described in the data dictionary and documentation for the project (include link). One of the notable steps was to geocode the address data included in the *Directory* for most journals in order to produce a map of the journals. Initial geocoding was performed in Esri ArcGIS software. Addresses that did not return a result in ArcGIS were entered manually into Google Maps and the best match was selected. The method of geocoding for each item is noted in the data.

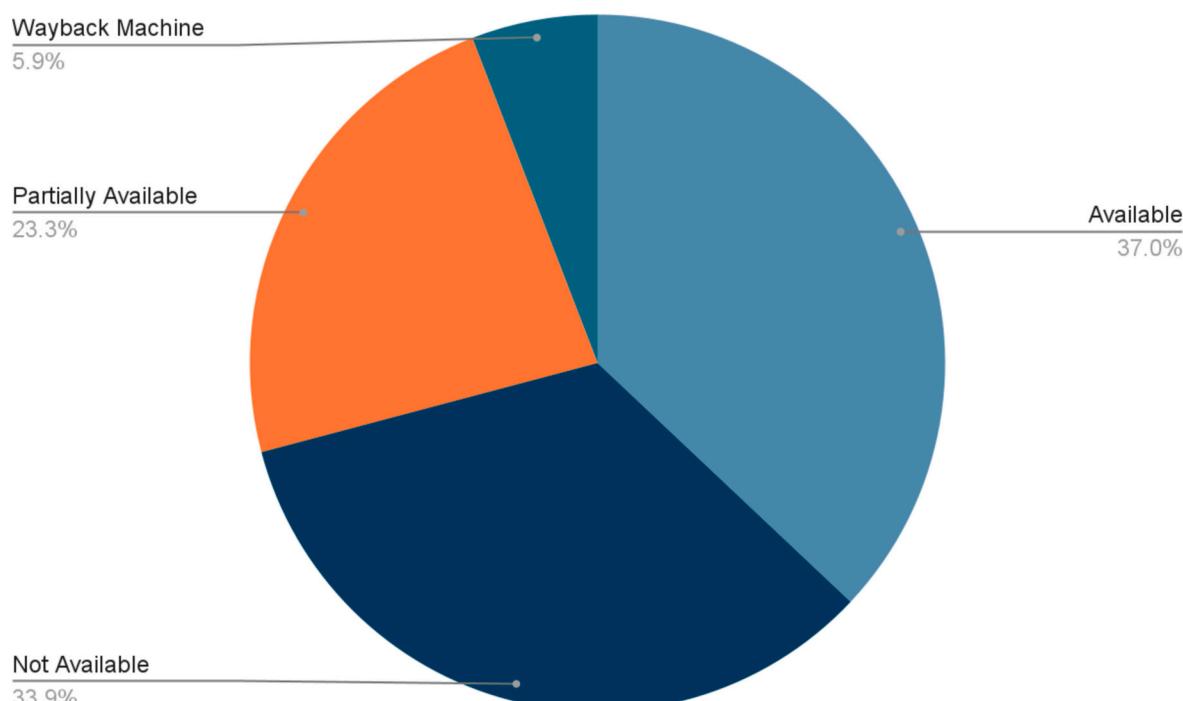
An additional round of coding was conducted to assign a mode of preservation to each journal, based on the earlier assessment of the journal's preservation status and the identified URLs and other sources where the content was available.

Data analysis and clean-up was performed in Google Sheets. Additionally, the project team is in the process of creating a website which will share the data for the study in an interactive format, using Omeka S. That website will also be a way for researchers to easily find the versions of these journals that do remain available, and it will be located at <http://exhibits.sjsu.edu/s/early-electronic-journals>.

## Analysis

### Availability

One of the most basic, but also most interesting questions about this set of journals is whether they continue to be available to researchers in 2024. As Fig. 1 depicts, out of the 443 journals included in the *Directory*, 150 (33.9 %) were not able to be located even after extensive internet and library catalog searching. 267 (60.3 %) of the journals were able to be located on the internet, either from the publisher or from a library or third-party archive. Of these, 164 (36.0 %) were available in full, and 103 (23.3 %) were only available in part, with some issues or other content missing from the archived version. Additionally, 26 journals (5.87 %) were available only from the Wayback Machine, but not from any current website.



**Fig. 1.** Availability of early electronic journals.

### World Wide Web adoption and other modes of distribution

Another important differentiating factor in this set of journals is the mode of distribution. This dataset captures the electronic publishing environment in the immediate aftermath of the introduction of the World Wide Web. At that moment, 101 (22.8 %) of the journals had a website. The remaining 342 (77.2 %) used other methods of electronic distribution, including email, the Gopher internet protocol, and file transfer protocol (FTP) (Fig. 2).

While 23 % of the journals shared URLs for their websites in the *Directory*, a smaller number (14.0 %) also listed the World Wide Web as a medium of distribution. This is understandable due to the variety of methods that these journals reported using to distribute their content, but for this reason, we use the presence of a URL, rather than the listed distribution methods, to identify journals distributed via WWW. Analyzing the distribution methods listed in the *Directory* shows that most journals (278, 62.8 %) used more than one method, while 176 (39.7 %) used three or more, and 82 (18.5 %) used four or more. The description of distribution methods was not highly standardized, with similar methods described differently by different journals.

By normalizing those descriptions into a smaller number of categories, we can see the technical environment of these journals more clearly. Even accounting for the seeming inconsistency in how WWW distribution was described, Email, FTP, Gopher, and Listserv all appear to have been more popular distribution methods than WWW at this juncture (Fig. 3).

The complexity of the distribution environment is one of the significant challenges to preserving these journals, and the normalized version of this data may actually conceal a good deal of additional nuance and complexity. As an example, the publication *Chaos Control* reported distribution via “Ftp [sic]; America Online; Full Hypercard [sic] version - on disk by mail.” While the text of early issues of *Chaos Control* is still available, the visual content included in the HyperCard version is not currently available. In 1994, different distribution methods allowed for very different features, especially for journals that included images or other visual content. Depending on which version of the journal has been archived, significant aspects of the original experience may or may not be recoverable.

### Geographic location of journals

The journals listed in the *Directory* are highly concentrated in North America (Fig. 4), and to a much smaller extent in Australia and England. While the aspiration of the directory was international, it is possible that North American publishers are significantly overrepresented due to their greater visibility to the editors of the *Directory*.

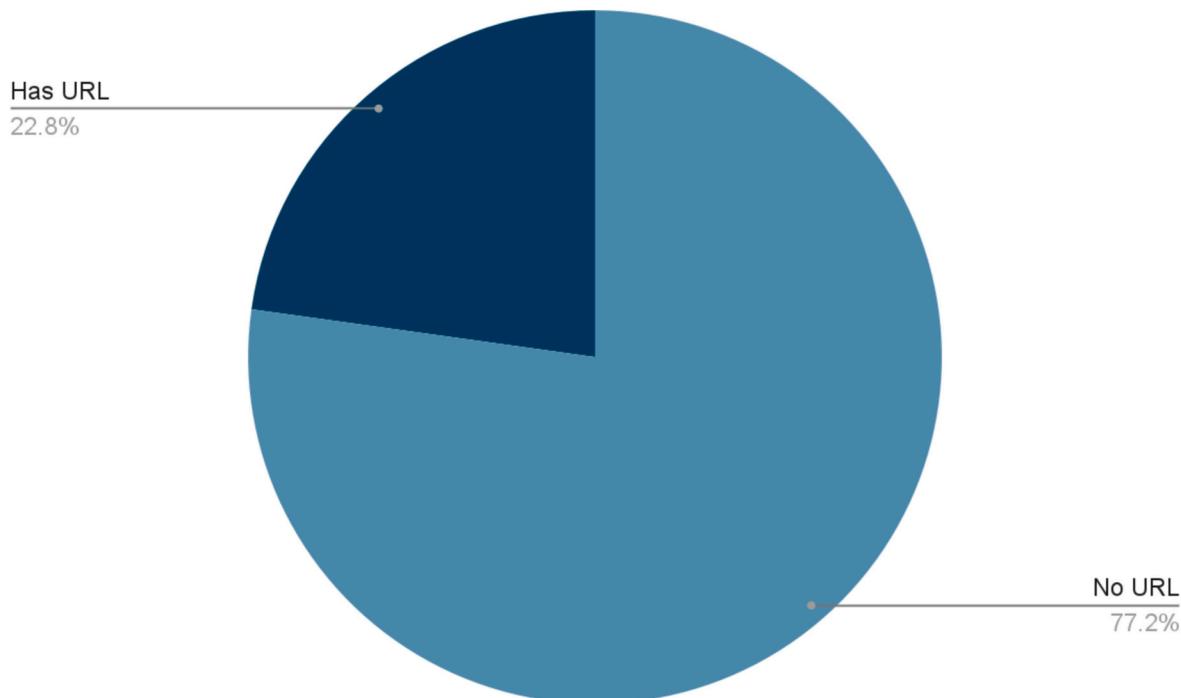
However, the outsized prevalence of United States journals may also reflect the roots of the internet in United States military technology, which resulted in greater availability of electronic publishing tools in the U.S. and at U.S. universities in particular. It is important to keep in mind that no central discovery tool for electronic journals, or other electronic content, was available at this time. The editors of the *Directory* were creating that tool, in print. The geographic distribution of the journals that the *Directory* indexed is a clue to the international distribution of electronic publishing, but we should be careful not to assume that the *Directory* was able to achieve its goal of comprehensiveness.

The map in Fig. 5 provides more detail on the location of the journals in the *Directory*, showing a strong concentration on the east and west coasts of the United States.

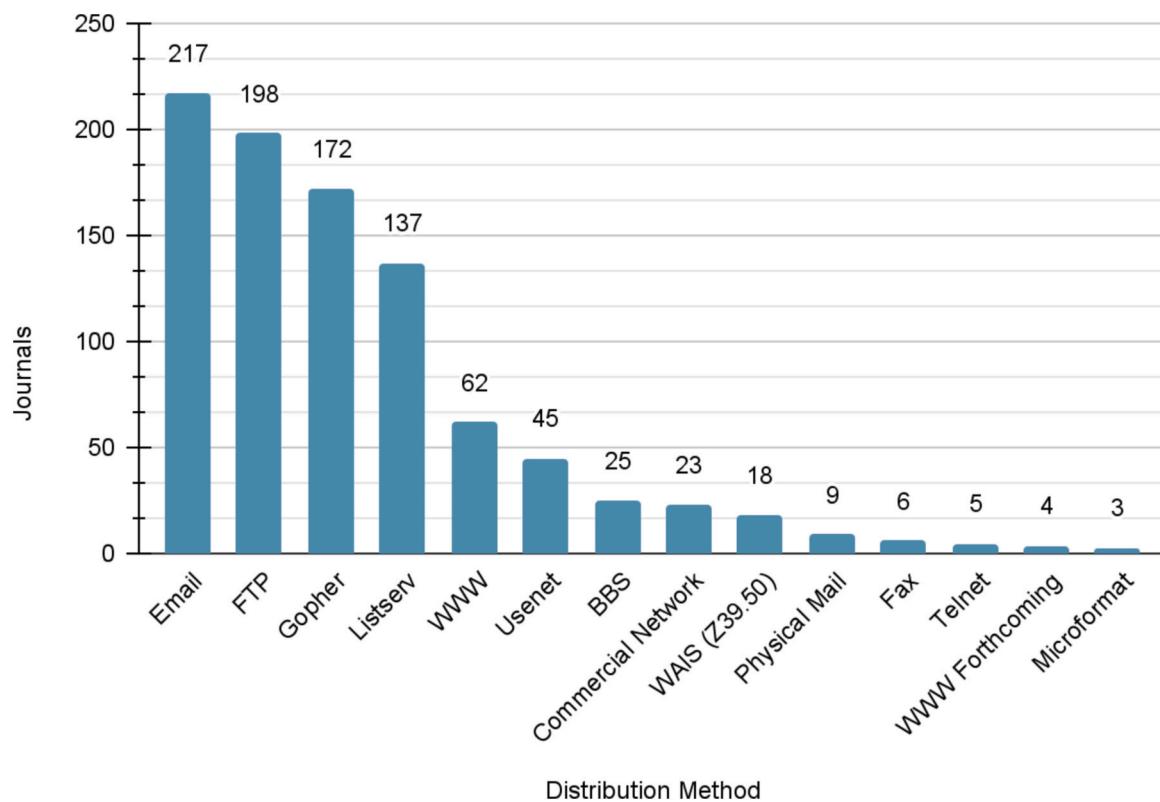
Considering geography very broadly, the data do not show a particularly strong relationship between the location of publication and the current availability of the journal for any of the continents with five or more journals (see Fig. 6).

WWW adoption however, as evidenced by the presence of a URL in the description of the journal, was at this time more common for North American and Australian journals than for journals published in Europe, or journals for which no address was available (Fig. 7).

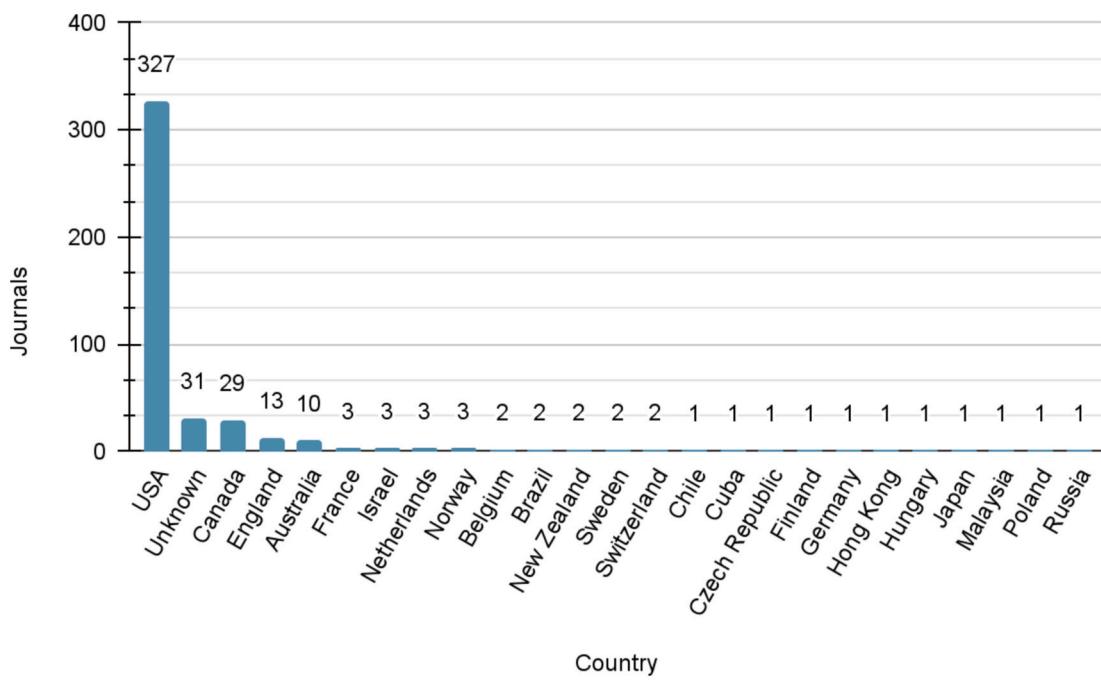
For journals that had adopted WWW distribution, the domain extension of the URL is an interesting indicator of their affiliation with a university, government, or commercial enterprise. Table 1 shows that 63 % of the journals with webpages were hosted on the .edu extension. The data about domain extensions also suggests that academic affiliation might be a predictor of current availability for these journals, as 51 % of the journals with .edu domains are currently available, compared to 33 % of journals on other top level domains. This is a helpful indicator, as many of the journals had ambiguous relationships to academic institutions that are difficult to determine with confidence from the



**Fig. 2.** Early electronic journals that adopted WWW distribution in 1994.



**Fig. 3.** Most frequent distribution methods of electronic journals in 1994.



**Fig. 4.** Early electronic journals by country.

available data. Put more simply, many of the journals seem to be published or distributed by people connected to universities, but at times it is not clear if the publication represents an academic or personal project.

#### Modes of preservation for early electronic journals

As the research team analyzed and coded the data, the mode of

hosting and preservation of each journal emerged as a possible subject of study. Fig. 8 displays the distribution of different modes of hosting, preservation, and access for currently available versions among the journals in the dataset.

To determine the mode of preservation, the research team first assessed any URLs at which the journal is currently available. For available journals, those URLs were distributed among a few frequent



Fig. 5. Early electronic journal locations.

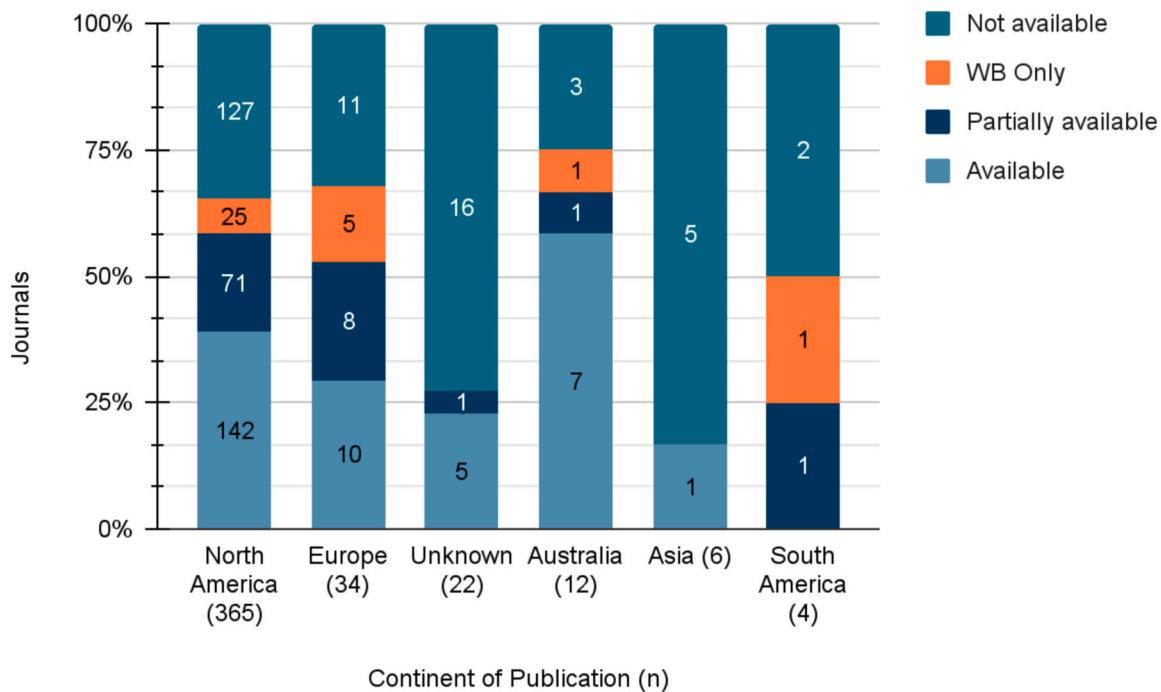
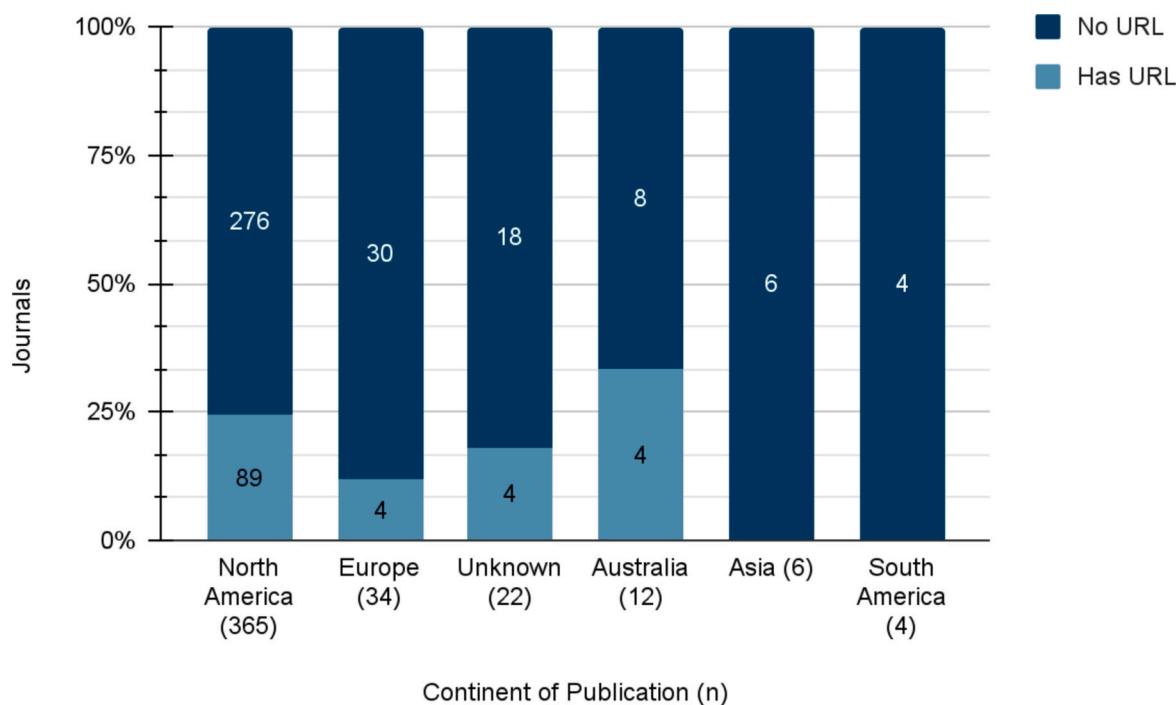


Fig. 6. Current availability of early electronic journals by publication location.

modes of access, with a large number of less frequent options that described fewer than five journals. The largest group (24.6 %) was described as hosted on the journal's own site. These journals are still primarily available from sites that appear to be affiliated with the same individuals, groups, or organizations that published the journal in 1994.

These are rarely published by the same mode of distribution or exact URL, so the presence of the journal on its own site speaks to some level of deliberate action taken in order to preserve and share the content. This category, like many of the others, also includes journals that are only available from the WayBack Machine. In this case, those journals were



**Fig. 7.** WWW adoption of early electronic journals by location.

**Table 1**  
Early electronic journals by URL domain.

Domain	Journals	
	Total	Percentage
.edu	61	62.89 %
.com	15	15.46 %
.org	10	10.31 %
.net	4	4.12 %
.ac	1	1.03 %
.gov	1	1.03 %
.int	1	1.03 %
.nl	1	1.03 %
.uib	1	1.03 %
.uk	1	1.03 %
.well	1	1.03 %
Total	97	100.00 %

archived from the journal's own site.

The second largest group of preserved journals (15.4 %) were preserved on websites affiliated with colleges or universities. The hosting method of these journals is actually quite diverse, with some hosted in the digital collections of university libraries and others hosted on the sites of academic departments, labs, or other units. Four journals (0.9 %) hosted on the personal sites of university faculty and/or staff were coded separately as "university personal webpage" but there may be some value to re-coding the other types of university-hosted pages with more granularity for further analysis, or for studies focused on the preservation environment in academic settings.

Table 2 demonstrates that a notable number of the journals were preserved either through the efforts of individuals and small groups ("Independent preservation effort", 12.4 %) or independent groups focused on preserving content on a particular subject or based on some other criteria ("Coordinated preservation effort," 2.5 %). The research team did not anticipate the importance of these grassroots efforts in preserving early electronic journals, and we will return to a more detailed analysis of their impact below. Additional notable modes of preservation included preservation by national libraries and governments, migration of the journal content to a Google Group, and

availability via print versions of the journal held by libraries or archives. Only six (1.35 %) of the journals were found to be available from commercial academic publishers, though some of the journals published independently (on their own sites) are currently available on a subscription basis.

The mode of preservation had a significant impact on the current availability of these journals. Fig. 9 shows the percentage of journals currently available for the three most prevalent modes of continued access: hosting on the journal's own site, hosting on a university-affiliated site, and access via an independent preservation effort. 75 % of university hosted-journals were fully available, versus 55 % of self-hosted journals and 58 % of journals captured by independent preservation efforts. Perhaps unsurprisingly, 36 % of journals archived through independent efforts were not available in full. In contrast, self-hosted journals were less likely to be incomplete (25 %) but more likely (20 %) to be only available from the Wayback Machine, and not from the journal's own server. This is particularly problematic due to the limited search functionality available on the Wayback Machine, and the fact that archival capture for these journals is not always complete. Generally the research team was only able to locate these journals on the Wayback Machine after finding a dead link or other reference to the journal's original URL. One reason we are excited to share our data online is to publish direct links to some of this archived content.

#### Independent and coordinated preservation efforts

The importance of independent and coordinated preservation efforts in preserving the earliest electronic journals was not something that the research team anticipated when beginning this study. In total, 66 (15 %) of the journals were preserved through these efforts, roughly equivalent to the number preserved on university sites and servers. Looking at these results in greater detail is helpful for illuminating the types of efforts that produced this surprising result.

One particular impactful site is [textfiles.com](http://textfiles.com/) (<http://textfiles.com/>) hosted by Jason Scott. 22 (5 %) of the journals in the dataset are available through the efforts of this site, which is supported by donations and announces its annual operating costs (\$1200) on its homepage. The research team wishes to highlight the value of this site to any researcher interested

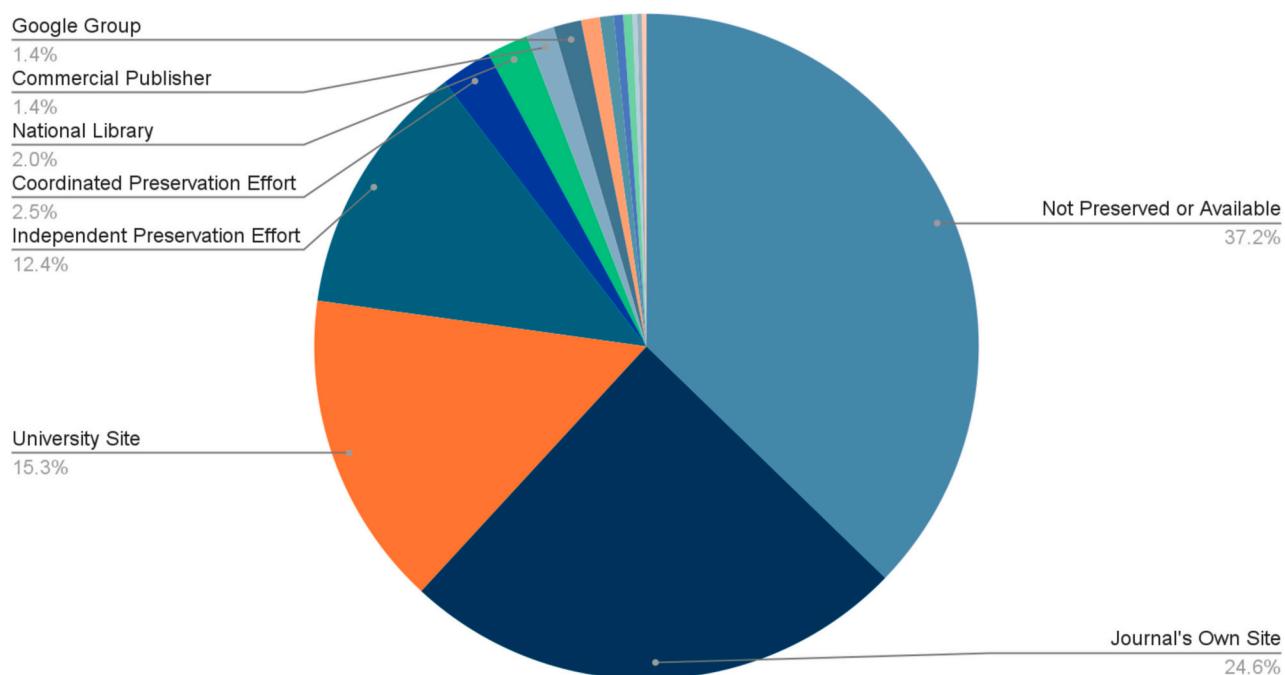


Fig. 8. Early electronic journals by preservation method.

**Table 2**  
Early electronic journals by mode of preservation.

Mode of preservation	Journals	
	Total	Percentage
Not Preserved or Available	165	37.25 %
Journal's Own Site	109	24.60 %
University Site	68	15.35 %
Independent Preservation Effort	55	12.42 %
Coordinated Preservation Effort	11	2.48 %
National Library	9	2.03 %
Commercial Publisher	6	1.35 %
Google Group	6	1.35 %
University Personal Webpage	4	0.90 %
National Government Site	3	0.68 %
Journal's Own Site on Community Hosting	2	0.45 %
Library Print Collection	2	0.45 %
Hosted on Commercial Platform	1	0.23 %
National Library Onsite Access	1	0.23 %
Physical Archival Collection	1	0.23 %
Total	443	100.00 %

in the early history of electronic communication and the internet. The site appears to represent a curated collection of resources primarily published from 1980 to 1995, and the early electronic journals represented there are a small portion of its collection.

Other notable efforts of individual collection and curation include Eric Lease Morgan's Infomotions, LLC (<http://www.infomotions.com/>), which held 6 (1.6 %) of the journals, and the currently unavailable Etext Archives ([https://web.archive.org/web/20230000000000\\*/http://www.etext.org](https://web.archive.org/web/20230000000000*/http://www.etext.org)) which was the only source for 4 (0.9 %) of the journals. Journals hosted on the Etext Archives are still available on the Wayback Machine, but would be quite difficult to locate without first identifying the original URL. In our data collection, these original URLs were typically identified by finding dead links to the source on other pages.

Of equal note 25 (5.6 %) of the journals were archived by sites or efforts that only archived a single one of these journals. These were often thematic archives or sites that selected a single one of these journals for their collections. Interesting examples include the Queer Resources Directory (<http://www.qrd.org/qrd/>), CTAN: The Comprehensive TeX

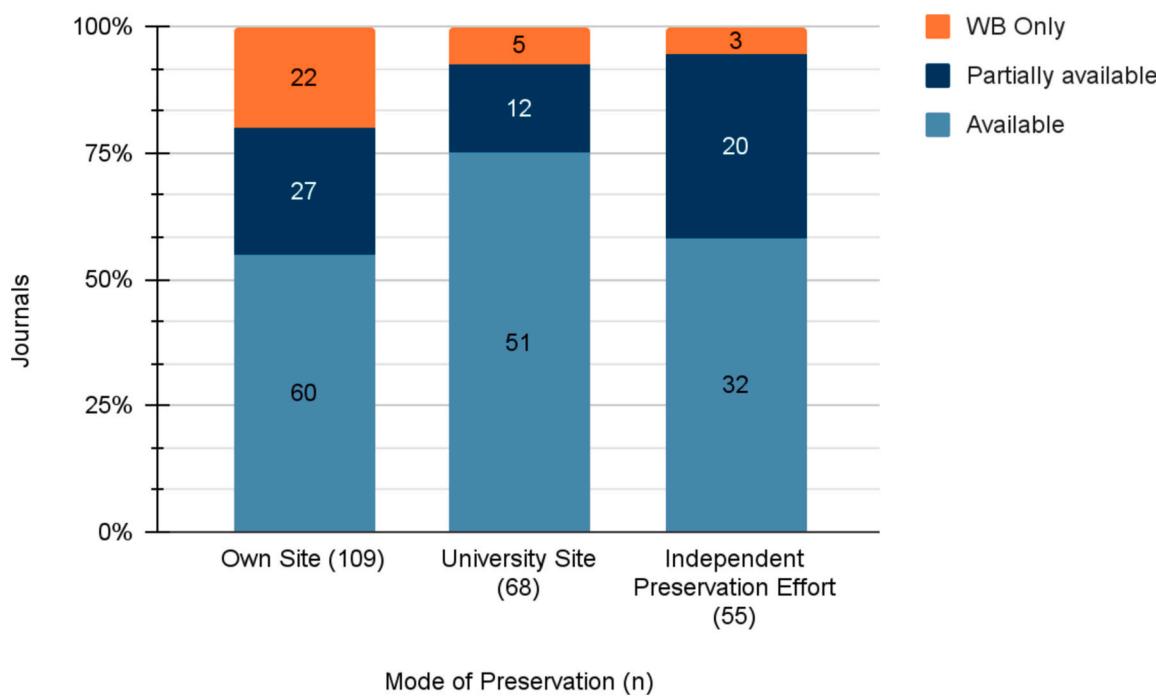
Archive Network (<https://ctan.org/>), and the Timor International Solidarity Archive (<https://timorarchive.ca/>). These efforts range from the highly coordinated, such as the Smithsonian-sponsored Biodiversity Heritage Library (<https://www.biodiversitylibrary.org/>) to the basically defunct: Spunk Library (<http://www.spunk.org/>), a collection of libertarian socialist resources, welcomes visitors to its homepage with an emphatic warning that “The actual text archive has probably not been updated since some time in the late 1990s.”

The importance of these independent and coordinated efforts should not be underestimated, but neither should their fragility. It is remarkable that these largely grassroots efforts can claim to have preserved nearly as many of the earliest electronic journals as the world’s universities. On the other hand, it is concerning that the preservation of these resources is so dependent on independent efforts - whether of small and self-publishers maintaining decades-old content on their own sites, or of citizen archivists building general or thematic collections. Upon closer inspection, even many of the university-hosted sites may raise concerns, especially those where the content appears to be resting unattended in an obscure subdirectory or subdomain. As is often observed, benign neglect is rarely an effective preservation strategy for electronic materials.

#### Discoverability of early electronic journals

Another area of interest for this study is the discoverability of these journals, in particular through library catalogs and discovery systems. In order to gather information on discoverability, the research team conducted searches in WorldCat to try to identify records for each journal, and collected the number of holding institutions recorded for each journal. 131 of the journals (29.5 %) could not be located in WorldCat by the research team. It is quite likely that there may be additional records for some of the journals that the team was unable to locate due to limited bibliographic information about the journals and journal titles that contain only common or generic words or phrases (e.g. *EJOURNAL*, *.dot.COM*, *Interface*, and *Inter/face*). Journals that are either absent from, or very difficult to locate in library catalogs are in some sense invisible, especially to many academic researchers.

However, records for the majority of the journals could be located. It



**Fig. 9.** Availability of early electronic journals by mode of preservation.

is not uncommon for libraries to catalog free electronic resources, and that practice was certainly less Sisyphean when the internet was a small fraction of its current size. As a result, the presence of records in WorldCat does not typically indicate that the holding libraries have preserved copies of the journal in question, but most often that they have chosen to highlight it as a free resource in their catalogs.

Still, it may be revealing to compare the availability of journals to their representation in WorldCat. **Table 3** shows a distinct, if unsurprising, correlation between the number of libraries reporting holdings for a journal and its current availability.

It is difficult to speculate about any causal relationship between the existence of catalog records for these journals and their preservation status. On the one hand, it could be argued that the journals with WorldCat records benefited from the attention of libraries. However, the inverse seems more likely - libraries may have tended to create records for electronic journals that thrived, and thriving, active journals were more likely to be preserved. Future studies that look more closely at the individual cases in question may be one way to resolve the relationship between WorldCat records and journal preservation status.

Revisiting the modes of preservation discussed above, it does appear that the mode of preservation or access has a relationship to the visibility of the journal via library catalogs. In particular, journals preserved via independent preservation efforts were only slightly more visible than journals that are not currently available at all, and were much less visible than any other category of preserved journals. As **Table 4** demonstrates, 56 % of independently preserved journals had a WorldCat record, with an average of 20 holdings per record. In contrast, among

**Table 3**  
Early electronic journals by number of WorldCat holdings.

WorldCat holdings	Number of journals	% of journals available
No record located	131	16.03 %
Record with 0 holdings	78	21.79 %
1 holding	42	35.71 %
3rd quintile (2–7 holdings)	64	43.75 %
4th quintile (8–159 holdings)	65	50.77 %
5th quintile (>159 holdings)	63	80.95 %
Total	443	37.25 %

**Table 4**  
WorldCat holdings of early electronic journals by mode of preservation.

Mode of preservation	Journals			WorldCat holdings	
	Lacks WorldCat record	Has WorldCat record	% with record	Total holdings in WC	Average holdings
Commercial publisher	2	6	75.00 %	2256	376.00
Coordinated preservation effort	1	10	90.91 %	433	43.30
Independent preservation effort	24	31	56.36 %	716	23.10
Not preserved or available	77	88	53.33 %	1779	20.22
Own site	19	90	82.57 %	25,006	277.84
University site	4	64	94.12 %	12,960	202.50
All others	4	23	85.19 %	5788	251.65
Total	131	312	70.43 %	48,938	156.85

the set of all journals, including those not currently available, 70 % had WorldCat records, with an average of 157 holdings per record. All other categories of preserved journals substantially exceed the values of the set of all journals, with the exception of journals preserved through coordinated preservation efforts, of which a very high percentage (91 %) had WorldCat records, but with only 43 holdings per record on average.

The low visibility of independently-preserved journals suggests that those journals, even more than journals hosted on their own sites, may be at significant risk, and might benefit from more targeted intervention by cultural heritage institutions.

#### *Impact of distribution method on availability*

While the World Wide Web would soon become the dominant method of distribution, electronic journals published in 1994 used a

dizzying array of distribution methods. Additionally, each journal was often available through multiple channels. This diversity creates significant preservation challenges.

Focusing on the five most popular distribution methods (Email, FTP, Gopher, Listserv, and WWW - see Fig. 3 above), Fig. 10 shows the current availability of journals based on their listed distribution method. The presence of a URL in the *Directory* record was used to indicate the availability of WWW distribution. For the other methods, the list of distribution methods in the record was used.

Journals using Gopher, FTP, and WWW are all available at rates slightly above the average for all journals. Journals using WWW were more likely to only be available via the WayBack Machine. Because the WayBack Machine only archives WWW content, this should not be surprising.

More than the specific distribution methods used, the availability of the journal through multiple popular methods correlates with a greater likelihood of current availability (Fig. 11). It is noteworthy that journals with WWW distribution were not necessarily more likely to be preserved than those with Gopher or FTP distribution. The one group with substantially better preservation records was those with four of the five primary distribution methods. This may reflect the distributed nature of preservation efforts directed at this set of journals, with different organizations focused on preserving journals distributed as text, or via the WWW.

#### Limitations

This study examines journals published at a particular juncture in the early history of electronic publishing. The 1994 edition of the *Directory* captures the online publishing environment in a moment of rapid change, just after the widespread availability of the World Wide Web, but before many journals had had a chance to take advantage of the new medium. While this is a particularly interesting moment, it does suggest that some of the statistical analysis performed for this study would yield different results in a study based on the 1993 or 1995 editions of the directory, and so care should be taken before extrapolating some of the data to other time periods or contexts.

In addition, despite the *Directory*'s intention of being an international

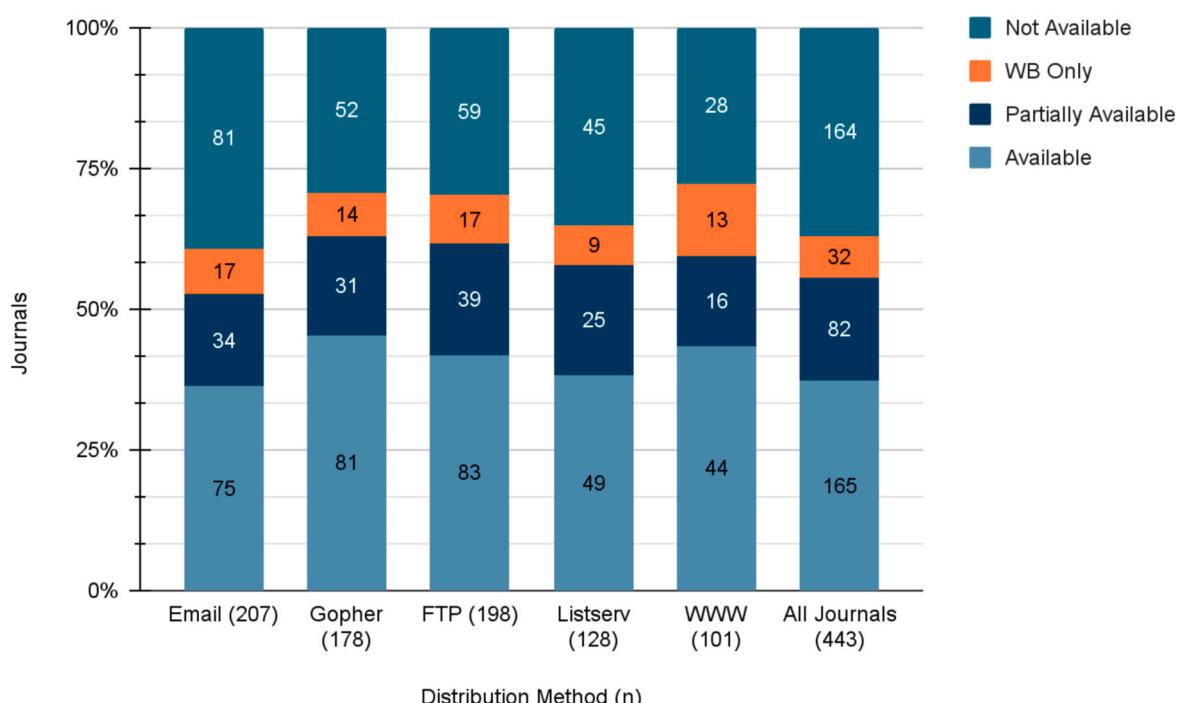
index, the journals included in the directory skewed heavily towards the United States, and especially United States universities. Given the limited distribution of computer networks at the time, some concentration of journals in the United States seems likely, but it is beyond the scope of this study to accurately assess the comprehensiveness of the *Directory*, and so we must be cautious and assume that non-U.S. journals are at least somewhat underrepresented in the dataset.

#### Conclusions

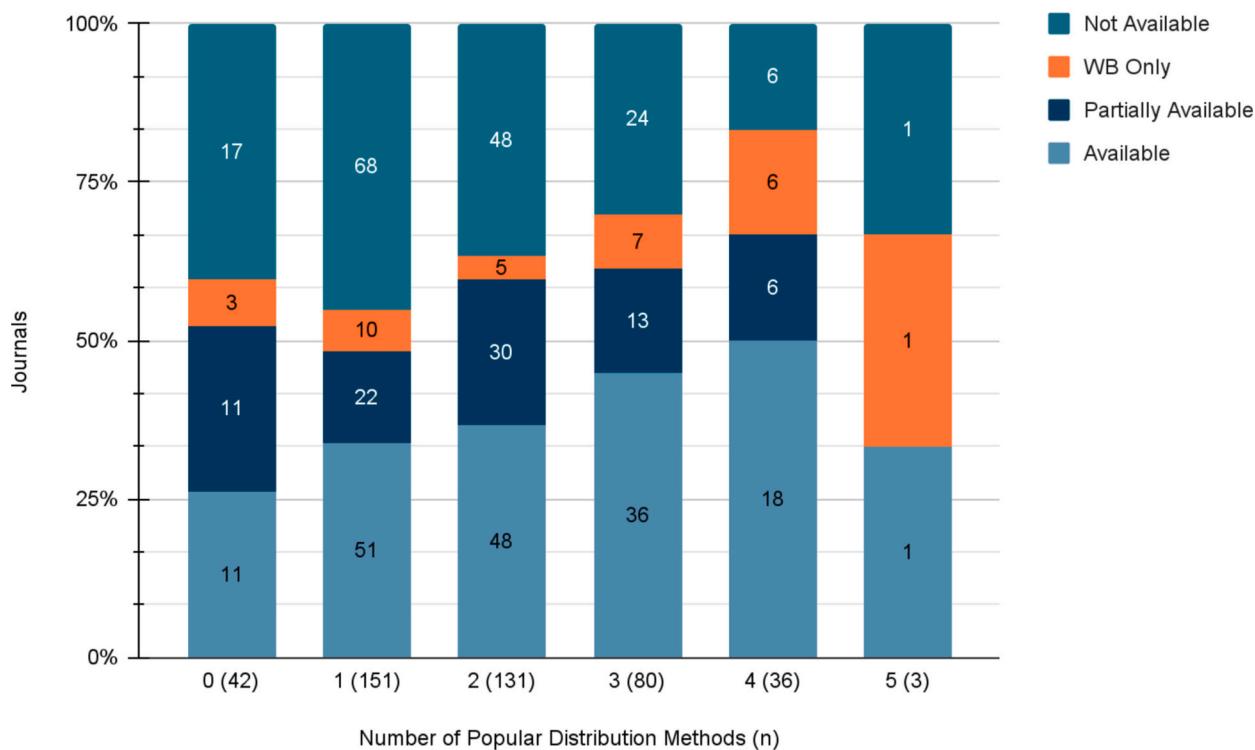
While many institutions and individuals have contributed to preserving the earliest electronic journals, 165 (37 %) of the publications are not currently available. While not all of these publications were long-lived or of great individual significance, it is striking that so many of the publications which merited inclusion in ARL's print directory of electronic journals are no longer available thirty years later. We can hope that some of these publications still exist offline, perhaps in the personal collections of those involved, and may even be transferred to institutional collections in the coming years.

Despite these gaps, the record of preservation efforts related to these journals is remarkable, and demonstrates the collective impact that is possible through highly distributed approaches to collection and preservation. For this group of journals, independent preservation efforts such as textfiles.com, Infomotions, LLC, and Etext Archives had a major impact on the current availability of the content, preserving 55 of the 443 journals in the dataset. These sites are rarely discussed in the literature around journal preservation, so documenting their impact is an important outcome of this study. Additionally, these preservation efforts are fragile, and the library community should make efforts to duplicate the content they preserve in a more sustainable manner.

The largest number of preserved journals exist on the publisher's own website. While this may be a stable arrangement for larger publishers, many of the self-preserved journals are part of websites that appear to be at risk. The same could be said of journals preserved on university servers, especially those hosted on personal websites of faculty and staff members, including on the tilde (~) servers long maintained by many universities. While the details of each journal's hosting arrangement are beyond the scope of this study, it is hard to avoid the



**Fig. 10.** Availability of early electronic journals by distribution method.



**Fig. 11.** Availability of early electronic journals by number of popular distribution methods.

conclusion that many of the journals rely on a small number of people for their continued availability, either through independent preservation efforts, or through websites archived and maintained by individuals involved in the journal.

This distributed mode of preservation presents significant challenges for discovery. 209 of the journals (47 %) had either no record in WorldCat (131), or a record with no holdings (78). The data on WorldCat holdings lend support to the conclusion that, despite the work individual libraries have done to preserve journals related to their home institutions, there has not been a coordinated effort on the part of libraries to preserve these early electronic journals.

While this study may highlight the challenges of accessing and preserving these journals, it also presents an interesting picture of the state of electronic journal publishing in 1994, at the moment when the WWW became widely available. The research team plans to explore this aspect of the project more through a digital exhibit, but it is fair to say that the data portrays a publishing environment in a state of transformation. The journals in this study seemed to use every method imaginable to share their content, including physical mail, fax, and microfilm. At the same time, these journals were on the cutting edge of new technologies that the academic publishing industry at large had not yet adopted.

Finally, this study points to the continued relevance of the preservation survey as a method for identifying gaps and needs in the preservation environment, and for understanding the mechanisms of digital preservation. Digital preservation is often discussed as a systematic process that must be done at large scale in order to be effective. While large-scale efforts like the Wayback Machine are an important part of the story for these journals, this study also reveals the small, labor-intensive efforts that have been needed to move these journals from the internet of 1994 to the World Wide Web of 2024 and beyond. Having documented their impact, the field of digital preservation would benefit from future research on these community-created and independent digital archives to better understand the motivations and practices of their creators, and to identify challenges and obstacles that they have encountered.

#### CRediT authorship contribution statement

**Nick Szydlowski:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rosa Salangsang:** Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – review & editing. **Daniella Kate Corpuz:** Data curation, Formal analysis, Investigation, Resources, Validation, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data is available at <https://exhibits.sjsu.edu/s/early-electronic-journals>.

#### References

- A short history of the Web. (2024, April 29). CERN. <https://home.cern/science/computing/birth-web/short-history-web>.
- Ahmad, R., & Rafiq, M. (2022). Assessing the preparedness of university libraries for digital preservation. *The Journal of Academic Librarianship*, 48(6), Article 102617. <https://doi.org/10.1016/j.acalib.2022.02617>
- Aturbar, M. (2020). *A framework for verifying the fixity of archived web resources*. Doctoral dissertation, Old Dominion University <https://www.proquest.com/docview/2451138951/abstract/A7987D3D1AED4400PQ/1>.
- Brichford, M., & Maher, W. J. (1995). Archival issues in network electronic publications. *Library Trends*, 43(4), 701–712. <http://hdl.handle.net/2142/7990>.
- Brown, N., Coppola, F., Modelli, A., Amicucci, F., Lichtblau, D., & Strlič, M. (2020). Non-destructive collection survey of the historical Classense Library. Part I: Paper characterisation. *Heritage Science*, 8(1), 88. <https://doi.org/10.1186/s40494-020-00432-w>
- Brunelle, J. F., Kelly, M., SalahEldeen, H., Weigle, M. C., & Nelson, M. L. (2015). Not all mementos are created equal: Measuring the impact of missing resources.

- International Journal on Digital Libraries*, 16(3), 283–301. <https://doi.org/10.1007/s00799-015-0150-6>
- Buchanan, S., & Coleman, S. (1979). Deterioration Survey of the Stanford University Libraries Green Library stack collection. Republished in Darling, P.W., (comp.). In , 1987. *Preservation Planning Program Resource Notebook* (pp. 189–221). Association of Research Libraries Office of Management Studies. [https://archive.org/details/ERIC\\_ED282561](https://archive.org/details/ERIC_ED282561).
- Durant, F. (2019). Preservation assessments. In K. A. Edwards, & M. Leonard (Eds.), *Assessment strategies in technical services* (pp. 183–217). ALA Editions. <https://doi.org/10.17613/8yq2-k252>.
- Eden, P., Bell, N., Dungworth, N., & Matthews, G. (1998). Preservation needs assessment in libraries and archives: Piecing together the national jigsaw. *Library Management*, 19(4), 228–237. <https://doi.org/10.1108/01435129810213325>
- Frank, R. D., & Rothfritz, L. (2022). Designated community: Uncertainty and risk. *Journal of Documentation*, 79(4), 880–897. <https://doi.org/10.1108/JD-07-2022-0161>
- Harnad, S. (1995, May 12). *The PostGutenberg galaxy: How to get there from here. Times Higher Education Supplement*. <https://eprints.soton.ac.uk/253349/1/thes.html>.
- Kelly, M., Nelson, M. L., & Weigle, M. C. (2014). The Archival Acid Test: Evaluating archive performance on advanced HTML and JavaScript. *IEEE/ACM Joint Conference on Digital Libraries*, 25–28. <https://doi.org/10.1109/JCDL.2014.6970146>
- Kenney, A. R., & Wesley, K. (2016). Building a social compact for preserving e-journals. *The Serials Librarian*, 70(1–4), 72–84. <https://doi.org/10.1080/0361526X.2016.1141630>
- Kiesel, J., Kneist, F., Alshomary, M., Stein, B., Hagen, M., & Potthast, M. (2018). Reproducible web corpora: Interactive archiving with automatic quality assessment. *Journal of Data and Information Quality*, 10(4). <https://doi.org/10.1145/3239574>, 17:1-17:25.
- King, L. A., Kovacs, D., & Okerson, A. (1994). *Directory of electronic journals, newsletters, and academic discussion lists*. Association of Research Libraries, Office of Scientific and Academic Publishing. [https://hdl.handle.net/2027/mdp.39015079927508?url\\_append=%3Bseq=3](https://hdl.handle.net/2027/mdp.39015079927508?url_append=%3Bseq=3).
- Kurth, M., Carroll, J., Kara, B., Kehoe, B., Witte, B., Spear, J., Rieger, O., & Renfro, P. (2011). Final report of the 2CUL LOCKSS assessment team. <https://www.2cul.org/sites/default/files/2CULLOCKSSFinalReport.pdf>.
- Lancaster, F. W. (1995). Attitudes in academia toward feasibility and desirability of networked scholarly publishing. *Library Trends*, 43(4), 741–752. <http://hdl.handle.net/2142/7983>.
- Mogge, D. (1999). Seven years of tracking electronic publishing: The ARL directory of electronic journals, newsletters and academic discussion lists. *Library Hi Tech*, 17(1), 17–25. <https://doi.org/10.1108/0737839910267154>
- Okerson, A., & O'Donnell, J. J. (1995). *Scholarly journals at the crossroads: A subversive proposal for electronic publishing*. Washington, D.C.: Office of Scientific & Academic Publishing, Association of Research Libraries. <https://hdl.handle.net/2027/mdp.39015034923758>.
- Rinio, T. (2016). Collection condition assessment in a midsized academic library. *Collection Management*, 41(4), 193–208. <https://doi.org/10.1080/01462679.2016.1227289>
- Roes, H. (1994). *Electronic journals: A survey of the literature and the net*. OiTiO. [https://oitio.eu/publications/ej\\_join.html](https://oitio.eu/publications/ej_join.html).
- Roes, H. (1996). *Electronic journals: A short history and recent developments*. OiTiO. [https://www.oitio.eu/publications/ej\\_1996.html](https://www.oitio.eu/publications/ej_1996.html).
- Segaetsho, T. (2014). Preservation risk assessment survey of the University of Botswana Library. *African Journal of Library, Archives and Information Science*, 24(2), 175.
- Spunaugle, E. D. (2022). Book condition survey for baseline assessment of under-resourced special collections. *Archival Issues*, 41(2), Article 2. <https://doi.org/10.31274/archivalissues.13881>
- VandeCreek, D. (2022). Where are they now? The 2020 status of early (1996–2003) online digital humanities projects and an analysis of institutional factors correlated to their survival. *Preservation, Digital Technology & Culture*, 51(3), 91–109. <https://doi.org/10.1515/pdtc-2022-0011>
- Walker, A. (2003). Preservation assessment surveys: An interdisciplinary approach. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 13 (3–4). <https://doi.org/10.18352/lq.7740>, Article 3–4.
- Walker, R. G., Greenfield, J., & Fox, J. (1985). The Yale Survey: A large-scale study of book deterioration in the Yale University Library. <https://www.ideals.illinois.edu/items/40813/bitstreams/124241/data.pdf>.
- Weibel, S. (1995). The World Wide Web and emerging internet resource discovery standards for scholarly literature. *Library Trends*, 43(4), 627–644. <http://hdl.handle.net/2142/7994>.
- Willer, A. M., & Szydlowski, N. (2012, January 20). *Preservation assessment of MIT E-journal holdings [conference presentation]*. ALA midwinter conference, preservation administrators interest group and digital preservation interest group, Dallas, TX, United States.
- Wolven, R. A. (2015). *Strategies for expanding e-journal preservation: Final report to the Andrew W. Mellon Foundation*. <https://doi.org/10.7916/D8KK9HJZ>