IJDC | *Research Paper*

Examining Persistence of European Open Repository Infrastructure and its Diffusion in the Scholarly Record

George Macgregor
University of Glasgow

Joy Davidson
Digital Curation Centre,
University of Glasgow

Abstract

This article seeks to determine the extent to which the principle of persistence is observed by repositories and the organizations that operate them. In addition, the impact that negative repository persistence levels may be having on the scholarly record is assessed. This is achieved by interrogating and combining data on European repositories from several repository registries and web-scraped sources, including the Internet Archive's Wayback Machine, which creates a unique dataset of historic repository locations and their OAI-PMH endpoints. Then, this data is used as the basis for text mining CORE, a vast corpus of scholarly outputs, to determine the extent to which impersistent European repository content has permeated the scholarly literature. The findings indicate that over one-fifth of European repositories (greater than 20 per cent) could be classified as "dead," with an even greater proportion (greater than 40 per cent) of the machine interfaces associated with these repositories similarly dead. Problematically, this analysis indicates that approximately 12,000 unique scholarly works cite, refer to, or actively use this repository content, amounting to approximately 19,000 unique repository locations, all of which are now dead and unretrievable from their stated resource locations. Partly owing to limitations in available repository registry data and the existence of "zombie" repositories, there are reasons to conclude that the total number of scholarly works referring to dead repository content is far higher. In addition, evidence of dead repository content entering the current scholarly record is found, which is described as "dead on arrival" referencing. The implications of these observations are considered, explanations are proffered, and possible policy interventions to address repository persistence are proposed. The dataset also enables several observations on the nature of impersistent repositories to be made, including their profile and their decay rate.

Submitted 24 December 2025 ~ *Accepted* 24 June 2026

Correspondence should be addressed to George Macgregor. Email: George.Macgregor@glasgow.ac.uk

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The expansion of open scholarly repositories within research institutions has significantly reshaped the way research is conducted and shared (European Commission: Directorate-General for Research and Innovation, DCC, DANS, EFIS & Visionary Analytics, 2022; Suber, 2012). The increasing accessibility of content made available by repositories enables a wider audience to engage with digital research resources, fostering an open research environment with its many advantages. A need for "trustworthy" repositories has emerged as an essential enabler of a functioning and healthy scholarly ecosystem, in which research resources demonstrate FAIRness (European Commission: Directorate-General for Research and Innovation, 2018). This is because repositories have become key components of global open scholarly infrastructure, supporting the sharing of publications, data, and other resources.

However, a growing crisis in the persistence of scholarly research resources has emerged. Despite the long-standing recognition that stable addressing on the web is critical to resource persistence (Berners-Lee, 1998), many scholarly resources are disappearing (Jones et al., 2016; Klein et al., 2014). This crisis coalesces with wider concerns about the reproducibility of scholarly research, the ability of scholars to verify findings, and broader questions about trust and integrity within scholarship more generally (Baker, 2016; Brembs, 2018; Gertler & Bullock, 2017; Helgesson & Bülow, 2023). Technical mechanisms devised to better support the persistence of scholarly resources have also been found to be unreliable, exacerbating these concerns (Klein & Balakireva, 2022).

Scholarly repositories, in their various permutations, have long been considered reliable nodes within the global open scholarly infrastructure. From the beginning, they were acknowledged as "essential infrastructure for scholarship" by enabling the discovery of digital content, delivering persistent access to the scholarly record, and supporting the long-term stewardship of digital resources (Lynch, 2003). Repositories have consequently tended to be operated by research organizations with expertise in digital object management, notably university research libraries. Although these organizations purport to observe digital object management best practice, this does not always appear to be reflected in practice. Emerging evidence suggests that some repository instances, and the organizations that operate them, are demonstrating poor repository management by moving content in unpredictable ways, mismanaging repository Uniform Resource Identifiers (URIs), or retiring repositories entirely (Bamgbose, Mohd, Tengku Wook, & Mohamed, 2025; Macgregor, 2025; Strecker, Pampel, Schabinger, & Weisweiler, 2023). The explanations for why this is happening can be complex (Bamgbose, Mohd, Tengku Wook, & Mohamed, 2025; Rothfritz, Matthias, Pampel, & Wrzesinski, 2025). Rather than mitigating the persistence crisis, many such organizations appear to be contributing to it, compromising the integrity of the scholarly record and hindering content discovery, research citation, verification, and reproducibility, all of which are essential to academic trust.

This article attempts to determine the extent to which the principle of persistence is observed by repositories and the organizations that operate them. A better understanding of this problem space motivates this research. The impact that negative repository persistence levels may be having on the scholarly record is evaluated. This is achieved by interrogating and combining data about European repositories from several repository registries and web-scraped sources, and using this data as the basis for text mining CORE[1] (Knoth et al., 2023), a vast corpus of scholarly outputs, to determine the extent to which impersistent European repository content has permeated the scholarly literature. The implications of these observations are considered, explanations are proffered, and possible policy interventions to address the issue of repository persistence are proposed. To assist the observations and analysis, the notion of "dead" and "zombie" repositories is defined. The dataset also enables several observations about the nature of impersistent repositories, their profiles, and their decay rates. Although the reason a repository may be dead cannot always be determined, this article contributes to an improved understanding

[1] CORE: Connecting Repositories: https://core.ac.uk/

of the threat to the digital scholarly record posed by scholarly infrastructure otherwise perceived as persistent. It also complements the wider, growing body of evidence examining "link rot," "reference rot," and the persistence crisis more generally.

Background and Context

URI Persistence

The growing fragility of some web resources and the need for their URIs to be managed responsibly were drawn to wider attention by Tim Berners-Lee (1998). Berners-Lee noted that: "When someone follows a link and it breaks, they generally lose confidence in the owner of the server [and] are frustrated from accomplishing their goal." Berners-Lee articulated a series of principles to which content publishers should adhere to ensure the persistence of resources. These included the allocation of URIs that "should not change" and were stable over a long time (i.e., 20–200 years) and that those creating URIs are mindful of their "design" such that they support long-term management (Berners-Lee, 1998). The centrality of persistent URIs to W3C initiatives such as the Semantic Web and Linked Data meant that Berners-Lee's principles were later formalised through W3C recommendations, including guidelines on "Cool URIs for the Semantic Web" (Sauermann & Cyganiak, 2008). Similarly, as nodes within open scholarly infrastructure and as scholarly data hubs, an adherence to responsible URI management within repositories has long been considered essential (Hockx-Yu, 2006; Nelson & Allen, 2002; Shreeves & Cragin, 2008). The importance of Cool URIs within the publication of FAIR data on repositories and the associated knowledge graphs arising from such publication has therefore also recently emerged (Thalhammer, 2024), highlighting the continued relevance of creating stable, secure, and persistent URIs.

Persistence and the Scholarly Web

Despite an early understanding of the problems impersistent URIs present for digital libraries (Phelps & Wilensky, 2000; Shreeves & Cragin, 2008) and the integrity of scholarship (Carnevale & Aronsky, 2007; Ducut, Liu, & Fontelo, 2008; Eysenbach & Trudel, 2005; Russell & Kane, 2008; Spinellis, 2003), research continues to uncover concerning levels of "reference rot" within the scholarly record. In a deep analysis of scholarly literature within the Science, Technology, and Medicine domains, Klein et al. (2014) reported that one-fifth of links to URI references suffered reference rot. In subsequent work, members of the same research group further revealed significant levels of "content drift" in URI references (Jones et al., 2016), with greater than 75 per cent of references pointing to web content that had changed since its original citation. This line of research was further refined by Massicotte and Botter (2017), who examined the extent of reference rot and content drift in doctoral theses and dissertations from a large university repository. Their findings were proportionally consistent with Klein et al. (2014), with 23 per cent of thesis references demonstrating link rot, although their detection of content drift was slightly lower, at approximately 50 per cent. The results from these studies and others like them (Martin-Segura et al., 2022) contribute to broader concerns about the integrity of the scholarly record and the ability of future scholars to verify assertions or findings reported in the scholarly literature.

Theoretical work performed by Romero (2025) concluded that trust is both a critical and a desired effect of improved transparency in scholarship. A lack of transparency, arising, for example, from the inability to verify scientific assertions in the literature, is highly damaging to trust in research findings and scholarship more generally.

Attempts to prevent link or reference rot from occurring in the first place are a complex sociotechnological challenge, requiring longer-term behavioural change across scholarship and digital publishing more generally (da Silva, 2021). This includes improved digital literacy among scholars, some of whom struggle to cite digital scholarly objects accurately (Van de Sompel, Klein, & Jones, 2016) or to correctly identify commonplace persistent identifiers, such as digital object identifiers (Macgregor, Lancho-Barrantes, & Pennington, 2023). Therefore, research has sought to mitigate the impact of reference rot by proposing innovative interventions. In an early study testing the retrieval of "decayed" online citations from works published in the early 2000s, Dimitrova and Bugeja (2007) reported that approximately 53 per cent of the citations they tested were retrievable using the Wayback Machine, the most reliable of the systems they evaluated. They concluded that the Wayback Machine was too unreliable to depend on for scholarship and noted concerns about how decaying citations "threatened" the scientific method.

However, as the Wayback Machine has evolved, so too have the possibilities for harnessing machine methods to access its data to resolve citation gaps, where they might occur. A rich stream of research has enabled a framework for time-based content negotiation, delivering the Memento Protocol (Jones, Klein, Van de Sompel, Nelson, & Weigle, 2021; Klein, Balakireva, & Shankar, 2019; Klein, Shankar, Balakireva, & Van de Sompel, 2019; Zhou, Grover, Klein, & Tobin, 2015). Initially a technical proposal (Van de Sompel et al., 2009), the Memento Protocol (Van de Sompel, Nelson, & Sanderson, 2013) has since been embedded in some scholarly tools to solve the content negotiation associated with reference rot (Mahanama, Balakireva, Jayarathna, Nelson, & Klein, 2022). "Time travel" and archiving within Memento-aware web browsers have also been demonstrated (Mabe, Nelson, & Weigle, 2021, 2022). Subsequent work has sought to improve the strength of hyperlinks by proposing "Robust Links," an approach that seeks to archive web resources at the point of their citation and include machine-actionable annotations in the link so that the resource can be accessed in both a live and an archived state (Jones, Klein, & Van de Sompel, 2021; Klein, Shankar, & Van de Sompel, 2018).

Persistence: Repositories as Scholarly Infrastructure

Work to ameliorate reference rot, such as the Memento Protocol and Robust Links, has largely arisen in response to the ephemeral nature of some web resources (Klein, Shankar, & Van de Sompel, 2018; Van de Sompel et al., 2009). Open scholarly repositories, whether institutional, subject-based, publication-focused, or data-centric, are usually contrasted with the wider ephemeral web as reliable and persistent components of scholarly infrastructure. From their inception, these repositories were considered as essential infrastructure, because they enhanced the discovery of digital scholarly content and supported the goals of the open access movement (Suber, 2012) and they ensured persistent access to scholarly materials and supported the long-term stewardship of digital resources (Crow, 2002; Francke, Gamalielsson, & Lundell, 2017; Hockx-Yu, 2006; Lynch, 2002, 2003). For these reasons, many open scholarly repositories have been operated under the auspices of research libraries. This has typically been explained by their operational overlap with other digital libraries or archives (Burns, Lana, & Budd, 2013). A recent survey of European repositories conducted by the Confederation of Open Access Repositories (COAR),[2] Ligue des Bibliothèques Européennes de Recherche (LIBER),[3] Open Access Infrastructure for Research in (OpenAIRE),[4] and Scholarly Publishing and Academic Resources Coalition Europe (SPARC) Europe[5] confirms the continued centrality of the university library to repository management (Shearer et al., 2023).

Repositories are beneficial services in which digital content can be deposited, stored, and managed, affording considerable discovery advantages to authors and organizations (Dong & Tay, 2023; Kelly, 2023; Thompson & Hoover, 2023). In general, they also remain conducive to URI persistence, as they are typically built with open-source software (e.g., DSpace, Invenio, and EPrints) and have been designed to support responsible URI management concepts (Macgregor, 2025). This will typically entail a URI syntax that predictably describes the scheme → subdomain → domain → folder/context or accession identifier (Thalhammer, 2024), resulting in simple, predictable, and less brittle URIs.

It might be expected that repository content, which uniquely underpins research and, ergo, the scientific method, would demonstrate greater persistence. However, a study by Strecker,

[2] COAR: https://coar-repositories.org/
[3] LIBER (Association of European Research Libraries): https://libereurope.eu/
[4] OpenAIRE: https://www.openaire.eu/
[5] SPARC Europe: https://sparceurope.org/

Pampel, Schabinger, and Weisweiler (2023) using data from the re3data data repository registry found that repository retirement was not uncommon, reporting that 6.2 per cent of research data repositories indexed in the registry had been retired. They described the negative implications for the scholarly record while also noting the lack of communication about "repository shutdown events." In an exploratory pilot study of UK-based repository persistence, Macgregor (2025) indicated that 45 per cent and 31 per cent of UK repository OAI-PMH endpoints and repository home Uniform Resource Locator (URL) locations, respectively, were "dead," with a combination of repository response errors reported (e.g., 404 or 400 responses and NXDOMAIN). Perhaps for these reasons, scholars are increasingly encouraged to use trustworthy digital repositories (TDRs) to share their work. However, organizational mismanagement of repositories significantly undermines trustworthiness.

On the theme of "trust" as it relates to data repositories, Yakel, Faniel, and Robert (2024) found that scholars' trust in reusing a data repository was predicted by the observable behavioural actions of the managing organization. Therefore, integrity, which is central to the concept of trust, is not established simply by stating, for example, a commitment to preserving data over time, but by behaving in a manner that demonstrates adherence to that commitment. The results described by Strecker, Pampel, Schabinger, and Weisweiler (2023) therefore represent an unhelpful development in the wider, positive need to ensure scholarly resources are deposited, disseminated, shared, and responsibly maintained over time. In addition, it undermines important pan-organizational initiatives, such as IMPACT-REPO, which seeks to safeguard Europe's repository infrastructure by "ensuring research remains accessible, trusted, and reusable" (Shearer et al., 2025).

As evidence emerges of their potential "untrustworthiness," the reliability of repositories as nodes within open scholarly infrastructure has recently been scrutinised. A systematic review performed by Bamgbose, Mohd, Tengku Wook, and Mohamed (2025) sought to understand the extent to which there was "trustworthiness" of the repositories operated by research libraries. In a systematic review, Bamgbose, Mohd, Tengku Wook, and Mohamed (2025) could not arrive at a measure of untrustworthiness; however, they were able to conclude that, while most repositories were deemed trustworthy, there were widespread concerns in the literature about the inadequate technical support or resources often afforded to repositories, as well as the financial constraints and librarian "disinterest." An overreliance by organizations on short-term funding arrangements for repository operations (sometimes between two and five years), as well as a shortage of suitably qualified human resources, has been observed to contribute to the challenges of longer-term repository management, threatening trustworthiness (Jouneau et al., 2025). Such an observation is arguably compounded by the importance of transparency in supporting trustworthiness (Romero, 2025) and the related importance of visible and understandable repository policies, which may be absent (Jouneau et al., 2025) because their transparency at least enables scholars to make informed decisions about where to deposit scholarly work (Yakel, Faniel, Kriesberg, & Yoon, 2013).

Rothfritz, Matthias, Pampel, and Wrzesinski (2025) conducted a broader systematic review to develop a greater understanding of the challenges confronting institutional repositories. The financial sustainability of repositories was again identified as a particular challenge in all global regions, with low acknowledgement of the unintended consequences of staffing constraints and resource limitations of repository mismanagement. In support of the IMPACT-REPO initiative, Shearer et al. (2025) tacitly acknowledge the challenging environment confronting repository infrastructure. They reiterate the need for organizations and the wider scholarly community to recognise their strategic importance and invest to secure their future, including superior observance of the FAIR principles, greater use of persistent identifiers, and the upskilling of staff to ensure repository management "excellence" (Shearer et al., 2025).

Research Motivation: Dead and Zombie Repositories

This research is motivated by a need to better understand the size and nature of an emerging repository persistence crisis. Rather than functioning as reliable nodes for the dissemination and management of scholarly research objects and helping address the persistence and reference rot pervading the scholarly web at large, emerging evidence suggests that some repositories are exacerbating the problem. The extent to which low levels of repository persistence may be affecting the integrity of the scholarly record is assessed. To achieve this, data about European repositories from multiple sources is aggregated and analysed, including various repository registries and web-scraped datasets. Europe is home to the largest concentration of repositories on any continent (Nazim & Bhardwaj, 2023). Therefore, it provides a sufficiently large indicator of global repository persistence while reducing the computational overhead associated with text mining all global literature. This approach enables a unique dataset to be constructed that captures historical repository locations and their OAI-PMH endpoints. Using this dataset, text mining of CORE (Knoth et al., 2023), a comprehensive collection of scholarly outputs, is conducted to evaluate the degree to which unstable European repository content has been referenced or incorporated into the scholarly literature.

To assist in the discussions and analysis, two impersistent repositories are identified: "dead" repositories and "zombie" repositories.

A "dead" repository can be defined as follows:

- A once active repository that had a history of successfully serving content, but which suddenly becomes unavailable over the Internet, either through active retirement or neglect. Therefore, these repositories will tend to demonstrate domain name registration errors (i.e., NXDOMAIN) or report HTTP 404 responses. These dead repositories will not display evidence of Cool URI management, normally because the repository service has been terminated or has been poorly maintained.

A subcategory of a dead repository is a "zombie" repository. A zombie repository can be defined as:

- A once active repository that had a history of successfully serving content before suddenly becoming unavailable. However, unlike a dead repository, it is often "reanimated" under a new domain, typically using different software and serving content from alternative locations. Limited Cool URI management is usually displayed (e.g., reusing existing URIs or formally redirecting users/machines to a new location). Users/machines are also unaware that the repository location has changed, and persistence is therefore disrupted.

Of note, zombie repositories are also dead ones; therefore, they are deemed a subcategory. This distinction is highlighted because these zombie approaches to repository management are equally as disruptive to content persistence as a terminally dead repository might be.

Methods: Data Collection and Preparation

Repository Registry Data

To test the persistence of repositories, it is necessary to consult an inventory of their existence. Such repository registries exist, although their data may have limitations (Walk, 2023). Baglioni, Pavone, Mannocci, and Manghi (2025) proposed a data model to improve the interoperability of scholarly repository registries, given their diversity and disparate data models. To build a suitable registry dataset for this study, a dataset from three sources using a combination of API interrogation and web scraping is curated:

1. Open Directory of Open Access Repositories (OpenDOAR): OpenDOAR[6] Data are exposed as JSON objects via the Jisc Open Policy Finder API (formerly the Sherpa API).[7] Using the API, registry data pertaining to European repositories were queried and

[6] OpenDOAR: https://opendoar.ac.uk/
[7] Jisc Open Policy Finder: https://openpolicyfinder.jisc.ac.uk/

extracted, capturing key registry data, including repository name, home URL, OAI-PMH endpoint location, and country code;

2. Registry of Open Access Repositories (ROAR): Although ROAR[8] does not provide a public API; it is possible to perform a full JSON export of its registry data on European repositories directly from its user interface. Data captured overlaps with OpenDOAR but includes key registry data, such as repository name, home URL, and OAI endpoint.

Both OpenDOAR and ROAR are primarily concerned with maintaining a contemporaneous registry of repositories and do not model changes arising from zombie repositories. Walk (2023) has, for example, reported staleness and gaps in OpenDOAR's registry data. Therefore, the OpenDOAR and ROAR data are combined with the Institutional Archive Registry (IAR):

3. The IAR was the precursor service to ROAR and was retired in the mid-2000s. A snapshot from the Wayback Machine is used, captured by the Internet Archive on 13 June 2006, when 750 repositories were registered with the IAR. Using web scraping techniques, 301 repositories identified as falling within the European scope were extracted and scraped for associated text and hyperlink data about each registry entry. Key registry data elements were gathered using this process, although country codes had to be assigned programmatically.

All three registry sources categorise repository types differently. OpenDOAR records five repository categories. These include institutional, disciplinary, aggregating, governmental, and undetermined. ROAR supports 12, although many of these could be interpreted as subcategories of those provided by OpenDOAR. To enable aggregate observations of repository type data, representative examples of the ROAR and IAR repository types were inspected and mapped to the OpenDOAR types, as given in Table 1.

Data from all three sources were merged, cleaned, and deduplicated, with redundant data removed. Superfluous URL query parameters, erroneously present in some repository registry entries, were removed to identify the next best stable nonparameterised URL. The resulting dataset, shared as part of this article, represents a unique inventory of active and historic repositories, including dead and zombie repositories. It forms a benchmark from which it is possible to make assessments about repository persistence over time and the scholarly impact of impersistence, where it exists.

Table 1. Mapping of repository types from ROAR and IAR registry data to the OpenDOAR repository type vocabulary.

IAR repository type	ROAR repository type	OpenDOAR repository type
Research institutional or departmental	Research institutional or departmental	Institutional
Not applicable	Research multi-institution repository	Institutional
Research cross-institution	Research cross-institutional	Institutional
Not applicable	Subject	Disciplinary
e-journal/publication	e-journal/publication	Undetermined
e-theses	e-theses	Institutional
Database	Database	Aggregating
Not applicable	Research data	Institutional

[8] ROAR: https://roar.eprints.org/

Not applicable	Open data	Aggregating
Not applicable	Learning and teaching objects	Institutional
Demonstration	Demonstration	Undetermined
Not applicable	Web observatory	Undetermined
Other	Other	Disciplinary

HTTP(S) Response Data

To identify dead repositories within this dataset, a script to gather HTTP status request codes for every repository domain URL and its associated OAI-PMH endpoint was deployed. OAI-PMH remains an "essential characteristic" of repository functionality (COAR, 2022), currently providing the only standard machine protocol for repository content discovery and data exchange (Knoth et al., 2023) OAI-PMH facilitates interoperability across heterogeneous repositories, enables metadata harvesting and aggregation, and remains a "cornerstone" protocol of repository workflows (Rubin, Bardi, Del Gratta, & Del Grosso, 2026). Its importance is reflected in repository registry records such as OpenDOAR. The absence of a functioning repository OAI-PMH endpoint, combined with a repository domain URL, is therefore a significant signifier of repository health.

HTTP responses were logged against the repository entries in the registry dataset. Common HTTP response codes are widely documented by the Internet Engineering Task Force (ITEF) and are available for reference (Fielding, Nottingham, & Reschke, 2022). To optimise response accuracy, GET HTTP requests were employed as the authoritative retrieval method and as a more reliable technique for obtaining response metadata and behaviour (Fielding, Nottingham, & Reschke, 2022). These requests were stateless and used distributed cloud infrastructure (Google Cloud) to best model repository service behaviour under disparate network conditions (e.g., production-scale routing behaviour). This is significant because many services dynamically tailor responses based on IP-derived geolocation and content delivery network routing. Therefore, the approach used provides a broader view of operational behaviour than a single fixed client location would allow.

Of note, many registry data in this dataset described repositories before the widespread adoption of HTTPS. HTTP to HTTPS redirects (HTTP 301 response) to essentially the same location, such as http://example-repository.uni.de to https://example-repository.uni.de, were therefore parsed by this script as a 200 response.

Although data pertaining to repository uptime was not something explicitly collected during these experiments, it is possible to report that many repositories demonstrated erratic availability during data collection. This could possibly be associated with recent "bad bot" behaviour by Large Language Model agents (Knoth, 2025; Sherrick & Pino Navarro, 2024). However, other repositories lacked valid SSL certificates, limiting browser or agent access. This meant that manual checks and retests of repositories were often necessary to verify whether unavailability was merely temporary, intermittent, or permanent.

Repository "Date of Decease" Data

The Wayback Availability JSON API[9] was queried to capture an approximate "date of decease" for repositories found to be returning unsatisfactory HTTP response codes. This API is partially based on the Memento Protocol and serves archived snapshot data on the last available website archive, including `archived_snapshot_URL` and associated time stamp. Data retrieved from the API were logged against the repository entries within the registry dataset.

[9] Wayback Availability JSON API: https://archive.org/help/wayback_api.php

CORE: Mining the Scholarly Literature

The final part of data collection concerned mining the scholarly literature to understand the extent to which any impersistent European repository content has entered the scholarly record. CORE (Knoth et al., 2023) was used as the basis for this text mining. CORE represents a vast open corpus of scholarly outputs, providing access to the world's largest collection of open publications. These publications are harvested and aggregated from the global network of repositories and journal titles, enabling user discovery via search but also scientific discoveries using text and data mining (TDM) techniques. CORE has become an important feature of the open scholarly infrastructure landscape (Jefferies et al., 2022) and is increasingly a platform on which value-added services are developed (Knoth et al., 2023).

CORE provides a sophisticated API to expose the scholarly resources it has harvested and enriched, using its CORE API Query Language.[10] A script was deployed to query version 3.0 of the CORE API to determine the extent to which impersistent European repository content has permeated the scholarly literature. This script sought to mine CORE's corpus for scholarly works that cite, refer to, or actively use dead repository content.

JSON responses from the CORE "Works" API were processed to log key bibliographic data elements. This included work title, authors, documentType, doi, identifiers, id (an internal CORE identifier), oaiIds (OAI identifiers associated with repositories), yearPublished, and depositedDate. Data arising from the fullText field were parsed, and any in-text references to dead repository content were extracted, using the URL prefix of offending repositories, with a wildcard. Extracted URLs underwent normalisation and validation. Of note, not all works in CORE have full text available for TDM, but many include data enrichments performed by CORE (e.g., reference data from CrossRef). This is a limitation discussed later in this article.

All data were captured and the dataset compiled during June 2025.

Results

Dead Repositories

Data gathered and merged from the repository registry sources identified 3,751 repositories falling within the European scope. The characteristics of these repositories are summarised in Table 2; the vast majority were "institutional" (approximately 82 per cent), with "disciplinary" accounting for the next-largest repository type (approximately 12 per cent). Overview results arising from the HTTP response data to identify dead repositories are also summarised in Table 2. A total of 824 repositories were deceased, equivalent to 22 per cent of all repositories within the dataset, leaving 2,927 active repositories. Of note, institutional repositories were the most common type of dead repository (n = 620), accounting for less than 75 per cent of all repositories found to be deceased. As a highly represented repository type within the dataset, this is perhaps unsurprising; however, it should be noted that as a proportion of all repositories within this dataset, this indicates that one-fifth (greater than 20 per cent) of all institutional repositories are now deceased. The strong presence of disciplinary repositories (n = 147) should be noted as being almost double the combined total of other dead repository types ("aggregating," "governmental," and "undetermined").

Table 2. Summary data by repository type within the analysed dataset and summary observations of dead repositories. Based on domain name HTTP response status data.

Repository type	Total by type	Total by type (%)	Dead count	Dead as % of total dead count	Dead as % of repository type	Dead as % of all repository types

[10] CORE API (3.0.0): https://api.core.ac.uk/docs/v3

Institutional	3,081	82.1	620	75.2	20.1	16.5
Disciplinary	445	11.9	147	17.8	33.0	3.9
Aggregating	104	2.8	25	3.0	24.0	0.7
Governmental	59	1.6	10	1.2	16.9	0.3
Undetermined	62	1.7	22	2.7	35.5	0.6
Totals	3,751	100	824	100	–	22

The HTTP response data overview in Table 2 includes assessments of repository death based on the home domain name of a repository. This method sought similar data based on the OAI-PMH endpoints associated with repositories in this dataset because support for this protocol is central to our understanding of repositories, their discovery, and data reuse (Macgregor, 2023). As described in the results and discussion, the response status of an OAI-PMH endpoint can also indicate a zombie repository. Data arising from analysis of OAI-PMH responses is given in Table 3.

A total of 1,643 OAI-PMH endpoints within this dataset were dead, with greater than 78 per cent (n = 1289) of these dead endpoints belonging to institutional repositories. This is equivalent to greater than 34 per cent of all repositories in the dataset. Combined with the dead endpoints identified for all repository types, well over one-third of repository OAI-PMH endpoints (approximately 44 per cent) in this dataset can be reported as dead. Significant proportional disparities were observed between repository types. Institutional and disciplinary repositories contribute to the overall volume measure of dead repositories, but within "governmental" and "undetermined" there are approximately 58 per cent and 52 per cent dead endpoints, respectively. For example, this dataset identified 59 governmental repositories, of which 34 appeared to have a dead OAI-PMH endpoint. These repository types may be small relative to this dataset but generate massive proportional results in their respective category type.

Table 3. Summary data by repository type within the analysed dataset and summary observations of dead repositories. Based on HTTP response status of associated OAI-PMH endpoints.

Repository type	Total by type	Total by type (%)	Dead count	Dead as % of total dead count	Dead as % of repository type	Dead as % of all repository types
Institutional	3,081	82.1	1,289	78.5	41.8	34.4
Disciplinary	445	11.9	243	14.8	54.6	6.5
Aggregating	104	2.8	45	2.7	43.3	1.2
Governmental	59	1.6	34	2.1	57.6	0.9
Undetermined	62	1.7	32	1.9	51.6	0.9
Totals	3,751	100	1,643	100.0	–	43.8

National profiles emerging from repositories in this dataset can be summarised, as given in Table 4. Data are organised by two-letter country codes, as per ISO 3166-1 alpha-2 (International Organization for Standardization, 2020), and ranked by dead repository count (domain). In total, 41 countries are represented. Larger countries have larger total repository counts but, perhaps disappointingly, also larger counts of dead repositories. The UK (gb) is the leader based on count, with approximately 24 per cent of its 528 repositories and approximately 41 per cent of associated OAI-PMH endpoints found to be dead. Percentage measures of dead repositories suggest significant impact for small categories (e.g., Luxembourg (lu), Belgium (be), and Bulgaria (bg)), where the proportional impact of dead repositories was as high as 80 per cent. This is shown in Figure 1, in which, at times, the tail of dead repositories by country (at the domain level) suggests an inverse proportional increase in some regions; however, regression analysis revealed no statistically significant association (R^2 = 0.006, p = 0.629).

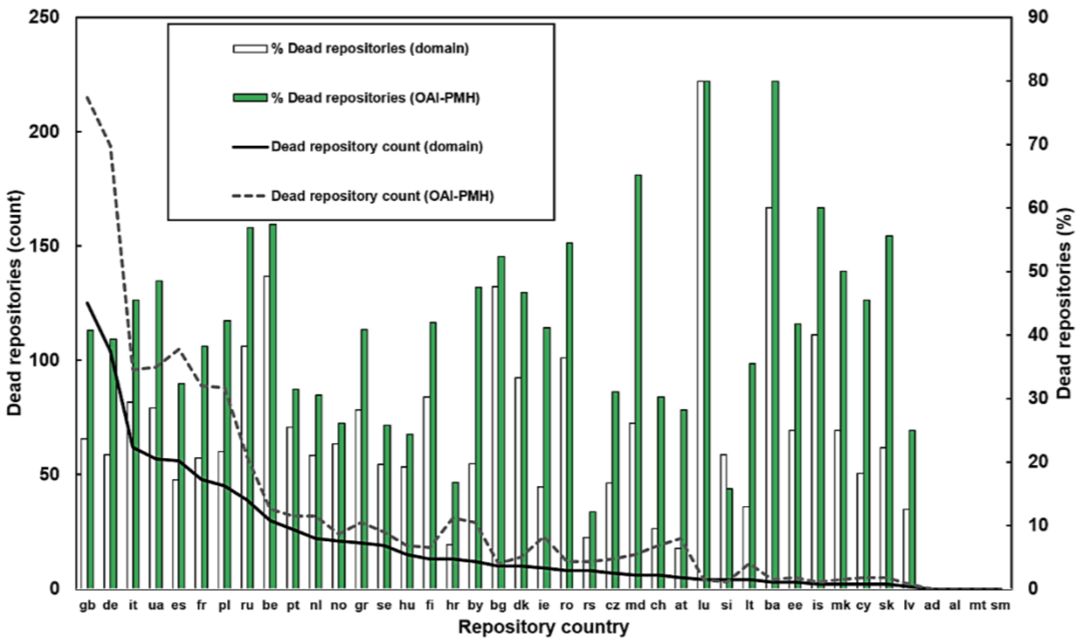

Figure 1. Number of dead repositories at domain and OAI-PMH level by country and as a percentage of total repositories within European countries.

The general disparity between dead repository domains and their associated OAI-PMH endpoints is highlighted in Tables 2 and 3. However, the differential between dead repository domains and endpoints by country is given in Table 4. Both the UK (gb) and Germany (de) report the highest number of dead repository domains and OAI-PMH endpoints, and the largest differential between the two. The significance of this differential is explored in the discussion. Slovenia (si) was the only region in which the importance of the OAI-PMH endpoint was recognised, and Cool URI management appeared evident.

Table 4. Summary counts and percentages of dead repositories by domain and OAI-PMH endpoint, categorised and ordered by two-letter country code, ISO 3166-1 alpha-2 (International Organization for Standardization, 2020).

Repository country	Total count	Dead count (domain)	Dead count (%) (domain)	Dead count (OAI-PMH)	Dead count (%) (OAI-PMH)	Domain/OAI-PMH differential	Domain/OAI-PMH differential (%)
ad	1	0	0.0	0	0.0	0	0.0
al	3	0	0.0	0	0.0	0	0.0
at	78	5	6.4	22	28.2	17	21.8
ba	5	3	60.0	4	80.0	1	20.0
be	61	30	49.2	35	57.4	5	8.2
bg	21	10	47.6	11	52.4	1	4.8

by	61	12	19.7	29	47.5	17	27.9
ch	63	6	9.5	19	30.2	13	20.6
cy	11	2	18.2	5	45.5	3	27.3
cz	42	7	16.7	13	31.0	6	14.3
de	494	104	21.1	194	39.3	90	18.2
dk	30	10	33.3	14	46.7	4	13.3
ee	12	3	25.0	5	41.7	2	16.7
es	325	56	17.2	105	32.3	49	15.1
fi	43	13	30.2	18	41.9	5	11.6
fr	234	49	20.9	89	38.0	40	17.1
gb	528	125	23.7	215	40.7	90	17.0
gr	71	20	28.2	29	40.8	9	12.7
hr	185	13	7.0	31	16.8	18	9.7
hu	78	15	19.2	19	24.4	4	5.1
ie	56	9	16.1	23	41.1	14	25.0
is	5	2	40.0	3	60.0	1	20.0
it	211	62	29.4	96	45.5	34	16.1
lt	31	4	12.9	11	35.5	7	22.6
lu	5	4	80.0	4	80.0	0	0.0
lv	8	1	12.5	2	25.0	1	12.5
md	23	6	26.1	15	65.2	9	39.1
mk	8	2	25.0	4	50.0	2	25.0
mt	2	0	0.0	0	0.0	0	0.0
nl	105	22	21.0	32	30.5	10	9.5
no	92	21	22.8	24	26.1	3	3.3
pl	208	45	21.6	88	42.3	43	20.7
pt	102	26	25.5	32	31.4	6	5.9
ro	22	8	36.4	12	54.5	4	18.2
rs	99	8	8.1	12	12.1	4	4.0
ru	102	39	38.2	58	56.9	19	18.6
se	97	19	19.6	25	25.8	6	6.2
si	19	4	21.1	3	15.8	-1	-5.3
sk	9	2	22.2	5	55.6	3	33.3
sm	1	0	0.0	0	0.0	0	0.0
ua	200	57	28.5	97	48.5	40	20.0

Recall the HTTP status request codes for every repository domain URL and its associated OAI-PMH endpoint that were gathered. HTTP responses were logged against the repository entries in the registry dataset. These data reveal the nature of dead repository responses and those that remain alive. They are set out in Table 5.

In total, 2,320 repositories in the wider dataset of 3,751 returned a 200 response, with fewer returning a 200 for their OAI-PMH endpoint (*n* = 1,755). A significant proportion of repositories returned redirection status codes (3XX) at both the domain (*n* = 607) and endpoint levels (*n* = 353). The significant number of 302 responses at the domain level is unusual (*n* = 580) and worthy of comment in the discussion section. However, such a code indicates that the requested resource has moved temporarily but will return to its original location later.

In total, 824 repositories were identified as deceased based on their domains. In Table 5, DNS resolution errors (NXDOMAIN) were recorded in 751 cases, 91 per cent of all negative responses. This NXDOMAIN count almost doubled for OAI-PMH endpoints (*n* = 1,400). A diversity of responses was logged across 5XX and 4XX, with the most common indicator of repository death after NXDOMAIN being a 404 response.

Table 5. HTTP response code counts against repository domains and OAI-PMH endpoints.

Response code categories	HTTP response code	Repository count (domain)	Repository count (OAI-PMH)
Nonexistent domain	NXDOMAIN	751	1,400
Server error responses (5XX)	504	1	0
	503	10	7
	502	3	1
	500	2	8
Client error responses (4XX)	422	0	4
	410	0	3
	404	38	106
	403	18	42
	400	1	67
Subtotal		824	1,638
Redirection status codes (3XX)	308	7	1
	307	13	10
	303	6	2
	302	580	339
	301	1	1
Subtotal		607	353
Successful response	200	2320	1,755
Total		3751	3,751

In separate but related analyses, when a redirection was found at the domain level, a corresponding redirection to the OAI-PMH endpoint was not implemented. Instead, 148 (Table 6) endpoints were dead, with NXDOMAIN the most common response for a repository domain reporting a 302 response. These examples accounted for 67 per cent of all those identified;

although it should be noted that NXDOMAIN arose from all the noted cases of 308, 307, 303, and 301. Therefore, NXDOMAIN was the response found in 110 cases (74 per cent). A mixture of client error responses (4XX) and server error responses covered the small remainder. Therefore, close to one-tenth (9 per cent) of all unresponsive OAI-PMH endpoints were associated with repositories that redirected at the domain level but not at the endpoint level, a possible indicator that these repositories may be zombie repositories.

Table 6. Recorded 3XX responses at repository domain level and corresponding responses at associated OAI-PMH endpoint.

Redirection response (domain)	HTTP response (corresponding OAI-PMH endpoints)	Repository count (OAI-PMH)
308	NXDOMAIN	5
307	NXDOMAIN	3
303	NXDOMAIN	2
302	NXDOMAIN	99
302	503	1
302	410	2
302	404	12
302	400	23
301	NXDOMAIN	1
Total		148

Date of Decease

Wayback machine availability

Of the 824 repositories identified as dead in this dataset, archived snapshots were available via the Wayback Machine Availability API for 680. The date of decease for these repositories begins in 2001 and continues to date; although to control for repositories experiencing recent availability problems, snapshots captured after 1 January 2025 were excluded.

To assess both the absolute growth and underlying dynamics of repository death, linear and logarithmic scaling were employed. Linear and log-linear regression models were evaluated to characterise the growth of cumulative counts over time. Linear was used to model overall trends and support regression analysis, while logarithmic was applied to highlight growth phases and detect potential exponential behaviour.

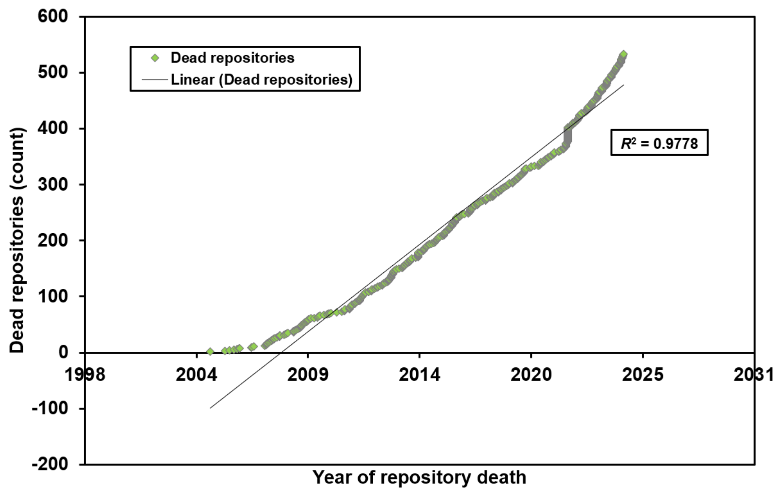

Figure 2. Growth of repository deaths over time, with linear regression included.

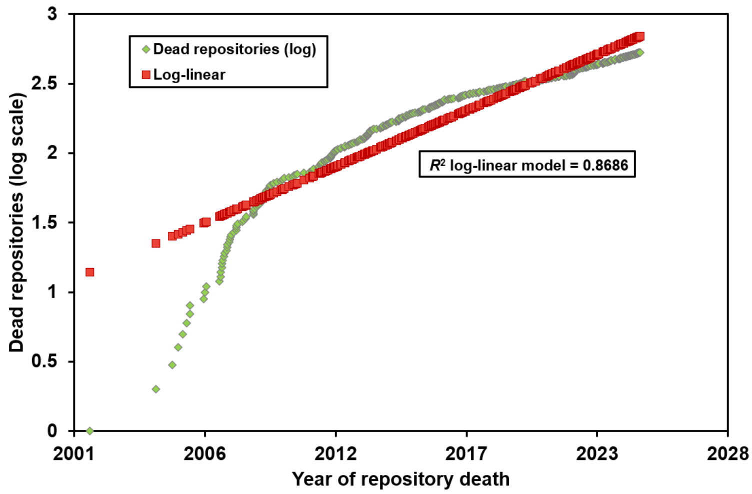

Figure 3. Growth of repository deaths over time (log), with log-linear regression included. *y*-axis values represent the natural logarithm of the data in Figure 2.

Figure 2 shows the temporal profile of dates of decease in the linear model, in which a slight but steady acceleration in the cumulative total of dead repositories in the years following 2011 can be noted. Another period of acceleration occurs in 2022. The linear model demonstrated a superior fit ($R^2 = 0.978$) compared with the log-transformed model ($R^2 = 0.869$) shown in Figure 3, suggesting a strong linear association ($F(1,531) = 21{,}049.13$, $p<0.001$). While the log-linear model revealed a statistically significant exponential component ($F(1,531) = 3{,}510.59$, $p<0.001$) and revealed interesting early-stage dynamics, the growth rate was modest (approximately 0.0202 per cent per time unit), indicating that the cumulative increase was more consistent with additive rather than multiplicative growth. These findings support the use of a linear model for predictive purposes, although it is noted that this observation is based on a subset of Wayback Machine Availability data relating to the wider repository dataset.

Rate of decay

In addition, repository death can be reconceptualised as the rate of decay of once-functioning, so-called “mortal” repositories. This is achieved by employing a classic decay function, which can be calculated where y represents the number of repositories remaining at the end of our temporal range, a represents the number of repositories at the beginning of that range, r represents the rate of decay expressed as a decimal, and *t* represents the time elapsed, which, in this case, is the number of years.

$$y = a(1 - r)^t$$

This yields an annual decay rate of 0.0107, meaning that over 23 years, the number of functioning repositories shrinks by 1.07 per cent every year. At this rate of decay, within a decade, it can be estimated that (i.e., by 2034) the number of functioning European repositories will shrink by a further 442, from the 2024 benchmark of 2,927.

Acknowledging that this data reveals that deceased OAI-PMH endpoints are decaying more rapidly, this analysis can be performed to better quantify OAI-PMH endpoint decay. Here, the decay rate is 0.0237, indicating that the death of OAI-PMH endpoints is far more pronounced at 2.47 per cent per annum. Therefore, the number of mortal OAI-PMH endpoints in 2034 could have declined by a further 681, bringing it down from its 2024 benchmark of 2,108.

Dead Repository Content in the Scholarly Literature

The final section of results concerns mining the scholarly literature to measure the extent to which dead European repository content may have entered the scholarly record through references or citations. The mining seeks to identify citations and references as they are understood in scholarly communication (Nicolaisen, 2007). That is, an acknowledgement of "intellectual debt" within academic text (i.e., citation), which refers the reader to details of a cited literature source (i.e., reference) within a list of works (Garfield, 1996). But they can also be understood as author A making "reference" to author B's work within an academic text, through an in-text reference; while a "citation" is when author A receives acknowledgement from author B through an in-text citation (Albarrán & Ruiz-Castillo, 2011). This mining seeks to identify where repository content may have been cited, referenced, or both. In addition, it identifies the citation of repository handles directly within academic text.

The results from mining the literature identified 12,040 unique scholarly works that cite repository content, which this analysis shows is dead. These unique works generated 19,248 references to dead content. This is because some works reference content in more than one dead repository, or multiple instances of scholarly content from the same dead repository. This is reflected in the associated measures of central tendency (mean (*M*) = 1.6, median (*Md*) = 1, *Range* = 1–70, *IQR* = 1). Most scholarly works reference just one (greater than 67 per cent) or two (greater than 20 per cent) instances of dead repository content (Table 7), but a notable observation is that references to dead repository content within unique works can often be multiple. The upper range is influenced by an extreme outlier (i.e., 70 references to dead content within a single work), and although such double-digit examples do not skew the median, accounting for fewer than 42 references, there are 1,614 examples of works citing or referring to greater than or equal to three dead repository sources.

Table 7. Number of references from a single scholarly work to dead repository content and number of scholarly works within this assigned category.

Number of dead references within work	Number of scholarly works with assigned number of dead references (count)	Number of scholarly works with assigned number of dead references (%)
1	8,075	67.07
2	2,351	19.53
3	906	7.52
4	354	2.94

5	177	1.47
6	79	0.66
7	37	0.31
8	14	0.12
9	5	0.04
>9	42	0.35

One observation is that dead repository content can remain alive in newly published literature for many years after a repository's demise. Of the scholarly works derived from CORE in this dataset, 10,375 had a reliable date of publication, which, combined with the established date of repository decease, could be used to calculate (in years) the extent to which dead content was or was not being cited in new literature. Key measures are summarised in Table 8.

Table 8. Summary of dead and alive repository references in newly published literature, where an accurate date of publication is available (from CORE). Negative values within the "Years" category denote the number of years after repository death that content was cited in published literature.

	All references	"Dead on arrival" repository references	"Alive" repository references
Total count	10,375	2,801	7,575
Total %	100.00	36.98	73.01
	Years (±)	Years (±)	Years (±)
Mean	3.01	−3.57	5.44
Median	3.00	−2.00	5.00
Standard deviation	5.50	3.78	3.77

Most scholarly works (approximately 73 per cent) referenced or cited repository content that was alive when the work was published. However, almost 37 per cent (n = 2,801) of these works cited repository content that was already dead on publication ("dead on arrival"). This repository content was already dead, on average, for greater than three-and-a-half years when the scholarly work was published (M = −3.57, Md = −2.00, and standard deviation = 3.78). In addition, the results appear to identify instances of dead repository content being cited in new literature as many as seven, ten, and sometimes even 17 years after a repository was known to be dead. This suggests that the consequences of repository death can linger in the scholarly record, with authors continuing to cite dead content in new literature long after it no longer exists. This is concerning. Explanations for it are offered in the discussion section.

The spread of "dead on arrival" references is shown in Figure 4, which illustrates the year of repository decease and the "distance" (number of years) between publication of the scholarly work and the year of repository death. This distance may be positive or negative. Positive values on the y-axis denote the number of years the repository was alive after its citation in the scholarly works. Conversely, negative values denote the number and extent of "dead on arrival" references in the published literature.

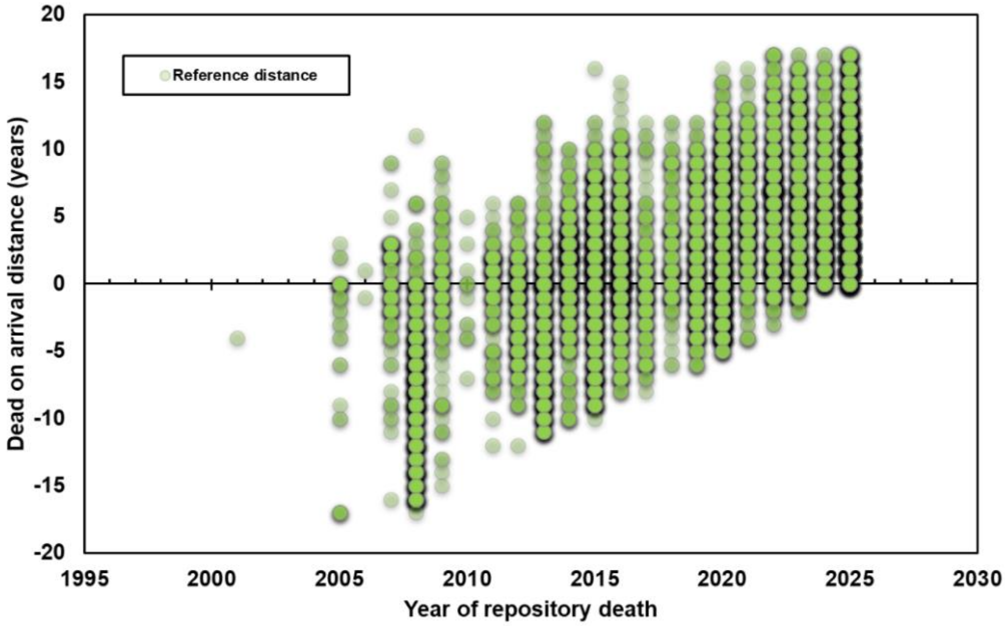

Figure 4. Dead and alive references to repository content within scholarly published literature. Figure highlights "dead on arrival" references. Negative values on *y*-axis indicate level to which repository was already dead (in years) when cited in scholarly literature.

Discussion

Attempts were made to measure repository persistence and the impact impersistance has on the scholarly record. The results make a series of important observations. In the discussion, some of the most significant are commented on.

This article indicates that over one-fifth of the repositories in this dataset are dead, with institutional repositories forming a significant proportion of these deaths (greater than 75 per cent). The machine interfaces of repositories in the dataset were dead in more than one-third of cases. Institutional repositories were highly represented, accounting for almost 80 per cent of all dead OAI-PMH endpoints; although some repository deaths were proportionally large in some category types (e.g., governmental). These findings corroborate emerging concerns that scholarly repositories are often mismanaged, despite their importance to scholarship and open research. This mismanagement, unfortunately, contributes to the persistence crisis and raises questions about the trustworthiness of the custodian organizations of such scholarly resources. This applies particularly to institutional repositories, a dominant organizational manager of which is research libraries.

The nature of this article limits the ability to comment on the organizational reasons for why certain organizations have precipitated repository death. However, the multitude of NXDOMAIN HTTP responses for repository domains and their OAI-PMH endpoints, as well as 4XX and 5XX responses, suggests these repositories have been neglected and poorly maintained, with termination ultimately considered a preferable course of action or indeed where no course of action has been considered at all. Rothfritz, Matthias, Pampel, and Wrzesinski (2025) and Bamgbose, Mohd, Tengku Wook, and Mohamed (2025) highlight the sustainability pressures threatening the operation of repositories, which thereby undermine their trustworthiness. This includes financial constraints and a level of organizational disengagement from repository operations, resulting in low awareness of the consequences arising from repository death. The volume of dead repositories identified in this article is such that a lack of understanding exists in many organisations about the core function of a repository, the implied commitment to long-term access, and the scholarly consequences of shutting one down. Concern about the technical capacity, digital skills, and knowledge transfer within repository-responsible teams, partly related to financial constraints, is probably a contributing factor. Research draws attention to an emerging skills deficit within research libraries, digital archives, and digital libraries, which would impact effective institutional repository management (Cope & Baker, 2017; Recker, Kleemola, & L'Hours, 2024; Tait, Martzoukou, & Reid, 2016), including the "platformisation" of repositories in some research libraries (Plantin & Thomer, 2025). Outsourcing of technical infrastructure usurps digital capacity (SCONUL, 2025), limiting the ability of some organizations to respond responsibly to long-term repository management. Organizations or groups responsible for disciplinary repositories may face separate challenges. These can include governance difficulties, a lack of funding, and disinterest from communities of practice (Björk, 2014; Rieger, 2012).

A notable additional factor, which is probably contributing to the growth of dead institutional repositories, will be the increased deployment of current research information systems (CRIS) at European universities (Biesenbender, Petersohn, & Thiedig, 2019; Bryant et al., 2018). Although CRIS software is designed to support institutional research monitoring (Fabre, Egret, Schöpfel, & Azeroua, 2021) and offers a system purpose and function distinct from repositories, many institutions have been observed merging their functions for organizational convenience (de Castro, Shearer, & Summann, 2014). In a global survey of CRIS deployment at research organizations, Bryant et al. (2018) confirmed a particular intensity of such software use in Europe, as well as a trend towards using it as a repository substitute, with CRIS software functioning as the "default" repository for scholarly literature in greater than 50 per cent of respondents. Schöpfel and Azeroual (2021), Schöpfel et al. (2024), and Fernandes and Pinto (2018) all report on the increased convergence between repositories and CRIS software and the resulting management challenges. Schöpfel et al. (2024) also note that this shift may not benefit open research and "underscores" the critical challenge arising from such a transformation.

This so-called "CRIS-as-IR" trend is worrisome because it often results in a repository's death. It also places primacy on software that is rarely open source, is often ill-suited to long-term digital object management, and rarely displays adequate support for Cool URI management. Therefore, this increases the likelihood that the future location of scholarly resources may again be disrupted, further exacerbating the diffusion of impersistent references in the scholarly literature. At the time of writing, from the Directory of Research Information Systems (DRIS)[11] (maintained by EuroCRIS), Europe is among the strongest regions for adopting CRIS solutions, with the UK, Italy, Spain, Germany, and Poland particularly strong. This may go some way to explaining why some of these countries were among the regions to host the largest concentration of dead repositories.

The analysis of HTTP responses elicited indications that some live repositories were zombie repositories. Recall that zombie repositories are those that are killed but then "reanimated" under a new domain, often using different software and serving content from alternative locations. Indications of this are observed in the implementation of some 3XX redirections, where URI management (via redirection) at the domain level is displayed but not at the OAI-PMH endpoint level, where NXDOMAIN and 404s are instead recorded. Because the OAI-PMH endpoints of these zombie repositories are dead, and the OAI identifiers associated with individual repository resources, it is probable that limited URI management will have been performed on the corresponding URI context or accession identifiers for individual repository resources. In these instances, managing organizations appear to have an awareness that users and machines will be disrupted by creating a zombie repository, and provide a redirection at the domain level. This action could be taken to redirect users who have bookmarked the repository location, or to avoid breaking domain-level hyperlinks elsewhere on the web. But either due to a lack of migration planning or to technical limitations, they do not seek to ensure URI persistence beyond the domain level. For these reasons, Berners-Lee's original contention was that organizations often displayed a "lack of forethought" in their URI management (Berners-Lee, 1998).

The findings from the analysis of repository date of decease and the rate of repository decay suggest additive growth in European repository deaths is occurring. Based on this data, there are indications that dead repositories emerge more quickly than new repositories are established, at

[11] DRIS: https://eurocris.org/services/dris

least within Europe. As the region with the highest concentration of registered repositories, Europe will considerably erode the global total number of active repositories. The predictive potential of these observations must always be weighed against the prospect that new, more persistent repositories may be launched in the future. In addition, the possibility that, were new repositories to be launched, they may simply perpetuate repository death, given the repository mismanagement observed, must be considered. Moreover, the possibility that a proportion of any "new" European repositories may be zombie repositories, reanimated following prior repository death, must be considered. An exact understanding of the wider impact of this observation of decay is therefore difficult to establish. Reference data from ROAR on recorded repository registrations (globally) since 1991 indicates that the launching of new repositories peaked in 2012, with a steady decline thereafter. Zenodo[12] was launched in 2013 as a repository solution to capture the "long tail" of European research work (Amorim, Castro, da Silva, & Ribeiro, 2015). A consequence may have been to usurp the launching of new repositories in the years since 2012. Zenodo fulfils a wide variety of repository use cases that hitherto would have necessitated the creation of a dedicated repository, and its impact on scholarly communication more generally has been vast (Crespo Garrido, Loureiro García, & Gutlebe, 2025).

From data collection, the parsing of HTTP 301 responses associated with HTTP to HTTPS redirects to a cognate domain was treated as a 200 response. This process surfaced frequent inappropriate use of the HTTP 302 response by repositories when a 301 response was expected. HTTP 302 indicates that a requested resource has moved temporarily but will return to its original location later (Fielding, Nottingham, & Reschke, 2022). Therefore, software agents, such as a search engine bot, will typically retain an original resource location and elect not to update their indexes with a location they have been informed is temporary (Mozilla Foundation, 2025). Of note, in this dataset, 580 repositories returned a 302 response instead of the expected 301 response for what appeared to be a permanent HTTP to HTTPS redirect. Usually, this was an HTTP to HTTPS redirect at the root domain, although some appeared as permanent redirects to new, unrelated locations. This observation suggests that many repositories are mistakenly configured to serve a 302 response when they should be serving a 301. Such local misconfiguration of repository infrastructure can negatively affect search engine discovery and is an issue for the repository community to address.

Understanding the extent of European repository impersistence was the first step to measuring the impact it might be having on the scholarly record. It was possible to surface almost 20,000 references to dead repository content from the mining of the scholarly literature. That these dead references arise from approximately 12,000 unique scholarly works reveals the proclivity some authors display in citing multiple repository resources in a single scholarly work. Unfortunately, it also reveals a level of citation confidence in repositories that should exist but does not. Instead, organizational mismanagement of repository infrastructure has undermined the scholarly record and contributed to the reference rot crisis. Despite its size and scope, CORE is not an exhaustive corpus of scholarly literature, nor are all scholarly resources available for TDM interrogation within CORE. Based on this, it could be assumed that the number of references or citations to dead repository resources in the literature is probably higher than reported in this article. Similarly, there is probably a higher proportion of "dead on arrival" references in the literature too.

A notable finding from this article is that the aftereffects of repository death can be found in the scholarly record years after a repository is known to have died. So-called "dead on arrival" references were therefore an interesting but concerning observation given the volume of such references detected (n = 2,801).

One possible explanation for this might be the increased use of reference management software and the growth of personal information management among scholars. Scholars producing systematic reviews have relied on reference management software for some time (Lorenzetti & Ghali, 2013), and key review guides in this space, such as the *Cochrane Handbook for Systematic Review of Interventions*, even propose their own software solutions for this (Higgins et al., 2024). Software solutions not only support the capture and management of metadata describing found resources but also the caching, saving, and annotation of associated

[12] Zenodo: https://zenodo.org/

full-text PDF documents, such as Zotero, Mendeley, and EndNote (Speare, 2018; Williams & Woods, 2024). More recent studies indicate that the ability to "save and organise PDF files" within solutions is an essential feature for many users (Nitsos, Malliari, & Chamouroudi, 2022). It is therefore conceivable that an overreliance on such solutions by authors has emerged in which authors continue to reuse dead repository content within the writing process without realising it is dead. This is because authors may only be referring to what has been previously captured or annotated locally or in the cloud, rather than verifying the reachability of the resource they are using during writing and citing. Some solutions support features that can mitigate link rot. For example, Zotero's support for application plugins has enabled the creation of "Zotero Memento" (leonkt, 2021), which observes the Memento Protocol (Jones, Klein, Van de Sompel, Nelson, & Weigle, 2021). However, a minority of scholars are likely to use these features because usage is predicated on understanding the issue of reference rot in the first place. The issue is that scholars generally expect repositories to be persistent and that the scholarly resources they have located in the past will remain available. This is, after all, the way in which repositories have been advocated to scholarly users since the beginning (Lynch, 2003). The death or zombification of a repository, therefore, further exacerbates the impact of repository impersistence on open research and scholarship more generally.

It could be suggested that the emergence of certification initiatives, such as CoreTrustSeal (CoreTrustSeal Standards and Certification Board, 2022), as well as projects designed to stimulate maturity in FAIR-enabling repositories (van Lieshout et al., 2025), is tacit sectoral acknowledgement of a problem that has hitherto been unquantifiable. Without adherence to trustworthy repository frameworks, research-performing organizations may always fail to acknowledge the long-term commitment arising from repository management. They may also lack cognisance of possible consequences to scholarship, users, or open scholarly infrastructure. Efforts by the European Open Science Cloud (EOSC) (Burgelman, 2021) to grow a resilient, open, and "trusted" environment in which scholars can easily publish, locate and reuse scholarly content have identified a need to improve FAIRness and trustworthiness in repository infrastructure. There is recognition that TDRs are central to the realisation of the EOSC (Directorate-General for Research and Innovation (European Commission), 2018). Projects emerging from the EOSC, such as FAIR-IMPACT (Dillo et al., 2024), concluded that greater transparency in repository processes was necessary to increase their overall trustworthiness (Grootveld et al., 2025). Subsequent similar work under the auspices of the FIDELIS project has proposed resources, including the Transparent Trustworthy Repository Attributes Matrix, which is designed to serve as a reference model and can assist in assessing the extent to which repository operations around digital object management, organizational infrastructure, and technology contribute to scholarly transparency and trustworthiness (L'Hours et al., 2025).

Limitations and future research

This article has limitations, some of which may motivate further research. In the methods, not all scholarly works in CORE have full text available for TDM interrogation, nor does CORE provide an inventory of all known literature. This means that an authoritative, universal picture of total dead repository content diffusion in the scholarly literature will never be possible. However, CORE remains the largest such open dataset available and a useful indicator. Future work could explore similar analyses while enlarging the TDM dataset to include other sources. An obvious expansion of this article would be to expand analyses beyond Europe to include all available repositories. Quantifying the global challenge of repository impersistence to scholarship would be possible, as would the observation of geographic or continental differences.

In relation to "dead on arrival" references, the limitation that CORE's aggregation of scholarly literature can span many different work types should be acknowledged (e.g., accepted manuscript, preprint, or version of record). Therefore, it is conceivable that a subset of the identified "dead on arrival" references in a preprint or accepted manuscript was later identified and corrected during editorial steps, typesetting, or both by a publisher within a Version of Record. This is difficult to detect and, ergo, quantify without significant additional analysis. Therefore, it is a suggestion for future research.

A final limitation of note concerns the HTTP GET approach used to capture repository response codes. It should be acknowledged that while it was advantageous in this article to use a distributed cloud infrastructure, it is conceivable that requests from specific cloud IP ranges may have been treated differently to fixed client requests. The influence of such variability was minimal in this instance because services exhibiting erratic availability, errors, or both were retested and spot-checked.

The findings of this article could be used to define recommendations, guidance, or both that can support organisations and the repositories they host to develop and implement policies and practices that ensure greater sustainability and limit the number of dead repositories emerging in the future.

Finally, a productive area for future research would be to better understand how dead repository content has been cited or referenced in the scholarly literature. Approximately 20,000 references to dead repository content have entered the scholarly record as part of this article. However, it has long been known that not all references or citations are equal (Moravcsik & Murugesan, 1975). This is because scholarly works are cited by authors in a variety of ways and for a variety of different purposes; for example, is the reference "organic" or "perfunctory," or "evolutionary" or "juxtapositional"? Even the location of a citation in a scholarly work can be a predictor of its academic utility (Cano, 1989). This means the negative scholarly impact of dead repository content will vary depending on how the author has referenced or cited it. A more granular quantification of impact is therefore possible by analysing, categorising, and measuring how dead repository content has been cited in the works themselves. Although computationally intensive to perform, the use of research methods harnessing sentiment analysis may help reveal how dead repository content has been cited, as well as provide a more nuanced understanding of its relative negative impact.

Conclusion

Open repositories have become essential components of open scholarly infrastructure in recent decades. They provide infrastructure to support the storage, discovery, and impact of research. Therefore, the persistence of this infrastructure is critical to scholarship, supporting key aspects of the open research agenda. Instead of constituting reliable nodes in open scholarly infrastructure, this article has exposed concerning weaknesses in the persistence of European repository infrastructure, with significant levels of repository impersistence detected. These examples of impersistence have compromised the scholarly record in thousands of cases (including in "dead on arrival" references), hindering content discovery, research citation, verification, and reproducibility. In other words, infrastructure that has been designed and managed by organizations to better support the goals of open scholarship is, in many cases, damaging it.

From this article, many organizations operating repositories experience challenges in managing them, including misconfigured HTTP redirects, poor URI management, and technical naivety in the strategic management of repositories, resulting in the death or zombification of otherwise healthy repositories. The causes for this predicament are complex, as described earlier. As the findings of this study highlight, the consequences of repository mismanagement to scholarship are significant. These consequences need to be better acknowledged and understood by managing organizations. This is particularly relevant to university research libraries, which, as noted in this article, have hosted a disproportionate number of found-dead repositories and generally purport to be leaders in open research policy and practice.

Data Availability Statement

All data and software code underpinning the research documented in this article are available at: https://doi.org/10.5525/gla.researchdata.2101

Acknowledgements

We wish to acknowledge insightful online discussions about repository persistence with Paul Walk (Antleaf), Michael P. Taylor (Index Data and University of Bristol), and Herbert Van de Sompel (DANS and University of Ghent).

Joy Davidson acknowledges funding from the European Union's Horizon Europe Framework Programme for Research and Innovation, under grant agreement 101188078. FIDELIS: Establishing a European Network of Trustworthy Digital Repositories (https://doi.org/10.3030/101188078).

References

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology*, *62*(1), 40–49. https://doi.org/10.1002/asi.21448

Amorim, R. C., Castro, J. A., da Silva, J. R., & Ribeiro, C. (2015). A comparative study of platforms for research data management: Interoperability, metadata capabilities and integration potential. In A. Rocha, A. M. Correia, S. Costanzo, & L. P. Reis (Eds.), *New Contributions in Information Systems and Technologies*: Vol. *353*. Springer, Cham. (pp. 101–111). https://doi.org/10.1007/978-3-319-16486-1_10

Baglioni, M., Pavone, G., Mannocci, A., & Manghi, P. (2025). Towards the interoperability of scholarly repository registries. *International Journal on Digital Libraries*, *26*(1):2. https://doi.org/10.1007/s00799-025-00414-y

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Bamgbose, A. A., Mohd, M., Tengku Wook, T. S. M., & Mohamed, H. (2025). Are digital repositories for university libraries trustworthy? A systematic review. *Business Information Review*, *42*(2). https://doi.org/10.1177/0266382125132830

Berners-Lee, T. (1998). *Cool URIs don't change*. W3C. https://www.w3.org/Provider/Style/URI

Biesenbender, S., Petersohn, S., & Thiedig, C. (2019). Using Current Research Information Systems (CRIS) to showcase national and institutional research (potential): Research information systems in the context of Open Science. *Procedia Computer Science*, *146*, 142–155. https://doi.org/10.1016/j.procs.2019.01.089

Björk, B.-C. (2014). Open access subject repositories: An overview. *Journal of the Association for Information Science and Technology*, *65*(4), 698–706. https://doi.org/10.1002/asi.23021

Brembs, B. (2018). Prestigious science journals struggle to reach even average reliability. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00037

Bryant, R., Clements, A., de Castro, P., Cantrell, J., Dortmund, A., Fransen, J., Gallagher, P. & Mennielli, M. (2018). *Practices and patterns in research information management: Findings from a global survey*. OCLC Research. https://doi.org/10.25333/BGFG-D241

Burgelman, J.-C. (2021). Politics and open science: How the European Open Science Cloud became reality (the untold story). *Data Intelligence*, *3*(1), 5–19. https://doi.org/10.1162/dint_a_00069

Burns, C. S., Lana, A., & Budd, J. M. (2013). Institutional repositories: Exploration of costs and value. *D-Lib Magazine*, *19*(1/2). https://doi.org/10.1045/january2013-burns

Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, *40*(4), 284–290. https://doi.org/10.1002/(SICI)1097-4571(198907)40:4%3C284::AID-ASI10%3E3.0.CO;2-Z

Carnevale, R. J., & Aronsky, D. (2007). The life and death of URLs in five biomedical informatics journals. *International Journal of Medical Informatics*, *76*(4), 269–273. https://doi.org/10.1016/j.ijmedinf.2005.12.001

Confederation of Open Access Repositories. (2022). *COAR community framework for good practices in repositories, version 2*. Zenodo. https://doi.org/10.5281/ZENODO.7108101

Cope, J., & Baker, J. (2017). Library carpentry: Software skills training for library professionals. *International Journal of Digital Curation*, *12*(2), 266–273. https://doi.org/10.2218/ijdc.v12i2.576

CoreTrustSeal Standards and Certification Board. (2022). *CoreTrustSeal requirements 2023–2025*. Zenodo. https://doi.org/10.5281/zenodo.7051012

Crespo Garrido, I. del R., Loureiro García, M., & Gutleber, J. (2025). The value of an open scientific data and documentation platform in a global project: The case of Zenodo. In J. Gutleber, & P. Charitos (Eds.), *The economics of big science 2.0: Essays by leading scientists and policymakers.* Springer, Cham. (pp. 181–200). https://doi.org/10.1007/978-3-031-60931-2_14

Crow, R. (2002, August). *The case for institutional repositories: A SPARC position paper.* SPARC. https://web.archive.org/web/20190307175704id_/http://pdfs.semanticscholar.org/da92/25f01aced7586efb7ea8f82406371374b06f.pdf

da Silva, J. A. T. (2021). Reference rot degrades information preservation and induces the loss of intellectual integrity. *Epistēmēs Metron Logos, 6*, 1–6. https://doi.org/10.12681/eml.25792

de Castro, P., Shearer, K., & Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. *Procedia Computer Science*, *33*, 39–46. https://doi.org/10.1016/j.procs.2014.06.007

Dillo, I., Ulrich, R., Huber, R., L'Hours, H., Davidson, J., Neidiger, C., Parkes, O., Everhardt, M., Reijnhoudt, L., Meijas, G., Ramezani, P., Verburg, M.& Priddy, M. (2024). *Exposing repository information to foster connections and trust: evaluating and implementing guidelines (Version 1)* [Conference poster]. Zenodo. International Conference on Digital Preservation (iPRES2024), Ghent & Flanders, Belgium. https://doi.org/10.5281/ZENODO.13495834

Dimitrova, D., & Bugeja, M. (2007). Raising the dead: Recovery of decayed on-line citations. *American Communication Journal*, *9*(2). http://ac-journal.org/journal/2007/Summer/2RaisingtheDead.pdf

Directorate-General for Research and Innovation (European Commission). (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*. Publications Office of the European Union. https://doi.org/10.2777/1524

Dong, D., & Tay, C. H. A. (2023). Discoverability and search engine visibility of repository platforms. In L. Woolcott & A. Shiri (Eds.), *Discoverability in Digital Repositories* (pp. 157–175). Routledge. https://doi.org/10.4324/9781003216438-11

Ducut, E., Liu, F., & Fontelo, P. (2008). An update on Uniform Resource Locator (URL) decay in MEDLINE abstracts and measures for its mitigation. *BMC Medical Informatics and Decision Making*, *8(0)*:23. https://doi.org/10.1186/1472-6947-8-23

European Commission: Directorate-General for Research and Innovation. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*. Publications Office. https://doi.org/10.2777/1524

European Commission: Directorate-General for Research and Innovation, DCC, DANS, EFIS & Visionary Analytics. (2022). *European research data landscape: Final report*. Publications Office. https://doi.org/10.2777/3648

Eysenbach, G., & Trudel, M. (2005). Going, going, still there: Using the WebCite service to permanently archive cited web pages. *Journal of Medical Internet Research*, *7*(5):e920. https://doi.org/10.2196/jmir.7.5.e60

Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Quantitative Science Studies*, *2*(1), 42–64. https://doi.org/10.1162/qss_a_00111

Fernandes, S., & Pinto, M. J. (2018). From the Institutional Repository to a CRIS system: what challenges? *Qualitative and Quantitative Methods in Libraries*, *7*(3), 431–487. https://www.qqml-journal.net/index.php/qqml/article/view/494

Fielding, R. T., Nottingham, M., & Reschke, J. (2022). *HTTP Semantics [RFC 9110] (Request for Comments No. RFC 9110)*. Internet Engineering Task Force. https://doi.org/10.17487/RFC9110

Francke, H., Gamalielsson, J., & Lundell, B. (2017). Institutional repositories as infrastructures for long-term preservation. *Information Research*, *22*(2). Retrieved from https://informationr.net/ir/22-2/paper757.html

Garfield, E. (1996). When to cite. *The Library Quarterly*, *66*(4), 449–458. https://doi.org/10.1086/602912

Gertler, A. L., & Bullock, J. G. (2017). Reference rot: An emerging threat to transparency in political science. *Political Science and Politics*, *50*(1), 166–171. https://doi.org/10.1017/S1049096516002353

Grootveld, M., Fink, A. S., Jonquet, C., González Guardia, E., Dillo, I., Nordling, J., Davidson, J., Marjamaa-Mankinen, L., Verurg, M., Priddy, M., Grau, N. & Pittonet Gaiarin, S. (2025). *D1.3 recommendations for a FAIR EOSC [White paper] FAIR-IMPACT Synchronisation Force (Version 1.0).* Zenodo. https://doi.org/10.5281/zenodo.14979704

Helgesson, G., & Bülow, W. (2023). Research integrity and hidden value conflicts. *Journal of Academic Ethics*, *21*(1), 113–123. https://doi.org/10.1007/s10805-021-09442-0

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (Eds.). (2024). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*. Cochrane. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current

Hockx-Yu, H. (2006). Digital preservation in the context of institutional repositories. *Program*, *40*(3), 232–243. https://doi.org/10.1108/00330330610681312

Jefferies, N., Freeman, A., Notay, B., Knoth, P., Pentz, E., Buys, M., & Stewart, S. (2022, March 7-18). *Open scholarship infrastructures and sustainability.* [Conference presentation]. Oxford Festival of Open Scholarship 2022. Oxford, UK. https://ora.ox.ac.uk/objects/uuid:7108677d-3092-47ff-b749-3479e4be0020

International Organization for Standardization. (2020). *Codes for the representation of names of countries and their subdivisions* (ISO Standard No. 3166-1:2020). https://www.iso.org/standard/72482.html

Jones, S. M., Klein, M., & Van de Sompel, H. (2021). Robustifying links to combat reference rot. *Code4Lib Journal*, *50*. https://journal.code4lib.org/articles/15509

Jones, S. M., Klein, M., Van de Sompel, H., Nelson, M. L., & Weigle, M. C. (2021). Interoperability for accessing versions of web resources with the Memento Protocol. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 101–126). Springer, Cham. https://doi.org/10.1007/978-3-030-63291-5_9

Jones, S. M., Van de Sompel, H., Shankar, H., Klein, M., Tobin, R., & Grover, C. (2016). Scholarly context adrift: Three out of four URI references lead to changed content. *PLOS one, 11*(12):e0167475. https://doi.org/10.1371/journal.pone.0167475

Jouneau, T., Verburg, M., Horton, L., van Geest, G., Tujunen, S., Kallio, J., Paulsen, T., Duvaud, S., Recker, J., Holthe-Tveit, Å. J., Thorpe, D. E., Conzett, P., Kleemola, M. Kalaitzi, V., Huber, R., Jonquet, C., Aguilar, F., Forshaug, A. K., Alaterå, T. J. & Esteves, E. (2025). *FIDELIS landscape survey analysis (Version 1.0).* Zenodo. https://doi.org/10.5281/zenodo.15744996

Kelly, E. J. (2023). Discoverability beyond the library. In L. Woolcott & A. Shiri (Eds.), *Discoverability in Digital Repositories* (pp. 137–156). Routledge. https://doi.org/10.4324/9781003216438-10

Klein, M., & Balakireva, L. (2022). An extended analysis of the persistence of persistent identifiers of the scholarly web. *International Journal on Digital Libraries*, *23*, 5–17. https://doi.org/10.1007/s00799-021-00315-w

Klein, M., Balakireva, L., & Shankar, H. (2019). *Evaluating memento service optimizations.* [Conference paper]. 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Champaign, IL, United States. https://doi.org/10.1109/JCDL.2019.00034

Klein, M., Shankar, H., & Van de Sompel, H. (2018). Robust Links in Scholarly Communication. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. Association for Computing Machinery. 357–358. https://doi.org/10.1145/3197026.3203885

Klein, M., Shankar, H., Balakireva, L., & Van de Sompel, H. (2019). The Memento Tracer Framework: Balancing quality and scalability for web archiving. In A. Doucet, A. Issac, K. Golub, T. Aalberg, A. Jatowt. (Eds.) *Digital Libraries for Open Knowledge. TPDL 2019. Lecture Notes in Computer Science, (11799).* Springer, Champ. (pp. 163–176). https://doi.org/10.1007/978-3-030-30760-8_15

Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly context not found: One in five articles suffers from reference rot. *PLOS one, 9*(12), e115253. https://doi.org/10.1371/journal.pone.0115253

Knoth, P. (2025, July 3). Managing machine access to open repositories in the age of generative AI. *The 20th International Conference on Open Repositories*, [Conference Presentation] Chicago, IL, United States. https://doi.org/10.5281/ZENODO.15794663

Knoth, P., Herrmannova, D., Cancellieri, M., Anastasiou, L., Pontika, N., Pearce, S., Gyawali, B. & Pride, D. (2023). CORE: A global aggregation service for open access papers. *Scientific Data*, *10*(1):366. https://doi.org/10.1038/s41597-023-02208-w

L'Hours, H., Kleemola, M., Parkes, O., Recker, J., Duvaud, S., van Horik, R., Alaterå, T. J., Liberante, F., Conzett, P., Kaartinen, H., Båckman, S. & Esteves, E. (2025). *FIDELIS TTRAMatrix v01.00 Introduction and Overview*. Zenodo. https://doi.org/10.5281/zenodo.17144159

leonkt. (2021). *Leonkt/zotero-memento (Version 1.1.1)* [JavaScript]. Retrieved from https://github.com/leonkt/zotero-memento (Original work published 2019)

Lorenzetti, D. L., & Ghali, W. A. (2013). Reference management software for systematic reviews and meta-analyses: An exploration of usage and usability. *BMC Medical Research Methodology*, *13*(1), 141. https://doi.org/10.1186/1471-2288-13-141

Lynch, C. A. (2002). Preserving digital information to support scholarship. In M. Devlin, R. C. Larson & J. W. Meyerson (Eds.), *EDUCAUSE: The Internet and the university: forum 2002.* Boulder. https://www.cni.org/wp-content/uploads/2002/08/preserving-digital-info.pdf

Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Portal: Libraries and the Academy*, *3*(2), 327–336. https://doi.org/10.1353/pla.2003.0039

Mabe, A., Nelson, M. L., & Weigle, M. C. (2021). *Extending Chromium: Memento-aware browser.* [Conference presentation]. 2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Champaign, IL, USA, 2021 (pp. 310-311). https://doi.org/10.1109/JCDL52503.2021.00046

Mabe, A., Nelson, M. L., & Weigle, M. C. (2022). Extending chromium: based Memento-Aware web browser. In G. Silvello, O. Corcho, P. Manghi, G. M. Di Nunzio, K. Golub, N. Ferro, & A. Poggi (Eds.), *Linking Theory and Practice of Digital Libraries* (pp. 147–160). Springer, Cham. https://doi.org/10.1007/978-3-031-16802-4_12

Macgregor, G. (2023). Digital repositories and discoverability: Definitions and typology. In L. Woolcott & A. Shiri (Eds.), *Discoverability in Digital Repositories* (pp. 11–31). Routledge. https://doi.org/10.4324/9781003216438-3

Macgregor, G. (2025, April 2). *Revisiting the importance of persistence in the scholarly web.* [Conference presentation]. THE Digital Universities UK Conference, Lancaster, UK. https://eprints.gla.ac.uk/352395/

Macgregor G., Lancho-Barrantes, B. S., & Pennington, D. R. (2023). Measuring the concept of PID literacy: User perceptions and understanding of PIDs in support of open scholarly infrastructure. Open Information Science, 7(1). https://doi.org/10.1515/opis-2022-0142

Mahanama, B., Balakireva, L., Jayarathna, S., Nelson, M., & Klein, M. (2022). Memento validator: A toolset for memento compliance testing. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries.* Association for Computing Machinery (pp. 1–3). https://doi.org/10.1145/3529372.3533297

Martin-Segura, S., Lopez-Pellicer, F. J., Nogueras-Iso, J., Lacasta, J., & Zarazaga-Soria, F. J. (2022). The problem of reference rot in spatial metadata catalogues. *ISPRS International Journal of Geo-Information*, *11*(1):27. https://doi.org/10.3390/ijgi11010027

Massicotte, M., & Botter, K. (2017). Reference rot in the repository: A case study of electronic theses and dissertations (ETDs) in an academic library. *Information Technology and Libraries*, *36*(1), 11–28. https://doi.org/10.6017/ital.v36i1.9598

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, *5*(1), 86–92. https://doi.org/10.1177/030631277500500106

Mozilla Foundation. (2025, September 6). *302 Found—HTTP.* https://web.archive.org/web/20250906084515/https://developer.mozilla.org/en-US/docs/Web/HTTP/Reference/Status/302

Nazim, M., & Bhardwaj, R. K. (2023). Open access initiatives in European countries: analysis of trends and policies. *Digital Library Perspectives*, *39*(3), 371–392. https://doi.org/10.1108/DLP-06-2022-0051

Nelson, M. L., & Allen, B. D. (2002). Object persistence and availability in digital libraries. *D-Lib Magazine*, *8*(1). https://doi.org/10.1045/january2002-nelson

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609–641. https://doi.org/10.1002/aris.2007.1440410120

Nitsos, I., Malliari, A., & Chamouroudi, R. (2022). Use of reference management software among postgraduate students in Greece. *Journal of Librarianship and Information Science*, *54*(1), 95–107. https://doi.org/10.1177/0961000621996413

Phelps, T. A., & Wilensky, R. (2000). Robust hyperlinks and locations. *D-Lib Magazine*, *6*(7/8). https://doi.org/10.1045/july2000-wilensky

Plantin, J-C., & Thomer, A. (2025). Platforms, programmability, and precarity: The platformization of research repositories in academic libraries. *New Media & Society, 27*(1), 338–358. https://doi.org/10.1177/14614448231176758

Recker, J., Kleemola, M., & L'Hours, H. (2024). Closing gaps: A model of cumulative curation and preservation levels for trustworthy digital repositories. *International Journal of Digital Curation*, *18*(1). https://doi.org/10.2218/ijdc.v18i1.926

Rieger, O. Y. (2012). Sustainability: Scholarly repository as an enterprise. *Bulletin of the American Society for Information Science and Technology*, *39*(1), 27–31. https://doi.org/10.1002/bult.2012.1720390110

Romero, R. C. (2025). Transparency in open science: An actionable principle? *Open Information Science*, *9*(1). https://doi.org/10.1515/opis-2025-0016

Rothfritz, L., Matthias, L., Pampel, H., & Wrzesinski, M. (2025). Current challenges and future directions for institutional repositories: A systematic literature review. *Journal of the Association for Information Science and Technology*, 77(1), 301-322. https://doi.org/10.1002/asi.70016

Rubin, G., Bardi, A., Del Gratta, R., & Del Grosso, A. M. (2026). *Enhancing interoperability of SPARQL endpoints: RESTful and OAI-PMH API for the DH-ATLAS project.* Consiglio Nazionale delle Ricerche, CNR. https://iris.cnr.it/handle/20.500.14243/564781

Russell, E., & Kane, J. (2008). The missing link: Assessing the reliability of internet citations in history journals. *Technology and Culture*, *49*(2), 420–429. https://doi.org/10.1353/tech.0.0028

Sauermann, L., & Cyganiak, R. (2008, December 3). *Cool URIs for the Semantic Web.* W3C. https://www.w3.org/TR/cooluris/

Schöpfel, J., & Azeroual, O. (2021). Current research information systems and institutional repositories: From data ingestion to convergence and merger. In D. Baker & L. Ellis (Eds.), *Future Directions in Digital Information.* Chandos Publishing. (pp. 19–37). https://doi.org/10.1016/B978-0-12-822144-0.00002-1

Schöpfel, J., Azeroual, O., Chaudiron, S., Jacquemin, B., Kergosien, E., Prost, H., & Thiault, F. (2024). From open repositories to CRIS - A case study. *Procedia Computer Science*, *249*, 173–178. https://doi.org/10.1016/j.procs.2024.11.061

SCONUL. (2025). *The future of the systems librarian.* SCONUL. https://web.archive.org/web/20250429154425/https://www.sconul.ac.uk/SCONULDownloadController/Download?iType=1&iID=2420&cGUID=29251c00-74a5-4743-bc7a-51827748c4d3&cTempLocation=#expand.

Shearer, K., Nakano Koga, S. M., Rodrigues, E., Manola, N., Pronk, M. & Proudman, V. (2023). *Current state and future directions for open repositories in Europe*. Zenodo. https://doi.org/10.5281/zenodo.10255559

Shearer, K., Rodrigues, E., Manola, N., Malaguarnera, G., Pronk, M.& Proudman, V. (2025). *REPOSITORIES: Key infrastructure for Maintaining European research excellence.* Zenodo. https://doi.org/10.5281/zenodo.15119849

Sherrick, A. K. & Pino Navarro, D. A. (2024, June 28). *Creating a better balance: the need for tools and practices to combat AI harvests and resource flooding in repository environments* [Conference presentation]. Open Repositories 2024 (OR2024), Gothenburg, Sweden. https://doi.org/10.5281/zenodo.12579304

Shreeves, S. L., & Cragin, M. H. (2008). Institutional repositories: Current state and future. *Library Trends*, *57*(2), 89–97. https://doi.org/10.1353/lib.0.0037

Speare, M. (2018). Graduate student use and non-use of reference and PDF management software: An exploratory study. *The Journal of Academic Librarianship*, *44*(6), 762–774. https://doi.org/10.1016/j.acalib.2018.09.019

Spinellis, D. (2003). The decay and failures of web references. *Communications of the ACM*, *46*(1), 71–77. https://doi.org/10.1145/602421.602422

Strecker, D., Pampel, H., Schabinger, R., & Weisweiler, N. L. (2023). Disappearing repositories: Taking an infrastructure perspective on the long-term availability of research data. *Quantitative Science Studies*, *4*(4), 839–856. https://doi.org/10.1162/qss_a_00277

Suber, P. (2012). *Open Access*. The MIT Press. https://library.oapen.org/handle/20.500.12657/26065

Tait, E., Martzoukou, K., & Reid, P. (2016). Libraries for the future: The role of IT utilities in the transformation of academic libraries. *Palgrave Communications*, *2*(1):16070. https://doi.org/10.1057/palcomms.2016.70

Thalhammer, A. (2024). *Cool URIs for FAIR knowledge graphs*. arXiv. https://doi.org/10.48550/arXiv.2407.09237

Thompson, J., & Hoover, S. (2023). Discoverability within the library. In L. Woolcott & A. Shiri (Eds.), *Discoverability in Digital Repositories* (pp. 117–136). Routledge. https://doi.org/10.4324/9781003216438-9

Van de Sompel, H., Klein, M., & Jones, S. M. (2016, April 11-15). *Persistent URIs must be used to be persistent* [Conference poster]. Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, QC, Canada. https://doi.org/10.1145/2872518.2889352

Van de Sompel, H., Nelson, M. L., Sanderson, R., Balakireva, L. L., Ainsworth, S., & Shankar, H. (2009). *Memento: Time travel for the web* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.0911.1112

Van de Sompel, H., Nelson, M., & Sanderson, R. (2013). *HTTP Framework for Time-Based Access to Resource States – Memento (Request for Comments No. RFC 7089)*. Internet Engineering Task Force. https://doi.org/10.17487/RFC7089

van Lieshout, N., Ramezani, S., van Horik, R., Horton, L., Turner, D., Davidson, J., Marjamaa-Mankinen, L., Lager, L. & Nordling, J. (2025). *MS3.8 technical EOSC PID implementation guide & program (Milestone Report No. 3.8)*. Zenodo. https://doi.org/10.5281/zenodo.14779609

Walk, P. (2023, June 13). *Analysis of OpenDOAR data* [Conference presentation]. Open Repositories 2023 (OR2023), Stellenbosch, South Africa. https://doi.org/10.5281/zenodo.8091491

Williams, L., & Woods, L. (2024). Reference management practices of students, researchers, and academic staff. *The Journal of Academic Librarianship*, *50*(3):102879. https://doi.org/10.1016/j.acalib.2024.102879

Yakel, E., Faniel, I. M., & Robert Jr, L. P. (2024). An empirical examination of data reuser trust in a digital repository. *Journal of the Association for Information Science and Technology*, *75*(8), 898–915. https://doi.org/10.1002/asi.24933

Yakel, E., Faniel, I. M., Kriesberg, A., & Yoon, A. (2013). Trust in digital repositories. *International Journal of Digital Curation*, *8*(1), 143–156. https://doi.org/10.2218/ijdc.v8i1.251

Zhou, K., Grover, C., Klein, M., & Tobin, R. (2015). No more 404s: Predicting referenced link rot in scholarly articles for pro-active archiving. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 233–236). Association for Computing Machinery. https://doi.org/10.1145/2756406.2756940