Life and Death of DH Projects: A Preliminary Investigation of Their Lifecycles in Italy

Erica Andreose[1] , Giorgia Crosilla[1],
Remo Grillo[2,3] , Gianmarco Spinaci[2,3]

[1] Digital Humanities and Digital Knowledge, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy, erica.andreose@studio.unibo.it, giorgia.crosilla@studio.unibo.it

[2] Patrimonio Culturale nell'Ecosistema Digitale, Università di Bologna, Bologna, Italy

[3] I Tatti - The Harvard University Center for Italian Renaissance Studies, Firenze, Italy, rgrillo@itatti.harvard.edu, gspinaci@itatti.harvard.edu

ABSTRACT (ITALIANO)

La significativa proliferazione dei progetti di Digital Humanities (DH) a cui si è assistito negli ultimi anni ha comportato altrettante sfide relative alla conservazione digitale. Infatti, una parte dei progetti risulta ad oggi non più accessibile a causa di problemi quali l'obsolescenza tecnologica e il mancato mantenimento dei server.

Questo studio rappresenta un primo tentativo di disamina del ciclo di vita e della persistenza in rete dei progetti DH in ambito italiano, considerando 270 progetti realizzati dal 1998 al 2024, provenienti da fonti diverse quali atti di conferenze, piattaforme online di centri di ricerca e aggregatori di iniziative del settore. Dall'analisi è emerso che su 250 progetti con URL identificato, 213 sono raggiungibili, 31 sono accessibili parzialmente e 37 presentano problemi di risoluzione degli indirizzi web (con sito non funzionante o dove il reindirizzamento non è corretto). Inoltre, esaminando i progetti non disponibili, lo studio ha rilevato una loro durata in vita mediamente di 3 anni e 9 mesi. I dati evidenziano una crescita costante del numero di nuovi progetti a partire dagli anni 2000, con un incremento significativo nell'ultimo decennio, seguito da un calo dal 2022 in poi. Ciò che emerge sono delle fragilità legate alle infrastrutture digitali nel settore delle DH, che evidenziano la necessità di adottare strategie strutturate a garantirne la sostenibilità e l'accessibilità nel tempo. L'obiettivo finale del lavoro è sensibilizzare sui rischi sistemici connessi alla conservazione sul lungo termine dei progetti di ricerca del settore, al fine di garantire una maggiore persistenza e tutela dei contributi legati al patrimonio culturale e accademico.

Parole chiave: progetti DH; obsolescenza tecnologica; preservazione digitale;

ABSTRACT (ENGLISH)

In recent years, there has been a significant proliferation of Digital Humanities (DH) projects, but this growth has been accompanied by increasing challenges in digital preservation. In fact, many of these projects are not currently accessible due to issues such as technological obsolescence and server maintenance difficulties.

This study represents the first attempt to examine the online lifecycle and persistence of DH projects in Italy through an analysis of 270 initiatives from 1998 to 2024. The chosen research methodology combines data from multiple sources, including conference proceedings, research center websites, and existing project aggregators, to track projects' availability and longevity.

The analysis revealed that 250 projects had associated websites, with 213 currently functioning, 31 partially inaccessible, and 37 experiencing URL-related issues (with a non functioning website or having an incorrect redirection). Examining discontinued projects, the study found an average lifespan of 3 years and 9 months. Additionally, the data highlights a steady increase in new project launches since the early 2000s, with notable growth during the last decade, followed by a decline in net growth from 2022 onwards.

Our findings unveil the fragility of digital infrastructures in DH, highlighting the need for enhanced strategies to ensure the sustainability and accessibility of these initiatives over time. By addressing systemic risks and prioritizing preservation planning, the field can better safeguard its contributions to cultural and scholarly heritage.

Keywords: DH projects; technological obsolescence; digital preservation;

1. INTRODUCTION

The field of Digital Humanities (DH) is an interdisciplinary domain in constant evolution, with projects often tied to temporary funding and rapidly changing technologies. The lifespan of a DH project reflects not only technical aspects but also the sustainability of research practices. The loss of access to such projects entails the disappearance of data, tools, and knowledge, invalidating the time and resources invested and hindering progress in specific micro-disciplinary fields.

This research provides a preliminary analysis of the state of DH projects in Italy, focusing on their longevity and the factors influencing their lifecycle. Key questions include: What is the expected lifespan of a DH project? How many projects are “born” and ”die” over time?

It is worth noticing that this study focuses exclusively on DH projects whose output is publicly accessible online via a website or digital edition. As such, it does not account for projects that exist only locally or whose results are disseminated through non-digital means. This is a structural limitation, as it captures only a portion of the broader DH landscape, though likely the most visible and fragile over time. Using project URLs, we track their online activity to assess their status and establish their initial presence. This methodology facilitates systematic evaluation and lays the groundwork for scalable computational approaches applicable to international contexts.

For this study, a DH project’s "life" begins with its first verifiable online presence and ends when it becomes inaccessible or loses essential functionalities. To avoid the binary implications of "death", we describe projects no longer functional or accessible as "unavailable." We also identify intermediate states, including partially functional projects, irregularly maintained projects, and those no longer updated but still online.

Focusing initially on the Italian context has enabled us to conduct a controlled preliminary study, allowing for manual data verification and ensuring the accuracy of our findings. This approach has been instrumental in identifying methodological challenges and refining our framework. These insights will inform future expansions to encompass DH projects on a global scale, contributing to strategies for enhancing their long-term sustainability and accessibility.

2. RELATED WORKS

While extensive research exists on the disappearance of digital resources[1], our work focuses specifically on DH projects, an area where studies remain limited and primarily concern long-term sustainability methods. DH projects differ from general digital resources due to their purpose of preserving cultural and humanistic heritage, often acting as an alternative, complement, or enhancement to traditional physical preservation practices. If DH projects aim to achieve this goal of long-term preservation, their lifespan on the web should ideally be unlimited. However, this is often not the case, as many projects face challenges such as technological obsolescence, lack of active maintenance, and eventual inaccessibility.

Previous studies on the longevity and decline of DH projects have primarily focused on the North American context. For instance, VandeCreek (2022:99) highlights the importance of institutional support in mitigating project decline, noting that academic projects are 4.15 times more likely to sustain their digital resources than non-academic ones. Furthermore, it also reveals that 68% of early funded projects (1996-2003) remain accessible online.

The longevity of DH projects has also been analyzed by *Shelf Life: Identifying the Abandonment of Online Digital Humanities Projects* (Meneses & Furuta, 2019). Their study, based on an analysis of URLs from conference abstracts within the Alliance of Digital Humanities Organizations, concludes that the average lifespan of unmaintained online DH initiatives is approximately five years before disappearance. However, the methodology used in this work has its limitations: it includes URLs that are irrelevant to the analysis, such as those linked to external tools or unrelated projects, and excludes projects not presented at conferences. Furthermore, the exclusive reliance on HTTP responses provides only a partial understanding of a project's status, failing to capture complex issues such as functional failures of seemingly active websites.

Moreover, the inability of preserving digital humanities beyond a few years has been investigated further in the *Sustainability of digital humanities projects as a publication and documentation challenge* (Edmond & Morselli, 2020). They identify evolving needs, limited resources, inadequate sustainability planning, and

[1]Chapekis Athena, Bestvater Samuel, Remy Emma, and Rivero Gonzalo, *When online content disappears*, Pew Research Center Data Labs, May 17, 2024, https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears/. Accessed January 12, 2025.

"orphaning" as primary causes of project discontinuation[2]. Using the Bamboo Project[3] (Dombrowski, 2014) as a case study, they illustrate how late attempts at sustainability following the end of funding could not save the initiative. Key factors in the project's failure included insufficient funding, lack of planned interoperability, and a lack of strategic planning[4] and a cohesive vision among stakeholders. Broadening the discussion, technical neglect and the discontinuation of software, such as Adobe Flash Player, a phenomenon often referred to as "bit rot", pose significant threats to the sustainability of DH projects (VandeCreek, 2022: 93).
To our knowledge, this study represents the first effort to examine the sustainability and lifespan of DH projects in Italy, addressing this critical gap in the literature.

3. METHODOLOGY

To analyze the lifecycle of Italian DH projects, we defined geographical and thematic boundaries of what is considered part of this group (see the Introduction). We deemed these projects' URLs relevant for this analysis so that we can understand the projects' health state based on status codes and Internet Archive's Wayback Machine[5] snapshot dates of HTTP requests against their URLs. On the other hand, we imply that a project is its URL, and this is not always the case. Also, Internet Archive snapshots are estimates, not a comprehensively reliable point in time describing the health of a website. Currently, the literature lacks an aggregator of projects with the characteristics above. Hence, we adopted the methodology presented in (Spinaci et al., 2022) in an attempt to create such an aggregator in the form of a dataset, integrating external sources manually assessed.
With these premises, the experiment started by detecting the typologies of *sources* where information about projects could be found: aggregators, proceedings, and research center websites. For the aggregators, we scraped the DH Awards website[6], a website where the DH community can nominate resources, and an international committee vote for the nominated based on categories. Because of its main drawbacks (i.e., it lists international projects and tools alike), we drastically removed many entries that did not conform to our target projects. Other three project aggregators are the KNOT Catalogue[7], an archive containing digital scholarly resources among projects and objects, the dataset "Classification of web-based Digital Humanities projects" (Battisti, 2024) and the "Catalogue of Digital Editions" (Franzini, 2012) from where we extracted the ones relevant to the experiment[8]. For the proceedings, we manually assessed the AIUCD Conference Proceedings[9] dated 2013 and those ranging from 2016 to 2024, looking for articles describing or referencing projects, preferably with URLs. Another set of conference proceedings is IRCDL[10], the "Information and Research Science Connecting to Digital and Library Science", of which we manually assessed the period from 2005 to 2019. The last typology container considered is the research center websites. We started by collecting AIUCD-affiliated research centers[11] and integrating the list with the missing ones of our knowledge. We scraped each of their websites to get a list of their published projects. Although their coverage is far from "complete", these sources have been selected based on their availability and relevance to provide a comprehensive overview of DH projects in the Italian landscape.
During the manual assessment, each project was checked to agree to the conformity parameters defined in the previous sections. This step's output is a list of projects per source, each with its corresponding URL.
The second step is to merge the project lists. We decided to review the list's content before applying computational techniques to disambiguate and eliminate the duplicates. We noticed some projects repeated with different URLs, a behavior explainable by their physiological evolution over time. These cases were handled manually, keeping only the row with an existing website, i.e., with a resolving URL.

---

[2] Robin Camille, *The final death(s) of digital scholarship—An ongoing case study of DH2005 projects* [Conference presentation], Digital Afterlives Symposium, Bard Graduate Center, New York, NY, March 1, 2019. https://robincamille.com/presentations/death_of_digital_scholarship/. Accessed January 12, 2025.
[3] Dombrowski, Quinn. T*owards a Taxonomy of Failure*, January 30, 2019. https://quinndombrowski.com/blog/2019/01/30/towards-taxonomy-failure/. Accessed April 8, 2025.
[4] One of the attempts of formalization of good principles and practices to avoid early obsolescence and inaccessibility is *The Endings Project*, https://endings.uvic.ca/index.html. Accessed January 12, 2025.
[5] https://web.archive.org/. Accessed January 12, 2025.
[6] http://dhawards.org/. Accessed January 12, 2025.
[7] https://projects.dharc.unibo.it/knot/ . Accessed January 12, 2025.
[8] The ongoing project ATLAS: A Digital Cartography of DH Projects. GitHub Pages, 2024. https://dh-atlas.github.io , when completed, could be implemented as a source for enhancing the dataset.
[9] https://umanisticadigitale.unibo.it/pages/boa. Accessed January 12, 2025.
[10] https://dblp.org/db/conf/ircdl/index.html. Accessed January 12, 2025.
[11] https://www.aiucd.it/centri/. Accessed January 12, 2025.

After this preliminary review, we could automatically remove rows with duplicate URLs. The final result is a list containing 270 Italian Digital Humanities projects, of which 3.1% from Catalogue of Digital Editions, 4.4% come from the DH awards website, 12.2% from IRCDL proceedings, 15.9% from AIUCD proceedings, 24.8% from the KNOT Catalogue and 39.6% from Italian DH institutions.
The next step is to enrich the data with additional information. We encountered cases where a showcase website or a landing page existed (quite common when public grants fund projects) while, to the best of our knowledge, we could not find the project URL. This has been noted for 20 projects. For the remaining 250 projects, we controlled the status code of HTTP requests against their URL and analyzed their coverage over time by accessing the Wayback Machine APIs through a custom script. This script retrieves snapshots of each website, which, when filtered based on 2xx status codes, can provide an estimate of the period during which the website was or is available. Wayback Machines's crawlers do not give the exact date when a website appeared or disappeared from the web, but they offer a temporal guide. We can infer that a website was published before the first snapshot timestamp. Regarding the last snapshot, if a website is not resolved, we can infer that the website went offline after the last snapshot timestamp. In addition to the script, we manually inspected the Wayback Machine snapshots of unavailable projects to perform a qualitative check and determine whether the projects existed by analyzing the website's content. For each project, we evaluated whether it had a website, if it was accessible, whether its content was available, and if we encountered an outdated or inactive URL. In the latter case, we manually verified whether the project had been migrated to another domain and whether there was a redirect from the old one.

4. RESULTS

As of 11 April 2025 (the date cutoff that defines this analysis), this preliminary study identified 270 Digital Humanities projects (available online[12]) from Italy, spanning the years 1998–2024. Among these, 20 do not yet have a project website, although a platform had previously been announced. Of the remaining 250, 213 have a functioning website, while 31 are partially inaccessible (due to permission issues, obsolete technologies, or lack of maintenance). Additionally, 37 projects face URL-related issues: 26 do not resolve, and 11 lack a proper redirect after migrating to a new domain.

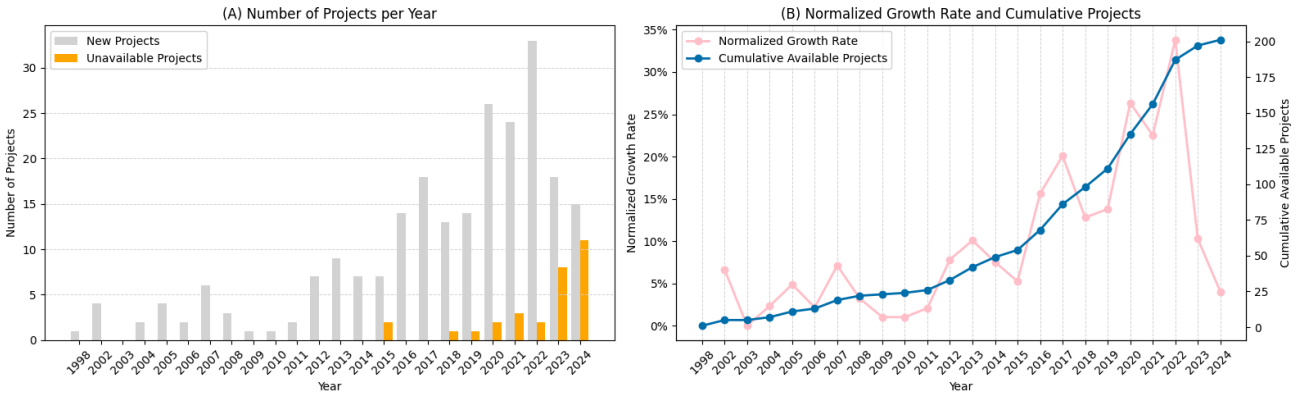

Figure 1. Growth Rate and Number of DH Projects in Italy per year

The resulting data outline an evolution of DH projects over time, as illustrated in (Fig. 1, A). The trend indicates a steady increase, with negligible fluctuations, in new online project launches per year starting from the early 2000s, with a significant surge in activity during the last decade. This increase is likely also correlated with the broader institutionalization of DH in Italy, including the establishment of new dedicated research centers and, therefore, funding opportunities (Sprugnoli et al., 2019). The surge in new projects also correlates with global trends in the DH field, as evidenced by Google Trends data reflecting increasing interest and activity in the domain nationally.[13]
The Growth Rate (Fig. 1, B), normalized by the number of cumulative projects, further contextualizes this trend. The chart reflects the annual difference between new and not-anymore-available projects,

[12] https://github.com/gspinaci/Vita-e-morte-DH-projects. Accessed April 11, 2025.
[13] https://trends.google.it/trends/explore?date=2004-01-01%202025-04-11&q=digital%20humanities&hl=it. Accessed April 11, 2025.

emphasizing the consistent growth of the DH landscape. Despite fluctuations, the positive net difference observed in most years indicates a general expansion of the field, especially during the period from 2016 to 2021, which saw the most substantial increase in the number of active projects in absolute terms. However, the declining growth from 2022 onwards might suggest an emerging gap between project launches and project discontinuations, possibly reflecting the challenges of maintaining long-term sustainability amidst growing complexity or resource constraints or that after more than 20 years, a critical mass of projects is being reached whose natural life cycle is coming to an end. It could also be due to natural project cycles or the conclusion of funding for upcoming releases tied to current national initiatives (e.g., PNRR/PRIN), potentially leading to a new increase in activity between 2025 and 2026. In either case, among discontinued projects, the average lifespan was calculated based on the period between the first and last snapshots recorded by the Wayback Machine. Preliminary findings suggest that, among the currently unavailable projects, their average lifespan was found to be relatively short, with the majority ceasing activity within 3 years and 9 months of their launch. However, this estimate should be interpreted with caution, as the current number of unavailable projects is relatively small.

While the increasing number of DH initiatives signals growth and innovation, the high rates of unavailability raise concerns about the durability of their contributions. The limited dataset and geographic focus constrain broader generalizations but provide a crucial starting point for more extensive investigations. Future expansions could integrate data from other contexts, facilitating cross-national comparisons and enabling comprehensive assessments of global DH sustainability.

5. CONCLUSIONS AND FUTURE DEVELOPMENTS

This study has provided an initial assessment of DH projects in Italy, identifying 270 initiatives and highlighting trends in their availability and sustainability. While the increasing number of projects reflects growing institutional and scholarly interest, the average lifespan of currently unavailable projects, from our list, — 3 years and 9 months — underscores challenges presumably tied to temporary funding cycles, technical obsolescence, or limited strategic planning, among others that might be worth investigating in a future study. One of the limitations of this project lies in the way data is aggregated and cleaned, as well as in the limited subset of DH projects analyzed. Due to the current lack of an aggregator that comprehensively tracks all past and ongoing DH projects, manual assessment was necessary. However, this approach is not a deterministic method for categorizing projects. Further standardized distinctions within DH project aggregators, such as the efforts undertaken by the KNOT Catalogue and ATLAS, could help eliminate the need for manual evaluation and reduce ambiguities during dataset cleaning.

To build on these findings, future research should move beyond aggregation and consider qualitative aspects of DH projects, such as their scale, funding sources, and institutional sponsorship. Understanding the internal histories of projects, including periods of downtime or partial unavailability, could provide valuable insights into the dynamics of their lifecycles. Additionally, developing a taxonomy of risk factors, such as reliance on outdated technologies or insufficient maintenance, could guide strategies to improve sustainability. While this research does not aim to offer a dynamic and continuously updated observatory, the dataset and the data collection workflow have been made openly available for replication and expansion. This "snapshot" approach provides a reproducible baseline for further studies, allowing for future analyses and tracking of trends over time. Scaling the methodology to automate parts of the project identification and verification process is also crucial for broader applications. While the manual process used in this study ensured accuracy through proper human investigation and implicitly assessing the principle of "Findability" as embodied in the FAIR principles (Tóth-Czifra & Edmond, 2020), it was time-intensive and error-prone, highlighting the need for more efficient approaches on a larger scale. Expanding the research internationally would enable cross-national comparisons and might uncover region-specific factors affecting DH project longevity. By addressing these areas, future studies can contribute to ensuring the sustainability and accessibility of DH projects, preserving their scholarly and cultural impact for the long term.

BIBLIOGRAPHY

Battisti, T. (2024). *Classification of web-based Digital Humanities projects leveraging information visualisation techniques* (1.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14192758

Dombrowski, Q. (2014). What Ever Happened to Project Bamboo? *Literary and Linguistic Computing*, *29*(3), 326–339. https://doi.org/10.1093/llc/fqu026

Edmond, J., & Morselli, F. (2020). Sustainability of digital humanities projects as a publication and documentation challenge. *Journal of Documentation*, *76*(5), 1019–1031. https://doi.org/10.1108/JD-12-2019-0232

Franzini, G. (2012). A Catalogue of Digital Editions. Zenodo. https://doi.org/10.5281/zenodo.1161425

Meneses, L., & Furuta, R. (2019). Shelf life: Identifying the abandonment of online digital humanities projects. *Digital Scholarship in the Humanities*, *34*(Supplement_1), i129–i134. https://doi.org/10.1093/llc/fqy079

Spinaci, G., Colavizza, G., & Peroni, S. (2022). A map of Digital Humanities research across bibliographic data sources. *Digital Scholarship in the Humanities*, *37*(4), 1254-1268.

Sprugnoli, R., Pardelli, G., Boschetti, F., & Del Gratta, R. (2019). Un'Analisi Multidimensionale della Ricerca Italiana nel Campo delle Digital Humanities e della Linguistica Computazionale. *Umanistica Digitale*, *3*(5). https://doi.org/10.6092/issn.2532-8816/8581

Tóth-Czifra, E., & Edmond, J. (2020). The risk of losing the thick description: data management challenges faced by the arts and humanities in the evolving fair data ecosystem. Digital Technology and the Practices of Humanities Research, 126.

VandeCreek, D. (2022). "Where Are They Now? The 2020 Status of Early (1996–2003) Online Digital Humanities Projects and an Analysis of Institutional Factors Correlated to Their Survival". *Preservation, Digital Technology & Culture*, *51*(3), 91–109. https://doi.org/10.1515/pdtc-2022-0011