Umanistica Digitale – ISSN: 2532-8816 – n. 19, 2025
C. Martignano – *A catalogue of software tools for digital scholarly editings*
DOI: http://doi.org/10.6092/issn.2532-8816/21093

A catalogue of software tools for digital scholarly editing

Chiara Martignano

University Ca' Foscari of Venice
chiara.martignano@unive.it

Abstract

In the field of digital philology, a large number of software tools have been, and continue to be developed over the years to assist (or replace) editors in the production of a Digital Scholarly Edition (DSE). The growing number of software tools for scholarly editing has created a need for a dedicated catalogue to help scholars and institutions select current tools that best suit their research purposes. A specialised catalogue would also serve as a valuable resource for studying technological trends and methodologies in digital philology. The creation of such a catalogue requires careful consideration of both practical and theoretical aspects. This paper evaluates existing tool discovery resources, examines the potential scholarly value of the catalogue, and addresses key implementation challenges. It then presents a prototype and explores possibilities for its future development.

Keywords: AIUCD2023, catalogue, digital scholarly editing, software tools, digital editing software

Nel corso degli anni, e in misura sempre maggiore negli ultimi tempi, sono stati sviluppati numerosi strumenti software per assistere (o sostituire) i curatori nella realizzazione di un'edizione scientifica digitale. Il crescente numero di prodotti software per la filologia digitale ha reso necessario un catalogo dedicato che possa aiutare studiosi e istituzioni nella selezione degli strumenti disponibili più adatti ai propri obiettivi di ricerca. Inoltre, tale catalogo rappresenterebbe una risorsa preziosa per studiare l'evoluzione tecnologica e metodologica della filologia digitale. La realizzazione di questo catalogo richiede un'attenta valutazione di aspetti pratici e teorici. Dopo aver esaminato le risorse attualmente disponibili per l'individuazione degli strumenti, questo articolo analizza il potenziale valore scientifico del catalogo e ne descrive le principali sfide implementative. Infine, viene presentato un prototipo di catalogo e le possibili strade da percorrere per proseguire lo sviluppo.

Parole chiave: AIUCD2023, catalogo, filologia digitale, strumenti software, software per l'editing digitale

1. Introduction

After initial experiments in the 1940s, the advent of the World Wide Web sparked a proliferation of software tools for digital scholarly editing ([11], [13]), which in turn established new practices and methodologies for producing and publishing digital editions. Today, editors can use a wide and diverse range of software tools, for example, to transcribe documents either manually or automatically, to collate and compare different versions of a text, and to publish their editions. However, finding accurate, up-to-date information about these tools and keeping track of all the software products that exist or have existed is challenging.

While catalogues of digital humanities tools, such as the Text Analysis Portal for Research (TAPoR) [16],[1] and data aggregators already exist, the academic community lacks a catalogue focused on digital scholarly editing. Investing time and resources in such a catalogue is worthwhile, as software development is a vital scholarly activity in the digital humanities. Many of the tools available for scholarly editing purposes are built by researchers "to make the general task of scholarship easier for other scholars" [15]. A dedicated catalogue would add value to the work of these scholars. The interest in software development as an important scholarly activity, and the variety of tools available for digital scholarly editing, calls for the creation of a specialised catalogue. If properly designed, this catalogue would benefit both scholars creating digital editions and software developers tasked with their technical implementation, helping them to select the tools that best meet their scientific and technical requirements.

This research stems from my doctoral studies on the development of software tools for digital scholarly editing. During my survey of the state of the art, I experienced first-hand the lack of a dedicated catalogue. My research led to the cataloguing of 58 different tools—detailed in an appendix to my thesis [10]—and the development of a prototype catalogue, which I presented at the 12th annual AIUCD conference in Siena. While this prototype doesn't fully address the identified gap, its development and the feedback from the conference helped to clarify both why and how to create an effective software catalogue for digital philology. In this article, I first outline the current state of the art and the reasons for creating the catalogue. I then present the main theoretical and practical considerations for creating such a catalogue. I conclude with a brief overview of the prototype and potential future developments of this research project.

2. State of the art

The two catalogues of scholarly digital editions, created by Patrick Sahle [17] and Greta Franzini [6] respectively, have proven to be indispensable research tools, allowing scholars to keep a wider perspective on the field of digital philology. Both catalogues provide information on tools and technologies used in digital editions. More specifically, Sahle's catalogue includes a separate page with brief descriptions of 24 major tools and infrastructures, whereas the description of individual digital editions rarely includes information about the tools used. For example, the description[2] of *Petrarchive* [12] mentions the visualisation

[1] Text Analysis Portal for Research (TAPoR). https://tapor.ca/home

[2] "The project aims at offering a 'rich text', interactive edition of one of the icons of western literature, Petrarch's *Rerum vulgarium fragmenta*. The prototypes consist of (1.) TEI-encoded documents from which we can render both diplomatic transcriptions and edited views of the text; (2.) Web-based presentations providing both diplomatic and edited views of the text, implemented with the TEI Boilerplate system; (3.) Facsimile page images for [the manuscript] Vat. Lat. 3195; (4.) a

tool TEI Boilerplate.[3] These descriptions are not searchable via full-text queries, and the catalogue is not available for download, making it impossible to systematically extract information about the use of software tools.

Franzini's Catalogue of Digital Editions includes an infrastructure field that lists programming languages, technologies, and tools used in the development of digital editions—with information available for more or less half of the records. Despite the greater granularity and detail of the information, identifying the software tools for textual scholarship mentioned in the catalogue is challenging, as the entries contain only names, without links to external resources or publications for verification. For example, in the description of the *Codice Pelavicino Digitale* [1], the infrastructure field contains the following data: "HTML, CSS, EVT, JavaScript, jQuery",[4] with the visualisation software EVT (Edition Visualization Technology)[5] mentioned in its abbreviated form, without any link to an external resource or a clear distinction from the programming languages.

Many of the tools listed in Franzini's catalogue appear obscure to users, either because of limited use in the academic community, discontinued availability, or because they are listed under outdated version names. For example, several entries list kiln,[6] an open source framework for building and deploying Web sites that primarily use XML content. Developed by a team at the Department of Digital Humanities (DDH), King's College, this tool was used to implement numerous digital editions, but it was last updated in 2012.

The aforementioned TAPoR is a collection of research tools for digital humanists. Developed by Geoffrey Rockwell and Milena Radzikowska with the support from the Arts Resource Centre at the University of Alberta, TAPoR is available as both a database and an online platform. Initially, as the name suggests, the collection focused on textual analysis, but it was later expanded to include the DiRT (Digital Research Tools)[7] Directory [7]—a collection of digital research tools developed at the University of California, Berkeley in 2013—and other tools that work with non-textual data and provide services used by digital humanists, such as publishing tools, GIS tools, and communication tools.[8] Several popular software products that are widely used in digital philology and scholarly editing are included in TAPoR, although the list is neither comprehensive nor fully up-to-date.

TAPoR provides a valuable example of how a catalogue of software tools should be implemented. Records are categorised according to the scholarly activities they enable, as defined

visual index, or map, to the Rvf as constituted in [the manuscript] Vat. Lat. 3195" (*Petrarchive*'s record in Sahle's catalogue). https://digitale-edition.de/e504

[3] http://teiboilerplate.org/

[4] Catalogue of Digital Editions: Codice Pelavicino, DOI: http://hdl.handle.net/21.11115/0000-000B-D153-B

[5] http://evt.labcd.unipi.it/

[6] https://kiln.readthedocs.io/en/latest/

[7] Digital Research Tools (DiRT) Directory. https://live-digital-humanities-berkeley.pantheon.berkeley.edu/resources/digital-research-tools-dirt-directory

[8] For a complete list of the types of tools included in TAPoR see the About page of the project: https://tapor.ca/pages/about_tapor.

by the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH).[9] This taxonomy includes the following main activities: analysing, capturing, creating, disseminating, enriching, interpreting and storing. Each main activity comprises a number of sub-activities, which are further subdivided into additional categories. For example, under the activity *enriching* one can find the sub-activity *editing*, and under editing the activities *adding*, *correcting*, *merging*, etc.

TAPoR's online platform provides an overview of the entire collection on the home page and a detailed description for each record, including:

1) general information, i.e., the name of the tool, the date of the last update of the record, the link to the official website of the tool and a short description;
2) information about the creator of the tool, and
3) attributes describing the main characteristics of the tool, such as the type of license under which it is available, and the TaDiRAH categories that best describe its functionality and features;
4) tags and related tools to help users browse the collection;
5) user reviews and comments.

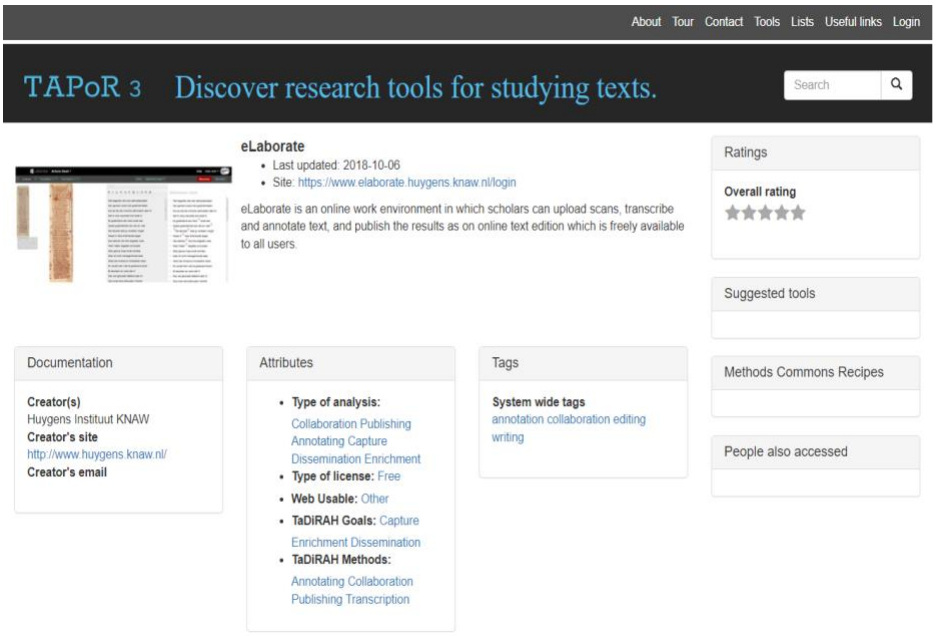


Figure 1 Web page dedicated to the eLaborate tool within the TAPoR platform.

The TAPoR platform also provides an advanced search interface that allows users to filter tools by other characteristics, such as the ease of use, development status, and possible warnings.

[9] Taxonomy of Digital Research Activities in the Humanities (TaDiRAH). https://vocabs.dariah.eu/tadirah/en/

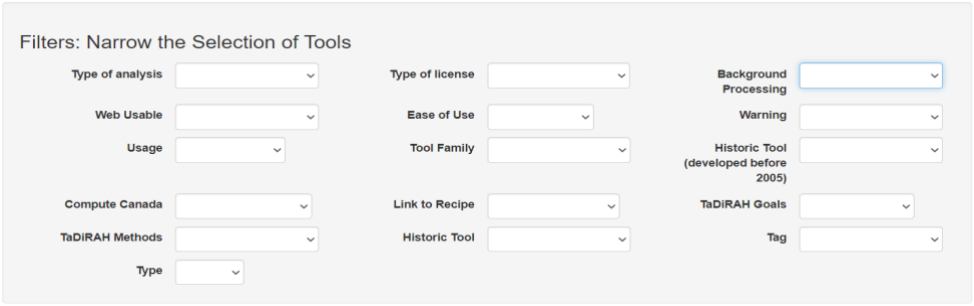
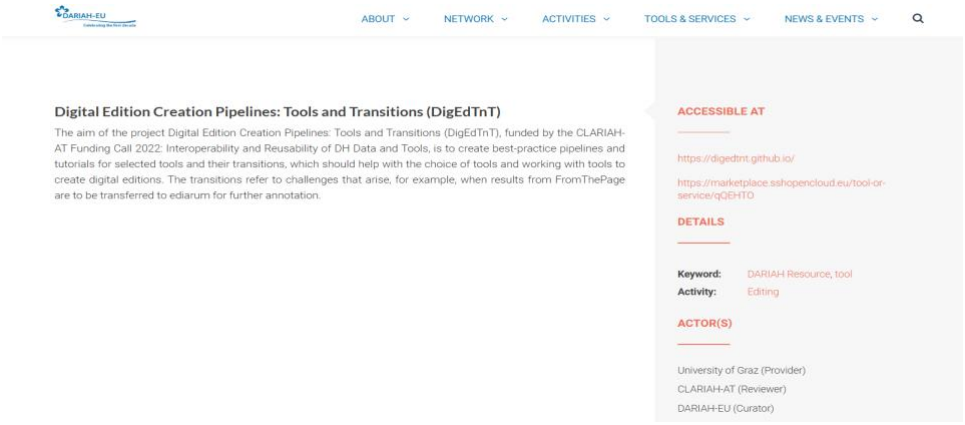


Figure 2 Search form on the TAPoR platform.

With its current 1718 records, TAPoR may be too large and diverse to prove itself as a practical tool. This problem also affects the OpenAIRE research information dataset,[10] which currently contains over 1200 research software[11] records tagged as pertaining to the Digital Humanities and Cultural Heritage field. To facilitate the discovery of tools, TAPoR provides lists of records, curated by associated editors, that focus on a particular area or theme in the field of digital humanities.

The DARIAH-EU's platform offers a catalogue of community-developed tools and services,[12] with a total of 292 records. For each indexed tool the catalogue provides the name and a short description, accompanied by URLs of the access points, keywords, afforded activities (as defined in the TaDiRAH taxonomy), and actors involved.



[10] OpenAIRE is a non-profit organisation which aims at promoting open science produced in Europe, by offering a set of public services such as a catalogue of scholarly data. https://explore.openaire.eu/

[11] OpenAIRE's dataset indexes four main types of research products: publications, research data, research software and other research products.

[12] https://www.dariah.eu/tools-services/tools-and-services/

Figure 3 Example of how a record is displayed in the DARIAH catalogue of tools and services.

The listed tools can be searched by title and filtered by category, afforded activity, and the actor involved in the development and maintenance. The categories are numerous, ranging from applications to aggregators, repositories to workflows.[13] In fact, a common problem of the three catalogues analysed so far is that it is not clear what kind of software tools are included.

Other valuable resources for discovering software tools for digital scholarly editing are:

- Lists curated by cultural institutions and associations, such as AIUCD[14] and EADH.[15]
- Lists curated by individual scholars or research teams, such as the one created by Cinzia Pusceddu as an appendix to a book on digital philology [3].
- Reviews published in the RIDE (A Review Journal for Scholarly Digital Editions and Resources) journal.[16]
- Proceedings of the main conferences related to DH.
- Publications and technical reports.

Reviews, conference proceedings, and other publications provide detailed information on the scientific and technical features of individual tools, while lists provide a broad but limited overview of the landscape. However, these resources are not suitable for systematic data mining and analysis because they present information in different formats and levels of detail.

3. Motivation

Having identified the lack of a catalogue of software tools for digital scholarly editing, a few remarks are needed in order to explain why this gap needs to be filled. Firstly, such a resource would help potential users—both individual scholars and cultural institutions such as libraries and archives—to find the tools best suited to their research goals. If well designed, the catalogue could quickly answer most of the questions potential users might have, especially those related to the stability, accessibility, and adequacy of the tools. Does the developer provide user support? What technologies is the tool based on? What are the system requirements? Is the tool available for free or for a fee? Does the tool meet my scientific needs?

With easier access to information about available software tools and guidance on how to integrate them into editorial workflows, more individual scholars and cultural institutions could produce effective, high-quality digital editions. In addition, the catalogue would help to bridge the perceived divide between traditional and digital philology by normalising the use of software

[13] To illustrate how long and varied is the list of categories used in DARIAH's tools and services catalogue, we fully report it. APIs Repository/Gateway, Access, Aggregators & Integrators, Applications, Archiving, Compute, Data Analysis, Data Archives, Data Exploitation, Data Management, Data Storage, Digital Preservation, Digitisation, Discovery, Education, Education & Training, Image/Data Analysis, Libraries, Machine Learning, Network, Platform, Processing & Analysis, Production, Publication, Related Training, Repository, Scholarly Communication, Scientific/Research Data, Services, Sharing & Discovery, Software, Software Libraries, Training & Support, Training Platform, Workflows.

[14] https://www.aiucd.it/progetti/

[15] https://eadh.org/projects

[16] https://ride.i-d-e.de/

tools for both digital and print editions. These tools could then become an integral part of shared, authoritative philological methodologies.

Secondly, it would make it easier for the research community to follow changes and developments in the various tools. Software products developed in this area tend to remain in use for a few years and then disappear quietly. This phenomenon is very common, both because of the inevitable and rapid changes that occur in technology and because their development and maintenance often depend on time-limited project funding [10]. In other cases, it is difficult to deduce from documentation and publications the status of a software, whether it is still maintained or whether it is compatible with current technologies. Another common case is when a software has been restarted by a different development team or agency and even under a new name. All these problems make it difficult to evaluate and select tools, especially for young researchers who are new to the field and do not know the history of how tools have been developed. A notable example is Collate,[17] the well-known automatic collation software developed by Peter Robinson. Originally released in 2000, along with Anastasia (Analytical System Tools and SGML/XML Integration Applications),[18] another tool developed by Robinson, was integrated into the SDE (Scholarly Digital Editions)[19] publishing system and platform. In 2010, Collate was re-launched as CollateX[20] by an international team of scholars and developers in order to adapt it to modern technologies. CollateX was then integrated into a new editing and publishing system developed by Robinson in 2018, called Textual Communities,[21] while in 2020, SDE was relaunched as Inkless Editions.[22] Keeping track of all these changes and knowing how to access and use Collate today is not easy.

Thirdly, the catalogue would serve as a highly valuable research tool for analysing the field of digital philology and its history from a more technical point of view. A catalogue focused on software for digital scholarly editing, would allow scholars to study the technological trends that are dominant today and have dominated the practice of scholarly editing in recent years. The data in the catalogue could be used to identify the software products that have been most widely used over a given period and, as a further step, to trace the most common workflows (or recipes). For example: between 2010 and 2015 most digital scholarly editions were produced using the editing software A, then analysed using software B and published as websites using software C. It would be possible to carry out a kind of archeological research, keeping a historical record of all the software tools that have been created and their impact on the field. In this regard, TAPoR already allows users to search for historical tools, developed before 2010, 2005 or 2000. Interestingly, TAPoR's advanced search interface also allows users to filter historical tools between *influential*, *no longer in active development*, *development sustained to present*, and *discontinued (no longer available)*.

[17] https://digitalmedievalist.org/2012/04/01/collate-text-editing-software/

[18] http://www.sd-editions.com/anastasia/index.html

[19] http://www.sd-editions.com/

[20] https://collatex.net/

[21] https://textualcommunities.org/

[22] http://www.inklesseditions.com/

4. Theoretical framework

Creating a catalogue of software products for digital scholarly editing requires careful consideration of both theoretical and practical aspects.

From a scholarly perspective, the first step is to define what types of tools belong in the catalogue. As shown in the previous section, existing catalogues contain diverse tools, services, and software products—from APIs to web applications to software libraries—each with different architectural and functional characteristics. To maintain flexibility, we could broadly classify all catalogue entries as software tools. A key consideration is whether to limit the catalogue to tools specifically designed for digital scholarly editing or to include widely used general-purpose tools such as WordPress. The line between general purpose tools and specialised scholarly editing tools is often blurred. Furthermore, including too many tools would be counterproductive, as the analysis of existing catalogues shows. I therefore propose two selection criteria. The first criterion includes tools specifically designed for digital scholarly editing tasks recognised by the academic community. These are typically created by humanities scholars with programming expertise or by computer scientists working under their supervision. The Classical Text Editor (CTE),[23] developed by the classicist Stefan Hagel in 1997 and still in use today, exemplifies this category. The second criterion includes tools that, while not specifically designed for textual criticism, have proven to be highly effective and have been widely adopted by the scholarly community. Oxygen,[24] a general-purpose XML editor, illustrates this category—it is frequently adopted by scholars creating digital editions using the XML/TEI standard.

The two selection criteria require a clear prior definition of textual criticism activities. The aforementioned TaDiRAH taxonomy offers guidance in this regard, by defining common activities performed by humanities scholars in a digital environment. In order to identify the activities specific to digital textual criticism, two aspects need to be considered. First of all, even though digital scholarly editions "are guided by a digital paradigm in their theory, method and practice" ([18], p. 28), activities of the traditional print-oriented editorial workflow—transcribing primary sources, collating, lemmatising, etc.—are also performed in digital workflows [9]. Therefore, all software tools that are either designed or successfully applied to perform one or more of these activities should be included in the catalogue. The second aspect is the existence of a technological and methodological model for producing digital scholarly editions, that has been widely adopted in the academic community over the last thirty years, the so-called "source-output model":

> "Secondo questo approccio, il testo annotato (o codificato) viene archiviato in uno o più file che nel loro complesso prendono il nome di sorgente (al maschile); nella maggior parte dei casi i file sorgente sono codificati nel linguaggio XML secondo le linee guida prodotte dalla TEI, che rappresentano lo standard di fatto nel campo delle edizioni digitali, oltre che un importante riferimento intellettuale [...]. Una serie di programmi informatici prendono il sorgente come punto di partenza (o input) e producono diversi oggetti digitali in uscita (detti, quindi, output): per esempio, partendo dallo stesso sorgente XML possono produrre una o più versioni HTML da

[23] https://cte.oeaw.ac.at/

[24] https://www.oxygenxml.com/

> pubblicare sul Web, un testo stampabile in PDF, un eBook da leggere su un tablet, degli indici interattivi, ecc."[25] ([13], p. 47).

Given the popularity of this model, marking up a document in XML/TEI format can be considered an activity of the textual criticism practice in a digital environment. For this reason, tools such as the Oxygen XML editor qualify as software tools for digital scholarly editing.

Having established the selection criteria for tools to be included in the catalogue, the next crucial question is what metadata to use to describe these software tools. Some basic metadata are easy to choose, as they are essential for identifying tools and their context. Using the Dublin Core metadata standard[26] as a reference, the catalogue needs to include the name of the tool, a brief description, the URL of the access page, the names of the individual scholars and institutions involved in the creation, development, and maintenance of the tool, and the license under which the tool is available. Temporal information—such as the date of first official release and last update—is essential for distinguishing historical tools and determining whether a tool is actively maintained. Ideally, the catalogue should track or reference all existing versions and official releases.

Beyond basic metadata, the catalogue should provide useful information for both target user groups: humanities scholars and software developers. For developers, key information includes programming languages and technologies used, source code availability (to enable code reuse and customisation), and input/output data formats to help identify possible workflows.

Scholars should be able to browse the catalogue to find tools that best suit their research goals. This requires categorising tools according to their objectives and activities. Using the "source-output model" as a reference, tools can be grouped into three main categories: production tools which assist scholars in editing; visualisation tools, which enable publication in user-friendly interfaces; and mixed tools, which support both editing and publication. The TaDiRAH taxonomy provides a workable framework for further describing the purpose of tools.

Different types of tools may require different additional metadata sets. For instance, automatic transcription tools should specify supported languages and accuracy rates. Following the example of how the latest version of Franzini's catalogue was implemented ([4], [5]), further properties can be identified through surveys to detect the attributes that the scientific community finds most useful. Finally, the catalogue should help users to evaluate the suitability of a tool by providing links and bibliographic references to publications, presentations, and reviews about the tool, as well as links and bibliographic references to existing digital editions that use it. This offers first-hand evidence of what can be achieved with each tool. Links alone would not ensure long-term access to external resources, as they may expire or change.

A final key issue is how the data will be presented to end users. The catalogue should provide a search interface and download options to facilitate information discovery and analysis. As the catalogue will include both current and historical tools—of varying stability and popularity—the

[25] According to this approach, the annotated (or encoded) text is stored in one or more files which collectively take the name of source; in most cases, the source files are encoded in XML following the guidelines produced by TEI, which represent the de facto standard in the field of digital editions, as well as an important intellectual reference [...]. A series of computer programs take the source as a starting point (or input) and produce various digital objects as output: for example, starting from the same XML source, they can produce one or more HTML versions for web publishing, printable PDF text, an eBook to read on a tablet, interactive indexes, etc.

[26] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

interface could offer two distinct exploratory paths. The first would be functional, aimed at scholars seeking tools for their editorial workflow. The second would support research into the history and methodology of digital philology. The function-focused interface should help users to identify suitable tools through effective filters, such as those offered within the DiRT repository website,[27] based on the research activities that the tools enable and the type of data that the tools can be used for (e.g., manuscripts, music, audio recordings).

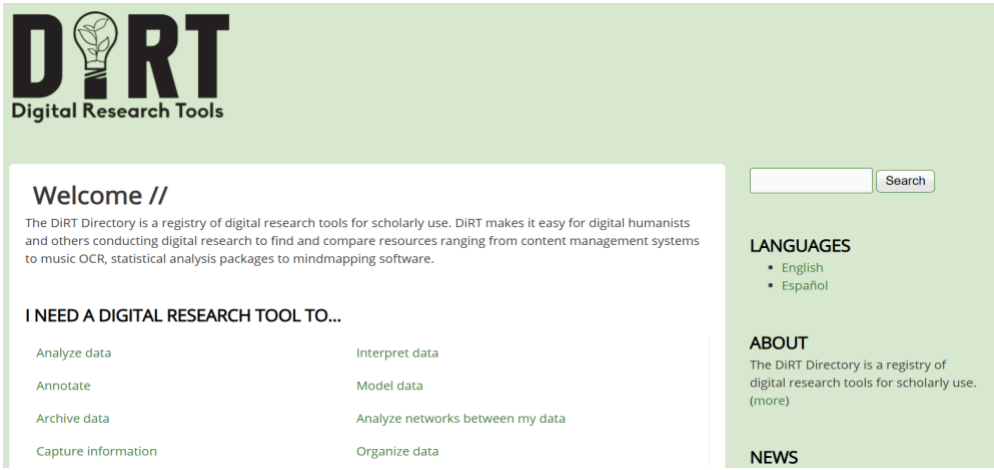


Figure 4 Home page of the DiRT website.

The function-focused interface should also guide users step-by-step through the creation of editorial workflows, offering recipes similar to those developed in the *Standardization Survival Kit* [14] by the PARTHENOS project.[28]

The research-focused interface would need to include a broader set of tools upstream to capture lesser-known software. This interface should provide date-based filtering (e.g., date of last update) and highlight tool relationships, such as when software A is derived from software B. The provenance ontology [8] can provide guidance for this.

5. Implementation

Many factors concerning the development and maintenance of the catalogue also need to be considered from a practical perspective, in order to make the cataloguing activity sustainable and durable, and to determine the technologies that will be used to build the catalogue.

The first aspect to consider is who should be responsible for the design and implementation of the catalogue. Ideally, a group of DH researchers should lead this effort, working with software developers and philologists, either as permanent team members or as external consultants. Collaboration between humanities scholars and software developers is essential to create a

[27] The website of the DiRT repository is no longer available, but its snapshots can be viewed through the wayback machine.

[28] https://www.parthenos-project.eu/

catalogue with appropriate metadata, search interfaces, and data visualisations. For specific issues, such as the selection of domain-specific metadata or the design of data visualisations, consultation with the wider scholarly community through questionnaires can prove effective.

The second aspect that requires careful consideration is data management and updating. With adequate funding, the optimal solution is to assign this responsibility to a small editorial team of two or three people. At the same time, regardless of the terms of employment of the editors (permanent or temporary), the catalogue can benefit from public participation through crowdsourcing systems. Web forms would allow software creators and others to submit or update tool records. However, all crowdsourced information requires editorial review before publication. While a system for end-users to report issues enables swift problem resolution, the catalogue would need in any case systematic, periodic reviews, given the rapid pace of technological change. Key areas requiring frequent review include embedded links, software status, latest versions, and new publications or digital editions. To streamline data management, automated information retrieval through APIs could pull data directly from software repositories containing source code and releases.

Regarding the publication of the catalogue, the expected outcome is a dataset that conforms to the FAIR principles [20] and is accessible in three main ways:

- File (or set of files) in a standard and non-proprietary format, downloadable and deposited in a repository such as Zenodo, to ensure long-term access to the data.
- User-friendly exploratory environment to consult information, perform searches, and get an overview of software development in the field of digital philology.
- API, so that external services can reuse or reference the data entered in the catalogue.

To ensure that the catalogue is truly FAIR-compliant, it requires comprehensive documentation describing the data model and persistent identifiers for individual records. These measures will help to establish the catalogue as a reliable, enduring scholarly resource.

Finally, once the catalogue is available, it will be crucial to pursue dissemination activities and to produce studies that demonstrate how the catalogue serves as a tool for the study of digital philology and its methodologies.

6. The prototype

The prototype catalogue comprises 58 records, and for each tool, the following information is provided:

- *Name* - The name of the tool.
- *Website* - The official website, repository, or wiki where the creators present the tool.
- *Category* - One of three main categories: *Visualization* (tools exclusively for displaying or publishing a DSE), *Mixed* (tools for both editing and publishing), or *Production* (tools that assist the philologist during specific editorial phases). Although the categories may overlap, the classification is based on primary functions and output types. Tools are labelled as visualisation or mixed if they can produce a complete digital scholarly edition.
- *TaDiRAH activity* - The relevant activities from the TaDiRAH taxonomy that best describe the main functions of the tool.
- *Begin date* - The release date of the tool's first public version. For experimental tools without an official release, this is the creation date of the source code repository or

the date given by the creators. If only the year is known, 1 January is used; if only the month and year are known, the first day of that month is used.

- *End date* - the date on which development of the tool ceased. For partial dates, the same convention applies as for begin dates.
- *Description* - The self-presentation of the tool from its website, repository, or wiki. This text outlines key features, offering insight into how creators conceived their tool.
- *Institutional partner(s)* - Organisations providing financial support, including universities, research centres, cultural institutions, and funding bodies. This information helps to assess the stability of the tool in terms of development, maintenance, and support.
- *Creator* - Individual developers or organisations responsible for the tool. For commercial software, the company name is given. For institutional tools, the name of the organisation or research group is provided. For academic projects, the names and affiliations of the creators are given.
- *Input format(s)* - Supported input formats. If sources list generic types such as "text" or "image", common formats are assumed (e.g., TXT for text, JPEG for images).
- *Output format(s)* - Available export formats for the DSE or other data. Note that some tools, particularly mixed or visualisation tools, may not support data export.
- *Technologies* - Programming languages, tools, libraries, and frameworks used for development. Some tools may be integrated with others (e.g., OpenSeadragon and VisColl in EVT).
- *System requirements* - The type of tool (desktop software, web service, library, web application) and technical requirements, including compatible operating systems.
- *Collaborative working* - Whether the tool supports multiple users working simultaneously on shared materials.
- *Open source* - Whether the source code is freely available.
- *Repository* - The location of the tool's source code.
- *License* - The license terms of the tool. "For a fee" indicates paid software; "Free" indicates free software without a specific open source license.
- *Current version* - The latest public release version number.
- *Editions* - DSEs created with the tool, with links to their official websites. This helps potential users to evaluate the tool's capabilities and results.
- *Publications* - References to publications, presentations, reviews, or case studies about the tool.

The catalogue prototype requires further refinement and completion, as the catalogue currently focuses on tools mainly developed in Italy, Europe, and the United States. Complete information is not always available from official sources or reliable publications for all tools, although some missing details could be obtained through additional research or by contacting the creators directly. Some information, including website links, may be out of date. The metadata listed above needs to be revised to optimise data organisation and simplify information management. For example, the catalogue needs an additional field to distinguish software products by their technical type—such as APIs, web applications, software libraries, and desktop applications. These categories should be created in collaboration with software developers, referencing current software engineering standards to create clear, effective terminology. In addition, each technology entry should include a brief description and a link to an external resource. This would make it easier to identify technologies used and maintain historical references even after specific technologies have become obsolete. Finally, some records need to be restructured to better present the data. For example, software EVT currently appears as a single record, when it should

be split into three separate records—one for each version—as these versions have different features and architectures.

The prototype catalogue was developed using Airtable,[29] an intuitive platform for creating and publishing datasets. This platform provides several useful features for building a catalogue—editors can collaborate on multiple records, while end users can comment on records, submit new entries via a web form, and export records in CSV format or print them. These features are currently only available to authenticated users, while other users can view or print the catalogue as a table with filter, sort and search options.[30] The catalogue is also available for download, as it has been deposited in Zenodo[31] in CSV format, to ensure long-term access and reach a wider audience.

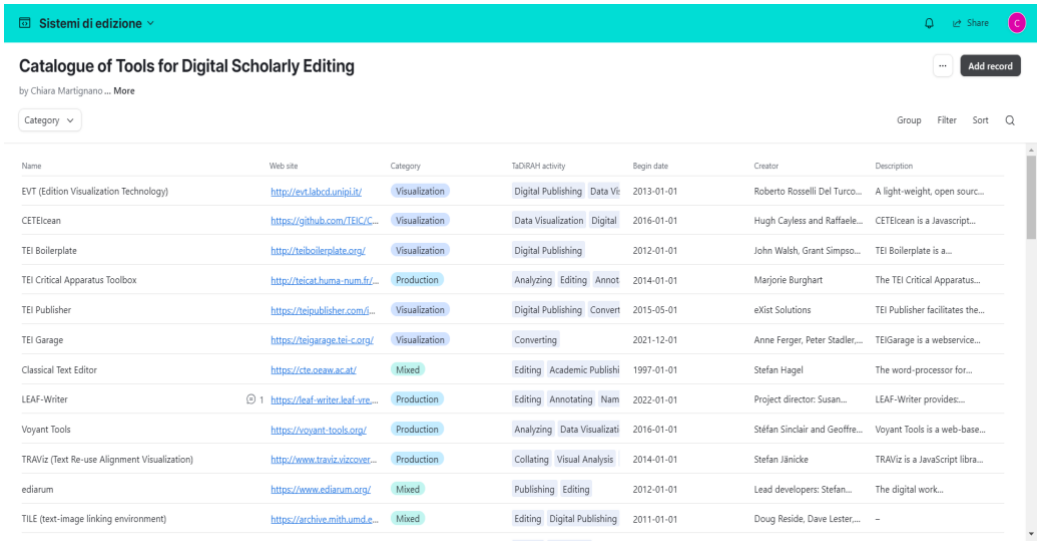


Figure 5 View of the prototype catalogue on the Airtable platform.

7. Future developments and conclusions

This research and the catalogue prototype aim to establish a foundation for a future, more comprehensive catalogue of digital scholarly editing tools, ideally developed as a stand-alone project. In the meantime, one possible approach is to enhance and promote the existing work by integrating the catalogued tools into established resources. For example, the prototype catalogue could become a curated list in the TAPoR collection. Another promising option is to

[29] https://www.airtable.com/

[30] Link to access the catalogue on Airtable: https://airtable.com/appmd2Z1LsaYMaYEF/shrb1dVa8QsJ4P8Ga. Link to comment on catalogue's records or add new records (authentication is necessary): https://airtable.com/invite/l?inviteId=invSJqDJNb34VNS1t&inviteToken=68e314020322720e2720897ab6a34f8d91926fd1970f04f519d3db2b4a33acb7&utm_medium=email&utm_source=product_team&utm_content=transactional-alerts.

[31] Link to access the catalogue on Zenodo: https://doi.org/10.5281/zenodo.8398700.

incorporate it into the knowledge graph of the ATLAS project.[32] This project—currently in its initial phase—aims to create a catalogue of digital scholarly objects in the field of Italian cultural heritage, leveraging semantic web technologies and standards [2]. The ATLAS prototype catalogue includes digital text collections, scholarly editions, linked open datasets, ontologies, and software tools [*ibid.*]. A notable strength of the ATLAS knowledge graph's data model is its inclusion of research projects, people, and organisations alongside various research products, effectively representing both scholarly outputs and their institutional origins. This data model could benefit the catalogue by streamlining information about different software tools developed within shared research contexts. Regardless of how this research continues, this article has demonstrated how a catalogue of digital scholarly editing tools can highlight and leverage the valuable development work of digital humanities scholars and their IT partners. I welcome any feedback on individual records in the prototype catalogue, as well as on its general definition and implementation.

Acknowledgements

I would like to thank professors Roberto Rosselli Del Turco and Lino Leonardi for providing me with references to several of the software tools catalogued. I would also like to thank Simone Zenzaro, Beatrice Nava, Christian D'Agata, and Marco Sartor for their valuable feedback during the AIUCD conference.

Funded by the European Union- Next Generation EU, Mission 4 Component 1 CUP J53D23013000006.

References

[1] Codice Pelavicino. Edizione digitale. 2020. Edited by Enrica Salvatori, Edilio Riccardini, Roberto Rosselli del Turco, Laura Balletto, Chiara Alzetta, Chiara Di Pietro, Chiara Mannari, Raffaele Masotti, Alessio Miaschi, 2nd ed., <http://pelavicino.labcd.unipi.it> ISBN 978-88-944430-2-8 DOI: https://doi.org/10.13131/978-88-944430-2-8.

[2] Daquino, Marilena, Alessia Bardi, Marina Buzzoni, Riccardo Del Gratta, Angelo Mario Del Grosso, Franz Fischer, Francesca Tomasi, and Roberto Rosselli Del Turco. 2024. 'The ATLAS: A Knowledge Graph of Digital Scholarly Research on Italian Cultural Heritage'. https://doi.org/10.6092/unibo/amsacta/7927.

[3] Fiormonte, Domenico. 2003. *Scrittura e filologia nell'era digitale / Domenico Fiormonte*. Nuova didattica Arte e letteratura. Torino: Bollati Boringhieri.

[4] Franzini, Greta, Melissa Terras, and Simon Mahony. 2019. 'Digital Editions of Text: Surveying User Requirements in the Digital Humanities'. J. Comput. Cult. Herit. 12 (1): 1:1-1:23. https://doi.org/10.1145/3230671.

[5] Franzini, Greta, Simon Mahony, and Melissa Terras. 2016. 'A Catalogue of Digital Editions'. In Digital Scholarly Editing: Theories and Practices. (Pp. 161-182). Open

[32] ATLAS is a two-year PRIN 2022 project funded by the European Union and a joint effort of the University of Bologna, the Ca' Foscari University and the Italian National Research Council. https://dh-atlas.github.io/

Book Publishers: Cambridge, UK. (2016), edited by E. Pierazzo and M. Driscoll, 4:161–82. Cambridge, UK: Open Book Publishers. https://doi.org/10.11647/OBP.0095.

[6] Franzini, Greta. 2012-. Catalogue of Digital Editions. https://zenodo.org/badge/latestdoi/42574907.

[7] Grant, Kaitlyn, Quinn Dombrowski, Kamal Ranaweera, Omar Rodriguez-Arenas, Stéfan Sinclair, and Geoffrey Rockwell. 2020. 'Absorbing DiRT: Tool Directories in the Digital Age'. Digital Studies / Le Champ Numérique 10 (1). https://doi.org/10.16995/dscn.325.

[8] Lebo, Timothy, Satya Sahoo, and Deborah L McGuinness. 2013. 'PROV-O: The PROV Ontology'. OWL 2. https://www.w3.org/TR/prov-o/.

[9] Leonardi, Lino. 2021. 'Filologia digitale del Medioevo italiano'. Griseldaonline 20 (2): 77–89. https://doi.org/10.6092/issn.1721-4777/12817.

[10] Martignano, Chiara. 2023. 'Un modello concettuale per le edizioni critiche digitali'. Doctoral thesis, Siena: Università degli Studi di Siena. http://dx.doi.org/10.25434/chiara-martignano_phd2023.

[11] Orlandi, Tito. 2010. Informatica testuale: teoria e prassi. Manuali Laterza 308. Roma [etc.: GLF editori Laterza.

[12] Petrachive: An Edition of Petrarch's Songbook. 2013-. Edited by Isabella Magni, Wayne Storey, John Walsh, and Francesco Marco Aresu. https://dcl.ils.indiana.edu/petrarchive/.

[13] Pierazzo, Elena, and Tiziana Mancinelli. 2020. Che cos'è un'edizione scientifica digitale. Le bussole 599. Roma: Carocci.

[14] Puren, Marie, Charles Riondet, Dorian Seillier, and Laurent Romary. 2017. 'The Standardization Survival Kit (SSK)'. July. https://hal.science/hal-01587687.

[15] Ramsay, Stephen, and Geoffrey Rockwell. 2012. 'Developing Things: Notes toward an Epistemology of Building in the Digital Humanities'. Debates in the Digital Humanities, 75–84.

[16] Rockwell, Geoffrey, Stéfan Sinclair, and Milena Radzikowska. 2020. 'Text Analysis Portal for Research (TAPoR) Archive'. ERA. 1 January 2020. https://doi.org/10.7939/r3-tnha-4924.

[17] Sahle, Patrick. 2020. 'A Catalog of Digital Scholarly Editions'. http://www.digitale-edition.de/.

[18] Sahle, Patrick. 2016. 'What Is a Scholarly Digital Edition?' In *Digital Scholarly Editing*, edited by Matthew James Driscoll and Elena Pierazzo, 1st ed., 4:19–40. Theories and Practices. Open Book Publishers. https://www.jstor.org/stable/j.ctt1fzhh6v.6.

[19] Schreibman, Susan, and Ann M Hanlon. 2010. 'Determining Value for Digital Humanities Tools: Report on a Survey of Tool Developers'. *Digital Humanities Quarterly* 4 (2).

[20] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.