Software solutions for web information systems in digital humanities: review, analysis and comparative study

Maurizio Toscano; Manuel J. Cobo; Enrique Herrera-Viedma

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86613

Recommended citation:

Toscano, Maurizio; **Cobo, Manuel J.**; **Herrera-Viedma, Enrique** (2022). "Software solutions for web information systems in digital humanities: review, analysis and comparative study". *Profesional de la información*, v. 31, n. 2, e310211.

https://doi.org/10.3145/epi.2022.mar.11

Article received on July 18th 2021
Approved on February 14th 2022

Maurizio Toscano ✉
https://orcid.org/0000-0001-5418-3859

Fecyt
European Office
C/ Pintor Murillo, 15
28100 Alcobendas (Madrid), España
maurizio.toscano@fecyt.es

Manuel J. Cobo
https://orcid.org/0000-0001-6575-803X

University of Granada
Department of Computer Science and Artificial Intelligence
Research Institute DaSCI
Periodista Daniel Saucedo Aranda, s/n
18014 Granada, Spain
mjcobo@ugr.es

Enrique Herrera-Viedma
https://orcid.org/0000-0002-7922-4984

University of Granada
Department of Computer Science and Artificial Intelligence
Research Institute DaSCI
Periodista Daniel Saucedo Aranda, s/n
18014 Granada, Spain
viedma@decsai.ugr.es

Abstract

Research in the humanities increasingly depends on how information is structured and managed and how, on the basis of that information, new knowledge is produced. Additionally, participatory approaches, which often rely on web information systems as their supportive infrastructure, have made an impact on the most recent historiographical trends, in particular in the methodological framework of digital humanities. The aim of this paper was to produce, from an operational and implementation perspective, a review of software solutions frequently used to develop web information systems for research projects in humanities and cultural heritage, in order to provide an understanding of the various possibilities available and their positives and limitations, also based on different users' requirements. An individual and comparative analysis of sixteen different application frameworks commonly used in these fields, either generic or developed for a specific research domain, has been carried out, considering their main functionalities, strengths, and weaknesses. The achieved results facilitate critical and reasoned decision-making among several available options, guiding the makers of those systems, both researcher(s) and developers(s), and providing them also with a common ground of terms and use cases to facilitate their necessary dialogue.

Keywords

Web information systems; Digital humanities; Content management systems; Digital assets management; Virtual research; Environment; Crowdsourcing; Comparative analysis; Software review.

Funding

This article is a result of the projects: "Disruptive group decision making systems in fuzzy context: Applications in smart energy and people analytics", 2020-2023, Reference: PID2019-103880RB-I00, funded by MCIN/AEI/10.13039/501100011033, PR Enrique Herrera Viedma; and "New fuzzy systems for decision making: Applications in digital environments", 2021-2023, Reference: P20_00673, financed by *Junta de Andalucía*, PR Enrique Herrera Viedma.

1. Introduction

An information system is an arrangement of people, data, processes, and information technology that interact to collect, process, store, and provide as output the information needed to support an organization (**Whitten** *et al.*, 2004). They originate in the business environment to manage, store, process and retrieve data, in order to produce information. They are collaborative environments by design, fostering sharing and teamwork, which is becoming increasingly important in the humanities. They predate the Internet (**Hirschheim**; **Klein**, 2012), but today the Web constitutes the ideal environment to deploy their technological component, thanks to user-friendly browsing interfaces, which employs tools widely familiar to the users, and the ability to operate almost independently, in terms of hardware and software, of both server and client-side operating systems. In addition to that, the Web is also an enabling factor that supports and favors opening these systems to other research groups or to society (**Toscano**, 2018).

The domain of application of this study are digital humanities (DH), a relatively recent trend in the study of the past, which stands at national and international level, both as an area of research and as a preferred subject for funding (**Burdick** *et al.*, 2013; **Toscano** *et al.*, 2020). In this context, web information systems (WIS) constitute a central piece in the process of digitization of research: through the different phases of planning, modeling, implementation, and use, they propitiate, and in a certain way force, the raising of new questions and the reformulation of existing ones, which ultimately constitutes the main challenge of DH at present (**Rodríguez-Ortega**, 2018). These systems represent, in many cases, the starting point of that process and often the first opportunity for the researcher using traditional methods to face the series of problems and opportunities generated by the digital turn.

In the development process of a WIS (Figure 1), once the data model has been designed and the functionalities needed to meet the user community requirements have been evaluated, it is necessary to choose the most appropriate working environment for the following software implementation. This paper focuses specifically on that phase, through a critical review of existing software solutions for the implementation of web-based information systems in humanities and cultural heritage (CH). An individual and comparative analysis of different application frameworks commonly used in these fields, either generic or developed for a specific research domain, has been carried out, considering their main functionalities, strengths, and weaknesses.

Figure 1. Infographic of the core phases in the development of a customized WIS.

Humanities can no longer ignore IT solutions to manage the rising amounts of data and information they produce, and to support knowledge generation through quantitative approaches, which have progressively grown in popularity. Research in the humanities increasingly depends on how information is structured and managed and how, based on that information, new knowledge is produced through interpretative processes. Additionally, participatory approaches, which often rely on WIS as their supportive infrastructure, have made an impact on the most recent historiographical trends, in particular in the methodological framework of public history and digital public history (**Gallini**; **Noiret**, 2011; **Noiret**, 2018; **Pons**, 2018). A growing interest from society to participate in co-creative and collaborative projects in the humanities and CH has been detected (**Bocanegra-Barbecho** *et al.*, 2017; **Bocanegra-Barbecho**, 2020), through social networks and within the framework of Citizen Science initiatives (**Hedges**; **Dunn**, 2018; **Ridge**, 2014; **Terras**, 2016). For these reasons, new public historians, and digital humanists in general, represent the main target audience of this paper, as promoters and makers of digital resources and infrastructures.

> Humanities can no longer ignore IT solutions to manage the increasing amounts of data and information they generate, and to support knowledge generation through quantitative approaches, which have progressively grown in popularity

Data, particularly in disciplines that study the past, do not tend to become outdated and, once stored in a structured way, can be exploited and reworked over decades. It is not the same for WIS, which belong to a continuously evolving environment, strictly dependent on the technologies available and used at a given time. In this regard, any review remains a snapshot of the current state of the technological evolution in a specific sector. Despite this, or precisely for this reason, it is necessary to comparatively review existing software solutions frequently used for research projects in the humanities and CH. Relying only on familiarity and previous experience with a specific software, which anyway are subjective factors worth considering, entails the risk to fall into the so-called golden hammer issue (**Brown** *et al.*, 1998), i.e. the obsessive application of a familiar technology to solve a wide variety of different problems. As stated by software developer **Gilgado** (2014):

> "The problem with using the same tools every time you can is that you don't have enough arguments to make a choice because you have nothing to compare to and is limiting your knowledge".

The aim of this paper was to produce such a review from an operational and implementation perspective, in order to provide readers with an understanding of the various possibilities available to build information systems, their positives and limitations, also based on different users' requirements. The final goal was to facilitate a critical and reasoned decision making among the available options, guiding the makers of those systems, both researcher(s) and developer(s), and providing them also with a common ground of terms and use cases to facilitate their necessary dialogue (**Scheuermann**; **Kroeze**, 2017).

The text is organized in five chapters, beyond this introduction: the methodology, which introduce the type of products under analysis and explain the criteria followed for their selection; the description of the selected software packages, organized in four sections; the comparative analysis, which present the discussion of results; a chapter devoted to limitations, and then the conclusions.

2. Methodology

Among software solutions used to build WIS, at least three main categories can be identified:

- Content Management System (CMS);
- Digital Assets Management (DAM);
- Virtual Research Environment (VRE).

CMSs are mainly used to collect, store, and publish on the Web alphanumeric digital content generated collaboratively by users, according to a certain workflow and in a user-friendly environment (**Martinez-Caro** *et al.*, 2018). DAM is a type of CMS focused on objects collections management: assets and media, along with their metadata, rather than alphanumeric content (**Jiménez**, 2003). VREs are similar to CMS, but built specifically for research, as virtual laboratories to support collaboration and provide built-in analytical and visualization tools (**Candela** *et al.*, 2013). Because of their specificities and their relevance for our domain of application, we decided to treat software solutions aimed to crowdsourcing (**Oomen**; **Aroyo**, 2011) as an additional separate category.

In order to select specific packages within these four groups, we established basic inclusion and exclusion criteria:

- Free and OpenSource Software (FOSS), allowing custom modifications, integrations, and adaptations without any restrictions, in order to meet the needs of DH projects;
- software under active development and maintenance[1];
- evidence of use cases and proven applications in the field of humanities or CH[2].

This review excludes, despite their popularity, software packages that offer by default limited customization and where the management of unstructured content prevails over data and metadata management and manipulation, as is the case, for example, of *Joomla* or *WordPress*.

Criteria used for the analysis come from the comparative examination of official documentation, combined with the approach taken in recently published similar reviews (**Nishanbaev** *et al.*, 2019). The resulting set is quite wide, comprising 15 parameters, but that does not exclude that some relevant factors may have been left out.

They have been grouped into three blocks;

1) The first includes essential features, like file formats handled, mapping, translatable interface, batch importing and exporting, and features extensibility via modules or plugins.

2) The second block focuses on data modelling and information retrieval functionalities, which is the distinctive set of features a WIS is expected to provide (**Lenz**; **Oberweis**, 1998).

3) The third block comprise some additional features that we considered worth including either because they respond to an interesting trend toward decoupled applications –REST API and Mobile app– or because they facilitate initial adoption and long-term maintenance –respectively online demo and SaaS (Software as a Service).

Identified criteria are not exempt from possible limitations linked, for example, to the wide set of software selected. That means that we cannot exclude that analysing, for instance, each category of software with a specific set of criteria would have provided some additional insight. In this regard, we decided to favour the intercomparability across a set of solutions that are often used as alternatives in similar contexts. We decided to proceed this way also because, thanks the progressive incorporation of new functionalities, the differences between software of different categories are gradually diminishing.

To carry out the comparative analysis, a combined approach has been taken testing each software package. Whenever possible, the protocol involved the analysis of the official documentation combined with the use of the online demo. In the few cases where a demo was not provided (Table 7), the software has been installed on a remote server. In addition to that, most of the selected software has been already subject of individual reviews published in research papers, which have been also taken into consideration during the analysis and cited in following chapter 3.

3. Description of selected software packages

There is a relevant number of technological solutions developed to address the needs of managing structured information and documentation on the Web. While some are clearly alternatives to each other, others cover a variety of different user cases, with functionalities that overlap only in part, and are more difficult to compare. A representative, albeit incomplete and temporary, list of FOSS applications used in the field of DH includes the following:

1. Content Management Systems:

- *MediaWiki*
- *Drupal*
- *BackdropCMS*
- *WissKI*
- *Mukurtu*
- *Omeka Classic*
- *Omeka S*
- *AlchemyCMS*

2. Digital Assets Management:

- *Islandora*
- *Arches*
- *CollectiveAccess*

3. Virtual Research Environment:

- *ResearchSpace*
- *Nodegoat*

4. Crowdsourcing:

- *Zooniverse*
- *Pybossa*
- *EnrichEuropeana*
- *CrowdHeritage*

As we will see in the following chapters, all the selected packages provide the developer with a working environment ready for configuration, with a greater or lesser degree of customization.

3.1. Content Management Systems

MediaWiki is the base software for *Wikipedia* and other *Wikimedia Foundation* projects. Launched in 2003, since then it grown a lot in popularity and has been used within a wide range of contexts. Its use in DH projects is widespread, for example for collaborative transcription activities (**Causer**; **Wallace**, 2012; **Transcribe Bentham**, 2021), catalogs of heritage assets (**Chevalier** *et al.*, 2012; **Care**, 2020), text corpus (**Rutherford** *et al.*, 2018) and all kinds of digital collections, in particular with the use of its extension *Semantic MediaWiki*, which adds functionality for structured data handling, SQL and SPARQL support and granular permissions control.

Drupal is a modular, multipurpose, and highly configurable software platform. It is the third most popular CMS[3], with a market share of 4.7%. Compared to its competitors, from which it differs by a steeper learning curve and greater flexibility, it is considered like a Content Management Framework (CMF), i.e. an environment for the development of customized web applications. In the latest versions[4] integrates by default functionality for handling and querying structured data. It has several strengths: the high number of available modules[5], which cover needs very relevant for the DH sector[6];

Table 1. Basic characteristics of selected CMS

Category	Software	Version analyzed	Updated on	Released on	Programming languages	Developer(s)	License
CMS	*MediaWiki*	1.35	25/09/20	2003	PHP, JavaScript	*Wikimedia Foundation*	GPLv2+
	Drupal	7.77	03/12/20	2011*	PHP, JavaScript	*Drupal* community	GPLv2/v3
	Drupal	8.9.11	03/12/20	2015*	PHP, JavaScript	*Drupal* community	GPLv2/v3
	BackdropCMS	1.17.4	26/11/20	2015	PHP, JavaScript	*Backdrop PMC*	GPLv2+
	Mukurtu	2.1.6	23/11/20	2012	PHP, JavaScript	*CDSC – WSU*	GPLv3
	WissKI	8.x-2.3	06/11/20	2012	PHP, JavaScript	*FAU-GNM-ZFMK*	GPLv2+
	Omeka Classic	2.7.1	28/01/20	2008	PHP	*CHNM – GMU*	GPL
	Omeka S	3.0.1	20/10/20	2017	PHP, JavaScript	*CHNM – GMU*	GPL
	AlchemyCMS	5.0	17/07/20	2010	Ruby	C. Fregin, T. von Deyen	BDS

* For *Drupal* we state the publication date of each version analyzed instead of the initial release date, that was 2001.

the possibility of combining modules to implement additional functionalities; its robust mapping and geocoding options; a very large community of users and developers. *Drupal* has been widely used in DH projects (**Dombrowski**, 2016a; 2016b; **Velios**; **Martin**, 2017), especially its version 7, launched in 2011 and whose official support has been extended until the end of 2023. The release of version 8, in 2015, has introduced some relevant changes in the base architecture[7], oriented to cover mainly the needs of large companies, which has increased development cost and web hosting requirements. Overall, these changes have slowed the adoption rate of *Drupal* 8[8] (Figure 2) and, in particular, its use in DH projects (**Dombrowski**, 2018). Additionally, it has led to the creation, in 2015, of a derivative fork: *BackdropCMS*. Version 9, which maintains a strong continuity with version 8, was launched on June 3, 2020, but its use is still marginal.

As we have pointed out, *BackdropCMS* was lunched as a fork of *Drupal 7*, with the aim to improve many of its shortcomings, under the principles of customization, affordability, speed, backward compatibility, and extensibility. It puts emphasis, in its development roadmap, on the editorial experience workflow and the support for multiple and right-to-left (RTL) languages. That, coupled with *Drupal 7*'s legacy in terms of data modeling and dynamic view creation, makes it particularly attractive in the context of DH and for custom WIS development. Currently, its major disadvantages are the level of adoption[9], still scarce if compared to *Drupal*, and the relatively low availability of modules[10].

An interesting phenomenon in the context of *Drupal* are its distributions[11]: pre-configured versions that bring together a set of modules to meet the requirements of a specific domain of application, some of which are very relevant for the scope of this paper. It can be considered new software that uses *Drupal* as the base component. Two of the most interesting distributions in the fields of CH and DH projects are *Mukurtu* and *WissKI*.

Mukurtu is a FOSS tool to provide digital access to CH. It has been co-designed to meet the particular needs of indigenous communities to manage, communicate and share their heritage digitally, in a culturally relevant and ethical manner. It has been developed and is currently maintained by the *Center for Digital Scholarship and Curation* at *Washington State University*. It is oriented in particular towards metadata curation, to correctly describe and tag community's indigenous heritage, providing differential access to information for community members and for the general public, through what it calls "cultural protocols" (**Wilberg**, 2014). Its development roadmap, which is public and has been recently updated[12], involves the migration to *Drupal 8* with *Mukurtu* version 4, scheduled for late 2021, and a one-year transition period for users to migrate to the new version, before official support for *Drupal 7* ends.

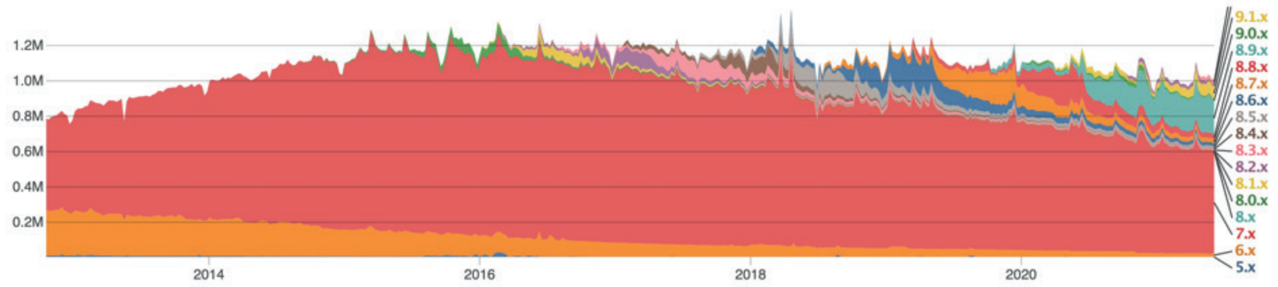

Figure 2. Adoption and usage rate of different *Drupal* versions, from 5 to 9.1. The use of version 7, with 64% of the total, continues to be prevailing. Source: *https://www.drupal.org/project/usage/drupal*

WissKI is defined by its creators as a web based VRE and CMS. Despite being multipurpose, it has an explicit focus on DH[13]. It has been developed by three scientific institutions, namely:

- the *Digital Humanities Research Group* of the *Department of Computer Science* at the *Friedrich-Alexander-University* of Erlangen-Nuremberg,
- the *Department of Museum Informatics* at the *Germanisches Nationalmuseum* in Nuremberg and
- the *Biodiversity Informatics Group* at the *Zoologisches Forschungsmuseum Alexander Koenig* in Bonn.

It implements semantic web concepts and is clearly oriented towards scientific research projects that collect, store, manage and communicate knowledge produced in those contexts, with a focus on long-term interoperability (**Scholz**; **Goerz**, 2012). It incorporates the ontology Erlangen CRM, which is an OWL-DL 1.0 (Web Ontology Language) implementation of *Cidoc-CRM*, even if it supports any other ontology. As RDF store, it uses the ARC *triplestore* that provides a SPARQL endpoint to query available data (**Nishanbaev** *et al.*, 2019).

> Data, particularly in disciplines that study the past, do not tend to become outdated; it is not the same for web information systems (WIS), which belong to a continuously evolving environment, strictly dependent on the technologies available and used at a given time

From a DH perspective, *Omeka* is considered one of the main alternatives to *Drupal* and its distributions. Its structure is designed primarily to manage digital collections of objects, enriched with metadata. It has been developed since 2007 by the *Roy Rosenzweig Center for History and New Media* at the *George Mason University* (Virginia), as a software solution for cataloging and presenting cultural objects from libraries, museums, archives, collections, and virtual exhibitions (**Alcaraz-Martínez**, 2012). It is currently available in two different versions: *Omeka Classic* and *Omeka S*. The first one, oriented towards individual projects and the education sector, is based on the *Dublin Core* metadata standard and its functionalities can be extended through plugins. The extension possibilities are much more limited than in *Drupal*, but there are modules very relevant to our area of interest, such as *Neatline*, to incorporate maps and timelines, and *Scripto*, a tool for collaborative transcriptions. *Omeka S*, whose first version was released at the end of 2017, has a more institutional scope, as a centralized resource from which to develop multiple platforms. Furthermore, it is not limited to *Dublin Core*, but can implement multiple vocabularies/ontologies and facilitates the use of linked open data. In both versions, the metadata standards-based data model architecture is the main difference with *Drupal*, where the data model is fully customizable. In addition to storing items directly in the *Omeka* database, it is also possible to import digital objects (XML files, such as TEI or EAD, images or videos) from a *Fedora Commons* repository, as linked data, through the *FedoraConnector* module (**Gruber**, 2020).

To complete this section, it is worth to include *AlchemyCMS*, the CMS used by *Europeana Virtual Exhibitions*. At the software level, it is a headless platform, i.e. it provides only a backend built as a repository of structured content, which it delivers through a *RESTful* API. Because of its own architecture, it focuses on storing pure content, without markup language, style, or formatting. It does not provide a ready-to-use platform but makes available the necessary components to develop a custom CMS. It integrates *ImageMagick* technology and is therefore particularly suitable for the creation of image-based systems. The project started as a proprietary application in 2007, under the name *washAPP*, and became an open-source initiative in 2010. It offers a comprehensive and well-structured documentation and has a solid and active developer base.

3.2. Digital Assets Management (DAM)

Table 2. Basic characteristics of selected DAM

Category	Software	Version analyzed	Updated on	Released on	Programming languages	Developer(s)	License
DMS	*Islandora*	7.x-1.13	03/09/19	2006	PHP, Java, JavaScript	*Islandora Foundation*	GPL
	Arches	5.1	27/10/20	2013	Python, JavaScript	*Getty CI – World Monuments Fund*	GPLv3
	CollectiveAccess	1.7.9	01/01/21	2007	PHP, JavaScript	*Whirl-i-Gig*	GPLv3

Turning to the category of DAM systems, *Islandora* represents a well-established software solution for the creation of digital collections and the long-term preservation of resources. It comes from the combination of *Fedora* and *Drupal*, to help institutions and organizations to manage and locate collections of digital objects, in a collaborative way with their public. Originally developed by the *University of Prince Edward Island Robertson Library*, it is now maintained through an international foundation. It uses *Fedora* as a repository for digital objects storage, access control and versioning; *Drupal* for web interface development, user interaction, metadata creation, taxonomies, and extendibility via modules[14]. *Islandora* adds to the combination between *Fedora* and *Drupal* a whole series of specific modules (microservices) for the configuration of the platform and its interoperability with external services. It can store and display a wide variety of objects: images, audios, videos, PDFs, collections, paged content (books, newspapers, serials) and binary files (**Ruest**; **Stapelfeldt**, 2014; **Stapelfeldt**; **Moses**, 2013).

CollectiveAccess software began to be developed in 2003 under the name *OpenCollection*, by the company *Whirl-i-Gig*, but released its first public version only in 2007 (**Alcaraz-Martínez**, 2014). It is designed to manage and publish on the Web large and heterogeneous collections of libraries, museums and archives that require support for different metadata standards (*Dublin Core*, *PBCore* and *VRA Core* are integrated), reference thesauri (*Library of Congress subject headings*, *Getty*, *GeoNames*), file formats (images, video, audio, multi-page documents, 3D scans, etc.) and a variety of visualization options (zoomable high resolution images, annotations, dynamic timelines, maps, etc.). At a software level, *CollectiveAccess* is based on two main components: *Providence*, as the backend cataloging platform, and *Pawtucket*, as the frontend querying and publishing interface.

Turning to the more specific field of CH management, a very complete solution is *Arches*. It has been developed jointly by the *Getty Conservation Institute* and the *World Monuments Fund*, with broad participation of international heritage professionals. It is primarily intended for institutional use, with the purpose of creating digital inventories that describe the type, location, extent, cultural periods, material culture and preservation conditions of heritage resources, establishing relationships among them. It includes: a system to define, structure and manage data; several query and visualization tools to search, identify and represent entities (also at the spatial, temporal, and fuzzy level); a project and tasks management tool, to organize data editing establishing customized workflows. It supports the use of multiple ontologies, starting from *Cidoc-CRM*, and the definition of controlled vocabularies (**Carlisle** *et al.*, 2014). In addition, it follows the *Open Geospatial Consortium* standards, which means that the platform is compatible with desktop GIS applications. Since version 4, it offers the mobile application *Arches Collector*, intended for projects involving field data collection and editing. Recently, *Arches* core capabilities and underlying technologies have been extended to cover a wider range of use cases[15]. For example, through the “*Arches for Science*” initiative, whose first version is scheduled to be released by summer 2021, it aims to support the community of researchers in the field of heritage science and conservation.

3.3. Virtual Research Environments (VRE)

Table 3. Basic characteristics of selected VRE

Category	Software	Version analyzed	Updated on	Released on	Programming languages	Developer(s)	License
VRE	*Research Space*	3.4.0	20/02/20	2018	JavaScript, TypeScript, Java	*The British Museum*	GPLv3+
	Nodegoat	7.3	17/09/19	2013	PHP, JavaScript, C++	*Lab1100*	GPLv3

ResearchSpace is a VRE developed by *The British Museum* in collaboration with the German company *Metaphacts* and is currently in pre-release. It employs graph-based data modeling and semantic Web technologies to support CH research. It emphasizes the graphical representation of knowledge, rather than cataloging, and promotes collaborative research (**Oldman**; **Tanase**, 2018). It provides an integrated environment of features designed to reproduce research methods, including *Cidoc-CRM*-based semantic search, clipboard and semantic annotation, image manipulation and annotation tools using *IIIF* (*International Image Interoperability Framework*), and knowledge maps. Although it is still in an early stage of development, its use has been documented in several disciplines, such as: in Archaeology, the *Gravitate* project, focused on providing an IT solution to establish cultural or physical relationships among heritage artifacts; in Art History, the project *Late Hokusai*; in the History of Science, the *CorpusTracer* project (**Kräutli**; **Valleriani**, 2018).

Nodegoat shares with *ResearchSpace* an approach oriented to dynamic information visualization and to the generation of knowledge. It is a VRE developed since 2011 by *LAB1100*, a *University of Amsterdam* spin-off. It includes data management, network analysis and graphical visualization tools. *Nodegoat* is designed to develop WIS focused on research, rather than for content storage and preservation. It allows users to co-create datasets based on custom-designed models. In the same management environment, it offers modules to analyze and visualize these datasets relationally, diachronically, and spatially (**Bree**; **Kessels**, 2013). It has been used in various humanistic disciplines, for a wide variety of use cases[16], particularly in projects where the research questions were closely linked to combined analyses of space and time variables. A free, but limited, version of *Nodegoat* is available as SaaS, while the other usage options require a subscription of an annual service contract. The complete source code is available to download for free, but rather to ensure the transparency of the project than for independent use.

3.4. Crowdsourcing

Table 4. Basic characteristics of selected software packages for crowdsourcing

Category	Software	Version analyzed	Updated on	Released on	Programming languages	Developer(s)	License
Crowdsourcing	*Zooniverse*	online	20/12/20	2009	JavaScript, Ruby	*Fingerprint Digital Media*	*Apache 2.0*
	Pybossa	3.3.0	20/12/20	2013	Python	*Scifabric*	GPLv3+
	EnrichEuropeana	1.0	01/12/20	2019	PHP, Java, JavaScript	*EnrichEuropeana* project	GPLv2/EUPLV.1.2
	CrowdHeritage	1.0	05/12/20	2019	JavaScript	*AILS-Lab*	EUPL V.1.2

> In terms of target audience, at least two trends can be identified: software developed with institutional customers in mind, versus application frameworks suited more for individual projects, with custom structure and functionalities

Zooniverse is a portal arisen in 2009 from the collaborative project *Galaxy Zoo*. It is not limited to the humanities but embraces initiatives from any field of scientific research. It is based on open peer-to-peer collaboration and offers the possibility to create projects for data collection and manipulation involving society. It has a very large community of users, which in 2019 reached 1.6 million volunteers. Despite publishing much of its code in open source, it is intended to be used as a service through its web platform. It offers ad hoc tools for assigning tasks to participants, as well as the ability to upload datasets (usually images) on which volunteers are asked to perform actions; the structure of its data model cannot be adapted to specific cases (**Simpson** *et al.*, 2014).

Pybossa is an open-source framework for the development of platforms for data collection, analysis, and enrichment in a collaborative environment. It was launched in 2013 by the Spanish company *Scifabric*, which is still in charge of its development. It offers functionalities for the transcription of handwritten documents, video, and audio; analysis of multimedia objects; geolocation of files; identification and tagging of objects, faces, sounds, etc., in images, audio and video; identification of entities in digitized works, such as animals, people, sentiments, etc.; enrichment of records and files with metadata. Its area of application is not limited to citizen science projects, and many of its use cases come from the humanities. It is worth noting that the technology offered by *Pybossa* constitutes the base infrastructure of the *BNE Community* initiative, promoted in 2019 by the *National Library of Spain* in collaboration with *Red.es* (**Sánchez-Nogales**, 2019). *BNE Community* is an open and distributed platform for collaborative work on the digital collections of the *National Library of Spain*, similar to *Zooniverse*, but with two relevant differences: researchers cannot provide their own data; new projects can be suggested, but not created directly. Users can be assigned tasks such as object identification, text, and audio transcription, georeferencing, tagging and labeling, metadata enrichment and OCR correction. *BNElab*, which oversees its development, has added new functionalities to the core technology that has made available for reuse.

EnrichEuropeana is the name of a software package developed as part of the homonymous project, funded by *Europeana* and the *European Commission* between 2018 and 2019, and then again up to 2022 as *EnrichEuropeana+*, in order to renew and extend the technological infrastructure of the *Transcribathon* platform (**Sciotti**, 2019). Like other crowdsourcing platforms, it offers features for transcribing, annotating and geo-referencing historical documents, although it has a narrower scope of application than the others described above. At the technological level, it is composed of three modules: a frontend application, made up of a customized version of *WordPress* and an *IIIF* viewer; an API for semantic enrichment; a data transfer infrastructure for the exchange of information between *Europeana* and the *Transcribathon* platform.

The *CrowdHeritage* platform is the latest initiative promoted by *Europeana* in this field. It was created to use crowdsourcing to improve the quality of digital CH metadata indexed in *Europeana*, as well as in other databases of heritage institutions with public APIs. It was developed by the *National Technical University of Athens*, in collaboration with the *European Fashion Heritage Association*, the *Michael Culture Association*, the French *Ministry of Culture* and the *Europeana Foundation* (**Kaldeli** *et al.*, 2021). Unlike other software solutions in this field, the tools available are somewhat more restricted, including content annotation features, semantic markup through thesauri and vocabularies, color analysis and collaborative geo-tagging. User participation is promoted through gamification, using a leaderboard and awarding badges. All the code is available as open source and can be used to implement similar platforms, although this is not the aim of the initiative.

4. Comparative analysis

Once the description of the different selected applications has been completed, Tables 5-7 show a comparison of their main characteristics, divided into three categories:

- essential features;
- information structuring and retrieval;
- additional features.

They group up to 15 properties, which are usually necessary for the creation of WIS and are often required by the promoters and users of those systems.

The intrinsic risk in the elaboration of comparative tables is to oversimplify for the benefit of getting an overall view. In some cases, the quality and completeness in the implementation of a certain feature are very different from one software to another and to point out the simple presence/absence with a tick mark may be reductive. To partly compensate this limitation, we introduced two variants: "L" for limited implementation, which means partial or experimental; "E" for extension, where the functionality depends on an external module, which by itself does not describe its quality, but indicates its level of integration and perspective of maintenance.

Table 5. Comparison of the essential features of the analyzed platforms

Software	Basic formats*	Advanced formats**	Maps	Multi-language	Import / export	Extensible via modules
MediaWiki	X	E	E	X	X	X
Drupal 7	E	E	E	X	E	X
Drupal 8/9	X	E	E	X	E	X
BackdropCMS	X		L	X	X	X
WissKI	X	E	E	X***	X	X
Mukurtu	X	E	X	X***	X	X
Omeka Classic	X	E	L	X	X	X
Omeka S	X	E	L	X	X	X
AlchemyCMS	X			L		L
Islandora	X	X	X	X***	X	X
Arches	X	X	X	X	X	X
CollectiveAccess	X	X	X	X	X	
ResearchSpace	X		X		X	
Nodegoat	X		X		X	
Zooniverse	X		X	L	X	
Pybossa	X	X	X	L	X	X
EnrichEuropeana	X					
CrowdHeritage	X			X		

Legend: E = extension; L = limited.
* PDFs, images, audio, and video. ** 3D objects, collections, paginated content, binaries, KML, etc. *** Multilanguage support based on *Drupal*, but not available for the distribution interface.

We divided the management of different formats between basic and advanced, since practically all the selected applications include the former by default, while more advanced formats support is guaranteed only in the case of DAMs, at core level, and in most CMSs, through extensions. Mapping functionalities, which are of unquestionable importance for the study of the past, are covered in the vast majority of cases, although at times to a limited extent. A similar situation is observed for the multilingual function, which we have included among the essential features because of the relevance that multilingualism and linguistic diversity have recently acquired in the field of DH (**Budin**, 2015)[17].

The most evident shortcomings are found between VRE and crowdsourcing platforms, with the exception of *Pybossa*, and can be explained by the fact that they are solutions for a more specific audience and with a narrower focus, which limit their functionality to those considered necessary in their field of application and do not foresee much extensibility.

There is clearly a tension in data modelling between standardization, compatibility, and interoperability, on the one hand, and personalization and peculiarities, on the other hand, with valid arguments in favor of both strategies (**González-Pérez**, 2018). A representation of these opposing trends is shown in the first two columns of Table 6, where some pattern can be identified. *Drupal* and *BackdropCMS* clearly favor customization over the support for metadata standards, while *Omeka* and *Europeana*-based applications do the opposite; a good balance of both features is offered by several solutions, with some of them adding interoperability with linked data (**Davis**; **Heravi**, 2021).

Basic search is included in all platforms, and most of them provide also support for additional information retrieval functionalities, through the integration of an external specialized search engine. *Apache Solr* clearly stand out as the preferred solution, but alternatives are present. They all provide improved performance and good scalability for very large datasets. The lack of advanced search capabilities in the group of software devoted to crowdsourcing can be explained with the fact that, by design, they prefer users to browse content following a preset workflow, instead of free searching, to optimize the performing of the assigned tasks.

> Virtual Research Environments (VRE) are characterized by a high degree of customization in the design of the data model and in the definition of taxonomies, together with little to no possibility of adapting the functionalities

Table 6. Comparison of data modelling and information retrieval functionalities of the analyzed platforms

Software	Customizable data model	Metadata standards	Linked data	Basic search	Advanced search
MediaWiki	E	E	E	X	Elasticsearch
Drupal 7	X	L		X	Solr
Drupal 8/9	X	L		X	Solr
BackdropCMS	X			X	
WissKI	X	X	X	X	Solr
Mukurtu	X	X	X	X	Solr
Omeka Classic	L	X		X	
Omeka S	L	X	X	X	Solr
AlchemyCMS	L			X	Ferret
Islandora	X	X	X	Solr	Solr
Arches	X	X	X	X	Elasticsearch
CollectiveAccess	X	X		X	X
ResearchSpace	X	X	X	X	X
Nodegoat	X	X	X	X	X
Zooniverse				X	
Pybossa	L			X	
EnrichEuropeana		X	L	X	
CrowdHeritage		X	X	X	

Legend: E = extension; L = limited.

The last set of features selected for comparison are not essential for the implementation of a WIS, but offer additional valuable options. SaaS is available, sometimes as the default choice, for all software devoted to crowdsourcing. That is understandable as they respond to specific workflows that can be centrally managed and deployed in several instances, which differ in content, but not in functionalities or data structure.

The availability of an official mobile app should not be underestimated, both to improve the browsing experience of large datasets on mobile devices (tablets in particular) and for projects involving data collection tasks requiring access to mobile hardware (camera, GPS, gyroscope, etc.). Very few platforms have one available, but most of the rest provides REST API, so custom mobile application can be relatively easily developed.

Official demos available online are not exactly a platform feature and are less relevant where good documentation or SaaS are available but can help evaluating software that is more complex to install. Official support, with a combination of paid and free services, is basically available for all selected platforms.

At a comprehensive level, a distinction must be made between 1) agnostic platforms and 2) those developed for a specific domain.

1) Clearly belong to the first group: *MediaWiki*, *Drupal*, *Islandora*, *CollectiveAccess*, *AlchemyCMS*, *Zooniverse* and *Pybossa*.

2) Are designed, conversely, for a more specific area of application: *BackdropCMS* (small businesses and NGOs); *WissKI*, *Nodegoat* and *ResearchSpace* (research in DH); *Mukurtu* (indigenous communities' CH); *Omeka* (digital collections); *Arches* (CH assets inventories); *EnrichEuropeana* and *CrowdHeritage* (transcription and tagging of collections indexed in *Europeana*).

Another variable that significantly affects the characteristics of some platforms versus others is the target audience. At least two trends can be identified.

1) A first group of software (to which the following belong: *CollectiveAccess*, *Omeka S* and *Arches*) has been developed with institutional customers in mind, such as GLAM: they are oriented towards the creation of institutional repositories and advocates a clear preference towards data models based on standardized metadata, interoperability (linked data and APIs) and the adoption of common protocols.

> The recommendation on using existing packages rely on the preference in using common code to implement customized web information systems (WIS) instead of custom code for WIS that share common characteristics and functionalities

2) A second group is suited more to develop web solutions for individual projects, with custom structure and functionalities. Some software, such as *WissKI*, *Islandora* or *Nodegoat*, provide a hybrid approach, combining a good level of compatibility and interoperability with consolidated standards, such as *Dublin Core* or *Cidoc-CRM*, with the possibility to personalize the data model and develop custom ontologies on top of the common standards. Because of this increase in functionalities, the weight and complexity of the source code has also grown, as well as the number of external dependencies, and the current scenario now offers few options that are easy to install and with basic requirements in terms of web hosting. Among them we can name: *MediaWiki*, *Drupal* 7, *BackdropCMS*, *Omeka Classic* and *Mukurtu*.

In general, all analyzed solutions are very mature, with relatively long development trajectories (Tables 1-4). Only two, *ResearchSpace* and *CrowdHeritage*, are new applications that have appeared in the last five years, as *Omeka S*, *BackdropCMS* and *EnrichEuropeana* are evolutions of pre-existing platforms. The level of adoption varies widely and is difficult to compare, especially between agnostic and specialized applications.

Table 7. Comparison of additional features of the analyzed platforms

Software	SaaS	Mobile app	REST API	Official demo
MediaWiki			X	
Drupal 7			E	
Drupal 8/9			X	
BackdropCMS			E	X
WissKI			X	
Mukurtu		X		X
Omeka Classic		X	X	
Omeka S	X	X	X	
AlchemyCMS			X	X
Islandora			X	X
Arches		X		X
CollectiveAccess	X		X	X
ResearchSpace			L	L*
Nodegoat	X		X	X
Zooniverse	X	X		X
Pybossa	X		X	
EnrichEuropeana	L**		X	L**
CrowdHeritage	L**		X	L**

Legend: E = extension; L = limited.
* An official demo package is available, but it should be manually setup with docker-compose.
** Available as SaaS only for their respective sponsor projects; those instances can be also considered software demo.

Finally, it is necessary to make a last distinction between initiatives that publish their code as open source to favor its distribution, downloading and reuse and those that use open source mainly to ensure the transparency of the initiative and the long-term stability of the code. In the first case, the installation process is usually relatively simple or very well documented and, if the application requires external dependencies, they are relatively common and widely used in web development environments. This group includes most of the solutions analyzed in this review, such as: *MediaWiki*, *Drupal*, *BackdropCMS*, *WissKI*, *Mukurtu*, *Omeka*, *AlchemyCMS*, *Islandora*, *Arches*, *CollectiveAccess* and *ResearchSpace*. The second group, instead, includes applications that are usually offered as SaaS and that provide active support to their user community (i.e., *Zooniverse*, *EnrichEuropeana* and *CrowdHeritage*) or to the development of the projects they host (i.e., *Nodegoat* and *Pybossa*).

5. Limitations

A review of this kind would be incomplete without a section dedicated to limitations that affect these types of software packages, being the need of routine maintenance one of the most relevant. In fact, none of the software solutions analyzed here can be safely kept online without at least corrective and adaptive maintenance in place (**Sommerville**, 2005), which involves both technical and human resources.

Software packages[18], especially those accessible on the Web, are affected by the identification and disclosure of security flaws that can expose to large-scale automated attacks, which compromise even systems that are not direct targets of malicious code execution. Noteworthy, in this regard, are the cases known as *Drupalgeddon 1* and *2*, which respectively occurred in 2014 and in 2018[19]. They exposed vulnerabilities that made possible the automatic hacking, in a few hours, of virtually every *Drupal 7*-based web-platform not updated to the latest version and the creation of backdoors that were not eliminated by the subsequent software update. The best protection against the public disclosure, and the subsequent possible malicious exploitation, of security flaws is the prompt installation of released security patches, in particular those relating to core components, which in turn may create compatibility problems with other components of the system, thus requiring active maintenance support.

In addition to that, it should be considered that the Web is a technological space in continuous transformation, both server-side and client-side. This continuous evolution causes, from time to time, incompatibilities of the systems developed with the underlying web technologies. A representative example of this has been the evolution of the PHP language, by far the most popular language on the Web[20]. The transition from version 5, supported until 2018, to version 7, released in 2015, brought significant improvements, but caused incompatibilities with existing applications that require a manual update of the source code. When choosing the most appropriate software solution to develop a WIS, it is essential to consider the presence of a large and active community of users and developers, as the existence of the former creates the necessary conditions for the latter to act in the resolution of the incompatibilities that arise.

Centralized academic infrastructures, backed by institutional investments, are as appropriate to provide these maintenance services as they are scarce. Individual projects in particular, often promoted by a scholar to tackle a specific research topic, inevitably reach in the long run a point of completion, while it makes sense to keep the valuable resources produced available online. It is out of the scope of this paper to review the long-term solutions available, such as the conversion from dynamic platforms to static websites[21] or the use of flat–file CMS[22], but digital humanist promoting the development of WIS should be aware of these issues from the very beginning and design a plan on time.

6. Conclusions

After reviewing the characteristics of the selected applications, we can conclude that all the software solutions analyzed have their advantages and disadvantages, so it is necessary to get an updated knowledge of the current situation, to help choosing the most appropriate option to meet the requirements identified in each case. There are solutions that, due to their multifaceted characteristics, can cover a wider spectrum of needs and allow the deployment of WIS that can enable new research processes, not foreseen at the design stage (for example, in terms of sources rescue and verification, corpus enrichment, collaborative mapping, etc.).

As it stands today, *Drupal*, either on its own or through a bundled software package (e.g., *WissKI*, *Mukurtu* or *Islandora*), occupies a very specific niche and still represents one of the best options available for the application domains covered in this paper. Its degree of flexibility in terms of data model and incorporation of advanced functionalities through modules makes it comparable to fully customized solutions, but at the same time it provides a user-friendly frontend and is backed by a very large community of developers and users. It is suitable both to manage structured research datasets and to create large repositories of digital objects, in particular with user-provided content in collaborative environments. In the latest versions (since 8.5) it has also improved the management of multimedia content, traditionally weak. *WissKI* in particular, thanks to its semantic backend and its long development trajectory (2012-2021), is a very valid option for DH projects of a large scale. For smaller projects with reduced or no maintenance budget, the *BackdropCMS* alternative is very convenient and only remains the question about its long-term sustainability, considering the relatively low adoption rate to date.

Outside the *Drupal* galaxy, the most interesting products in the DH context are three.

- *Omeka*, which maintains in both versions a preferential approach towards the creation of digital collections with standard metadata.
- *Pybossa*, for Citizen Science projects that foresee the involvement of society through the assignment of pre-configured tasks.
- *Nodegoat*, for the broad spectrum of pre-configured functionalities ranging from custom data modeling to online publishing, linked data support, spatio-temporal visualization, and network analysis. This application belongs to a relatively new category of software solutions oriented towards the generation, integration, dissemination, and conservation of scientific knowledge in a collaborative environment (**Blake** *et al.*, 2010), which are the VRE. They are characterized by a high degree of customization in the design of the data model and in the definition of vocabularies/taxonomies, together with little to no possibility of adapting the functionalities (e.g., workflow, search options, views, maps, etc.). In fact, the latter are considered to be more generalizable and applicable tout court to different disciplines and research domains. An interesting prospect, particularly for larger projects, would result from the combined use of certified data repositories and VRE: the former being the place for long-term storage and exchange of versioned datasets and the latter the ecosystem for dynamic processing and collaborative metadata enrichment.

For a more institutional audience, we cannot avoid highlighting the *Arches* platform, which has also recently expanded its scope beyond CH management. It features a dedicated mobile application for field work, very robust geo-referencing and mapping functionalities, access permissions granular control, a comprehensive and flexible workflow management and maintains a well-defined development roadmap for the coming years[23].

An additional conclusion we can draw from this analysis is that the range of available solutions is sufficiently wide and mature to prefer, in general, their use versus the option of programming the WIS from scratch, combining a database management system (DBMS) with custom backend and frontend (i.e., the *LAMP*[24] model). On the one hand, this approach allows the highest degree of customization and provides a higher level of security against external attacks, which represent, as seen above, an important factor to take in consideration using pre-built packages. On the other hand, it requires more specialized skills for the implementation of the system, which results in a higher cost, in particular for the initial development, but also for the adaptive maintenance in the medium to long term. In addition to

> The range of available solutions is sufficiently wide and mature to prefer, in general, their use versus the option of programming the web information systems (WIS) from scratch

that, it exposes to a higher risk against the so-called bus factor, since the complete customization of a system entails a higher degree of dependence on the initial developer(s), conditioning the possibility of transferring the maintenance of the system to other people. Other additional costs derive from the need to produce and maintain documentation on the generated code and to provide in-house updates to ensure compatibility among core, external libraries, and server-side technologies, which inevitably will be updated over time. The principle behind our recommendation on using existing solutions rely on the preference in using common code to implement customized WIS instead of custom code for WIS that share, in many cases, common characteristics and functionalities.

7. Notes

1. This criterion led us to exclude, for example, the *Archaeological Recording Kit* (*ARK*): an open-source software package for data management in Archaeology. It is the underlying technology used in the *Fasti Online* platform:
http://www.fastionline.org

an online database of archaeological excavations, surveys and archaeological conservation projects. Despite this, its latest stable version (v.1.1):
https://sourceforge.net/projects/arkdb/files

has been released in 2014 and is not compatible with recommended PHP versions. Version 2.0 has been under development for some time, but its repository has remained unchanged for the last four years:
https://github.com/lparchaeology/ark2

2. Several individual sources have been consulted to identify projects and initiatives that employ the selected software to build WISs and digital repositories. Examples of large thematic collections available online can be accessed at:
https://www.semantic-MediaWiki.org/wiki/Projects_in_eHumanities_running_Semantic_MediaWiki
https://digibug.ugr.es/handle/10481/55656
https://doi.org/10.5281/zenodo.3893546
https://cohistoria.es/proyectos
https://red.knowmetrics.org/en/artifacts-2
https://rrchnm.org/what-we-do

3. SaaS is the acronym for Software as a Service.

4. Open-Source CMS market share:
https://www.opensourcecms.com/cms-market-share

5. Version 7 incorporates in core the *Content Creation Kit* (*CCK*) module, for the creation of customized data models; version 8 incorporates the Views module, for the assisted configuration of SQL queries.

6. There are currently 47,211 modules available, about half of which are in active maintenance and are compatible with versions higher than 7.
https://www.drupal.org/project/project_module

7. For example, modules such as: *Features*, *Feeds*, *Teichi*, *Date YMD - Pre Epoch, Partial Date Values* (*Partial Date* for *Drupal* 7), *Autocomplete Deluxe*, *Views data export*, etc.

8. In particular, the adoption of the *Symfony* framework, the *PHP Composer* package management system and the object-oriented rather than the procedural programming language.

9. Version 7 is still used by 64% of all websites developed in *Drupal*, while version 8, after six years, has reached only 29% of the total
https://w3techs.com/technologies/details/cm-drupal

10. *BackdropCMS* has a usage statistic of 1,706 websites:
https://backdropcms.org/project/usage/backdrop

11. Currently 533:
https://backdropcms.org/modules

although more than 70 popular modules have been incorporated into the core:
https://backdropcms.org/upgrade-from-drupal/features-added-core

and almost all of the 100 most used *Drupal* modules have already been migrated:
https://backdropcms.org/upgrade-from-drupal/top-100-drupal-7-modules

12. There are currently 1380 official distributions
https://www.drupal.org/project/project_distribution

13. *https://us12.campaign-archive.com/?u=0b0b343dc81346ec97cfc51d9&id=acbdafa77c*

14. Examples of usage in DH comes, for example, from Linguistics (**Cimiano** *et al.*, 2020).

15. The Solr indexing and search engine
https://lucene.apache.org/solr
although presented as a core component by the project itself, is still an optional module.

16. Documented use cases include mapping heritage at risk, monitoring the effects of sea level rise on coastal assets, archaeological research, management and publication of information about the provenance of art objects, etc. A selection of projects implementing *Arches* can be consulted at
https://www.archesproject.org/implementations-of-arches

17. *https://nodegoat.net/usecases*

18. *https://multilingualdh.org/en*

19. Custom solutions, programmed from scratch, are less subject, although not exempt, from the negative effects of the public disclosure of security flaws.

20. More information in this regard can be found on:
https://www.drupal.org/forum/newsletters/security-advisories-for-drupal-core/2014-10-15/sa-core-2014-005-drupal-core-sql
and
https://www.drupal.org/sa-core-2018-002

21. According to *W3Techs* data, PHP is used by 77.8% of all websites with a known server-side programming language:
https://w3techs.com/technologies/details/pl-php

22. For example, using *Jekyll*, a static website generator, in combination with *Wax* or *CollectionBuilder*.

23. Flat-file CMS, such as *Grav*, are platforms that requires no database and are based on text files.

24. *https://www.archesproject.org/roadmap*

25. *LAMP* (*Linux*, *Apache*, *MySQL/MariaDB*, *PHP/Perl/Python*) is an acronym used to define a software bundle widely used for web applications. Each letter stands for one of its four open-source components: *Linux* for the operating system; *Apache* HTTP Server; *MySQL* for the relational database management system; *PHP* (or *Perl* or *Python*) for the programming language.

8. References

Alcaraz-Martínez, Rubén (2012). "*Omeka*: exposicions virtuals i distribució de col·leccions digitals". *BiD: Textos universitaris de biblioteconomia i documentació*, v. 28.
https://raco.cat/index.php/BiD/article/view/256877

Alcaraz-Martínez, Rubén (2014). "*CollectiveAccess*, un sistema de gestión y difusión de colecciones de museos, archivos y bibliotecas". *BiD: Textos universitaris de biblioteconomia i documentació*, v. 33.
https://doi.org/10.1344/BiD2014.33.23

Blanke, Tobias; **Candela, Leonard**; **Hedges, Mark**; **Priddy, Mike**; **Simeoni, Fabio** (2010). "Deploying general-purpose virtual research environments for humanities research". *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, v. 368, n. 1925, pp. 3813-3828.
https://doi.org/10.1098/rsta.2010.0167

Bocanegra-Barbecho, Lidia (2020). "Ciencia ciudadana y memoria histórica: nuevas pers- pectivas historiográficas desde las humanidades digitales y la historia pública Digital". In: Caro, Jorge; Díaz-de-la-Fuente, Silvia; Ahedo, Virginia; Zurro, Débora; Madella, Marco; Galán, José-Manuel; Izquierdo, Luis R.; Santos, José-Ignacio; Del-Olmo, Ricardo. *Terra incognita: Libro blanco sobre transdisciplinariedad y nuevas formas de investigación en el Sistema Español de Ciencia y Tecnología*. PressBooks.
https://doi.org/10.5281/zenodo.4034177

Bocanegra-Barbecho, Lidia; **Toscano, Maurizio**; **Delgado-Anés, Lara** (2017). "Co-creación, participación y redes sociales para hacer historia. Ciencia con y para la sociedad". *Historia y comunicación social*, v. 22, n. 2, pp. 325-346.
https://doi.org/10.5209/HICS.57847

Bree, Pim; **Kessels, Geert** (2013). *Nodegoat: a web-based data management, network analysis & visualisation environment*.
https://nodegoat.net

Brown, William J.; **Malveau, Raphael C.**; **McCormick, Hays W.**; **Mowbray, Thomas J.** (1998). *AntiPatterns: refactoring software, architectures, and projects in crisis*. John Wiley & Sons, Ltd. ISBN: 978 0 471 19713 3

Budin, Gerhard (2015). "Digital humanities, language industry, and multilingualism: global networking and innovation in collaborative methods". In: *CIUTI-Forum 2014: Pooling academic excellence with entrepreneurship for new partnerships*.

Peter Lang, pp. 423-448. ISBN: 978 3 034315708

Burdick, Anne; **Drucker, Johanna**; **Lunenfeld, Peter**; **Presner, Todd**; **Schnapp, Jeffrey** (2013). *Digital_humanities*. Cambridge, MA: MIT Press. ISBN: 978 0 262528863

Candela, Leonardo; **Castelli, Donatella**; **Pagano, Pasquale** (2013). "Virtual research environments: an overview and a research agenda". *Data science journal*, v. 12, pp. GRDI75–GRDI81.
https://doi.org/10.2481/dsj.GRDI-013

Care (2020). *Corpus architecturae religiosae europeae (IV-X saec.)*.
https://care.huma-num.fr/care/index.php

Carlisle, Philip K.; **Avramides, Ioannis**; **Dalgity, Alison**; **Myers, David** (2014). "The Arches heritage inventory and management system: a standards-based approach to the management of cultural heritage information". *Cidoc (International Committee for Documentation of the International Council of Museums)*, v. 8.
https://www.wmf.org/publication/arches-heritage-inventory-and-management-system-standards-based-approach-management

Causer, Tim; **Wallace, Valerie** (2012). "Building a volunteer community: results and findings from transcribe bentham". *Digital humanities quarterly*, v. 6, n. 2.
http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html

Chevalier, Pascale; **Granjon, Ludovic**; **Leclercq, Éric**; **Millereux, Arnaud**; **Savonnet, Marinette**; **Sapin, Christian** (2012). "Base de données annotées et wiki pour la constitution du corpus numérique CARE". *Hortus artium medievalium*, v. 18, n. 1, pp. 27-35.
https://doi.org/10.1484/J.HAM.1.102782

Cimiano, Philipp; **Chiarcos, Christian**; **McCrae, John P.**; **Gracia, Jorge** (2020) "Linguistic linked data in digital humanities". *Linguistic linked data*. Cham: Springer.
https://doi.org/10.1007/978-3-030-30225-2_13

Davis, Edie; **Heravi, Bahareh** (2021). "Linked data and cultural heritage: a systematic review of participation, collaboration, and motivation". *Journal on computing and cultural heritage*, v. 14.
https://doi.org/10.1145/3429458

Dombrowski, Quinn (2016a). "*Drupal* and other content management systems". In: Crompton, Constance; Lane, Richard; Siemens, Ray. *Doing digital humanities. Practice, training, research*. London: Routledge, pp. 289-302.
https://doi.org/10.4324/9781315707860

Dombrowski, Quinn (2016b). *Drupal for humanists (coding for humanists)*. Texas A&M University Press. ISBN: 978 1 623494728

Dombrowski, Quinn (2018). *Enterprise tools and DH*. Stanford | Digital humanities.
https://digitalhumanities.stanford.edu/enterprise-tools-and-dh

Gallini, Stefania; **Noiret, Serge** (2011). "La historia digital en la era del web 2.0. Introducción al dossier historia digital". *Historia crítica*, v. 43, pp. 16-37.
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-16172011000100003

Gilgado, José M. (2014). *Avoiding the law of the instrument*. Josemdev.
https://josemdev.com/avoiding-the-law-of-the-instrument

González-Pérez, César (2018). *Information modelling for archaeology and anthropology. Software engineering principles for cultural heritage.* Springer.
https://doi.org/10.1007/978-3-319-72652-6

Gruber, Ethan (2010). *Building Omeka exhibits with Fedora repository content | Scholars' Lab*.
https://scholarslab.lib.virginia.edu/blog/building-omeka-exhibits-with-fedora-repository-content

Hedges, Mark; **Dunn, Stuart** (2018). *Academic crowdsourcing in the humanities*. Chandos Publishing. ISBN: 978 0 081009413

Hirschheim, Rudy; **Klein, Heinz** (2012). "A glorious and not-so-short history of the information systems field". *Journal of the association for information systems*, v. 13, n. 4.
https://doi.org/10.17705/1jais.00294

Jiménez, Àngels (2003). "Digital asset management: la gestión de la información multimedia en las organizaciones". *El profesional de la información*, v. 12, n. 6, pp. 452-461.

http://profesionaldelainformacion.com/contenidos/2003/noviembre/4.pdf

Kaldeli, Eirini; **Menis-Mastromichalakis, Orfeas**; **Bekiaris, Spyros**; **Ralli, Maria**; **Tzouvaras, Vassilis**; **Stamou, Giorgos** (2021). "*CrowdHeritage*: crowdsourcing for improving the quality of cultural heritage metadata". *Information*, v. 12, n. 2, 64.
https://doi.org/10.3390/info12020064

Kräutli, Florian; **Valleriani, Matteo** (2018). "CorpusTracer: A Cidoc database for tracing knowledge networks". *Digital scholarship in the humanities*, v. 33, n. 2, pp. 336-346.
https://doi.org/10.1093/llc/fqx047

Lenz, Kirsten; **Oberweis, Andreas** (1998). "Design of world wide web information systems". In: *Classification, data analysis, and data highways*, pp. 262-269. Berlin, Heidelberg: Springer.
https://doi.org/10.1007/978-3-642-72087-1_29

Martínez-Caro, José-Manuel; **Aledo-Hernández, Antonio-José**; **Guillén-Pérez, Antonio**; **Sánchez-Iborra, Ramón**; **Cano, María-Dolores** (2018). "A comparative study of web content management systems". *Information*, v. 9, n. 2, 27.
https://doi.org/10.3390/info9020027

Nishanbaev, Ikrom; **Champion, Erik**; **McMeekin, David** (2019). "A survey of geospatial semantic web for cultural heritage". *Heritage*, v. 2, n. 2, pp. 1471-1498.
https://doi.org/10.3390/heritage2020093

Noiret, Serge (2018). "Digital public history". In: Dean, David (ed.). *A companion to public history.* John Wiley & Sons, Ltd., pp. 111-124.
https://doi.org/10.1002/9781118508930.ch7

Oldman, Dominic; **Tanase, Diana** (2018). "Reshaping the knowledge graph by connecting researchers, data and practices in researchspace". In: Vrandečić, Denny; Bontcheva, Kalina; Suárez-Figueroa, Mari-Carmen; Presutti, Valentina; Celino, Irene; Sabou, Marta; Kaffee, Lucie-Aimée; Simperl, Elena. *The semantic web – ISWC 2018*. Springer, v. 11137, pp. 325-340.
https://doi.org/10.1007/978-3-030-00668-6_20

Oomen, Johan; **Aroyo, Lora** (2011). "Crowdsourcing in the cultural heritage domain: opportunities and challenges". *Proceedings of the 5th international conference on communities and technologies - C&T '11*, pp. 138-149.
https://doi.org/10.1145/2103354.2103373

Pons, Anaclet (2018). "El pasado fue analógico, el futuro es digital. Nuevas formas de escritura histórica". *Ayer*, v. 110, pp. 19-50.
https://revistaayer.com/sites/default/files/articulos/110-1-ayer110_HistDigital_APons_MEiroa.pdf

Ridge, Mia (2014). *Crowdsourcing our cultural heritage*. Ashgate: Routledge. ISBN: 978 1 472410221

Rodríguez-Ortega, Nuria (2018). "Cinco ejes para pensar las humanidades digitales como proyecto de un nuevo humanismo digital". *Artnodes*, v. 22.
https://doi.org/10.7238/a.v0i22.3263

Ruest, Nick; **Stapelfeldt, Kirsta** (2014). *Introduction to Islandora*. YorkSpace Repository.
http://hdl.handle.net/10315/28006

Rutherford, Eleanor; **Hemati, Wahed**; **Mehler, Alexander** (2018). "Corpus2Wiki: a mediawiki based annotation & visualisation tool for the digital humanities". In: Burghardt, Manuel; Müller-Birn, Claudia (eds.). INF-DH-2018. Bonn: Gesellschaft für Informatik e.V.
https://doi.org/10.18420/INFDH2018-08

Sánchez-Nogales, Elena (2019). "ComunidadBNE: el proyecto de enriquecimiento colaborativo de la Biblioteca Nacional de España". *Mi biblioteca: la revista del mundo bibliotecario*, v. 59, pp. 68-72.

Scheuermann, Leif; **Kroeze, Jan H. Hendrik** (2017). "Digital humanities and information systems: innovating two research traditions". *AMCIS 2017 Proceedings*. 3.
https://aisel.aisnet.org/amcis2017/PhilosophyIS/Presentations/3

Scholz, Martin; **Görz, Guenther** (2012). "*WissKI*: a virtual research environment for cultural heritage". *ECAI*, pp. 1017-1018.
https://doi.org/10.3233/978-1-61499-098-7-1017

Sciotti, Elisa (2019). "Il transcribathon: un nuovo approccio alle lettere manoscritte risalenti alla Grande Guerra". *DigItalia*, v. 2, pp. 116-122.
http://digitalia.sbn.it/article/view/2174/1503

Simpson, Robert; **Page, Kevin R.**; **De Roure, David** (2014). "*Zooniverse*: observing the world's largest citizen science

platform". *Proceedings of the 23rd international conference on world wide web*, pp. 1049-1054.
https://doi.org/10.1145/2567948.2579215

Sommerville, Ian (2005). *Ingeniería del software*. Pearson Educación SA. ISBN: 84 7829 074 5

Stapelfeldt, Kirsta; **Moses, Donald** (2013). "*Islandora* and TEI: current and emerging applications/approaches". *Journal of the text encoding initiative*, v. 5.
https://doi.org/10.4000/jtei.790

Terras, Melissa (2016). "Crowdsourcing in the digital humanities". In: Schreibman, Susan; Siemens, Ray; Unsworth, John (eds.). *A new companion to digital humanities*, 2nd ed. Wiley-Blackwell, pp. 420-439. ISBN: 978 1 118 68059 9

Toscano, Maurizio (2018). "Where the researcher cannot get: open platforms to collaborate with citizens on cultural heritage research data". In: Romero-Frías, Esteban; Bocanegra-Barbecho, Lidia (eds.). *Ciencias sociales y humanidades digitales aplicadas*. Granada: Universidad de Granada, pp. 538-561.
https://doi.org/10.5281/zenodo.1469337

Toscano, Maurizio; **Rabadán, Aroa**; **Ros, Salvador**; **González-Blanco, Elena** (2020). "Digital humanities in Spain: historical perspective and current scenario". *Profesional de la información*, v. 29, n. 6.
https://doi.org/10.3145/epi.2020.nov.01

Transcribe Bentham (2021). *Transcription desk*.
http://transcribe-bentham.ucl.ac.uk/td/Transcribe_Bentham

Velios, Athanasios; **Martin, Aurelie** (2017). "Off-the-shelf CRM with *Drupal*: a case study of documenting decorated papers". *International journal on digital libraries*, v. 18, n. 4, pp. 321-331.
https://doi.org/10.1007/s00799-016-0191-5

Whitten, Jeffrey L.; **Bentley, Lonnie D.**; **Dittman, Kevin C.** (2004). *Systems analysis and design methods*. McGraw-Hill Irwin. ISBN: 978 0 071215213

Wiberg, Andrew (2014). "Murkuntu: information retrieval system engineered for indigenous individuals and communities". In: *IFLA WLIC 2014*, 16-22 August 2014, Lyon, France.
http://library.ifla.org/id/eprint/922