Perspectives on sustainable dislocated digital research resources

Andrea Alessandro Gasparini¹, Tom Gheldof²

¹Department of Informatics, University of Oslo, Gaustadalléen 23 B, Oslo, N-0373, Norway
²Research Unit of Ancient History, KU Leuven, Leuven, Belgium

Abstract
In difficult times, researchers often react promptly and adapt to a new situation, still focusing on the core values. In the upheavals of technology, researchers in Digital Humanities have also created innovative solutions. However, significant issues are associated with sustainable ingredients. First, software often emerges and disappears, leaving valuable data in the wild or causing costs in maintaining technological competences in the organization. Second, storing databases that have reached the end-of-life stage remains an unresolved issue. Third, knowledge and data can be closed inside a database and are often inaccessible to all. Finally, Digital Humanities must play a crucial role as a supporter of a sustainable world by rethinking its approaches. This paper uses ENCODE (an Erasmus+ funded project to bridge the gap between training and digital competences in the teaching/learning domain of ancient writing cultures) as a case study to examine some of the aforementioned issues.

Keywords
Digital Humanities, Sustainability, SDG, Systems Thinking

1. Introduction
Deconstructing and reconstructing knowledge are two of the major forces behind Digital Humanities (DH). Annotating old text to enrich forgotten languages, analyzing large bodies of images, or using artificial intelligence (AI) to predict missing parts of documents are a few of the various activities and technologies researchers are supporting today in the humanities in pursuing their tasks. During the last two decades, some radical changes have altered the technical infrastructure in which Digital Humanities may evolve. First, a fast and easy-to-connect network infrastructure allows anyone to contribute to digitization projects almost anywhere. Second, cheap computing resources are available at various sites, allowing the overproduction of projects and data. During this period, funding was easy to access, and a variety of small to very large projects started in academia and were occasionally supported by external software companies. As expected, this created a fragmented and complex landscape with several drawbacks [1]. For instance, it is very expensive to keep various projects alive, or
data disappear [2]. Digital sustainability in DH addresses all of these aspects, from software, data models, and interfaces to long-term access and security issues [3]. For instance, a well-designed service interface with a rich and user-friendly visualization of knowledge will ensure and maximize the use of DH tools. In addition, when visualizing datasets, the “Interfaces become performative environments where scholars can play with the data and build their own interpretations.” [4] Even though several activities in the context of digital sustainability do exist, what this paper emphasizes is an emerging, new, and critical role Digital Humanities need to expand and mature into: a supporter of a sustainable world by rethinking their approaches.

2. Digital Humanities for a sustainable world

The United Nations defined 17 Sustainable Development Goals (sdgs.un.org), covering issues, needs, and perspectives on poverty, gender, education, health, energy, consumption, production, peace, and climate change. In particular, problems concerning education, climate change, and energy consumption have implications for DH. Thus, the first question is how and why DH should consider the Sustainable Development Goals (SDG). What follows addresses the goals of developing long-term solutions for DH taking into consideration various SDGs and their relationship. Moreover, this paper will investigate DMP, FAIR, and Open Science [2] from a DH and SDG perspective.

2.1. Sustainability and software

DH projects use several technological infrastructures, including web software (e.g., front-end), sometimes a content management system, or a database platform (e.g., back-end). In addition, templates are necessary to register metadata in the system. The template is often a crucial part of the knowledge creation process, as it addresses the understanding and value of the data. When migrating to new platforms, reproducing the design process of building and testing the representation of data is quite difficult, and the lack of a “thick description” is often mentioned as a major risk for DH [5]. Preferably the interface has also been designed appropriately and tested with a User Experience (UX) approach.

All of the above can be developed in-house or by using off-the-shelf software. Nevertheless, all technological infrastructures have a life cycle [6], as software needs to be upgraded and, in the end, migrated to new ones. In addition, companies providing services may merge or disappear. Contextualizing the life cycle of technological infrastructures for DH in the academic world, the narrative is quite interesting. The anecdotal stories of professors with little funding, developing interesting services on computers under the desk in their offices, are often true and do not end well for the service.

Other issues, especially relevant for small projects, include security, as hacking into systems happens quite often and may destroy all the valuable data stored in the system. On the other hand, having goals of developing long-term solutions for DH projects already from the start, supports several SDGs. DH projects can foster innovation (SDG 9) when created in a professional environment. They support education for all (SDG 4) and reduce inequality (SDG 10) by supporting access to all the knowledge in the databases using, for instance, an Open Access approach when developing solutions for DH.
2.2. Sustainability and data

Data are another asset in DH, as they represent the work done during the research analysis, tuning, and cleaning of raw findings. Unfortunately, data are also stakeholders in DH, with several complicated aspects. Time is problematic for data stored in databases: first, occasionally one needs to migrate to new formats, and the long-term preservation of data needs to consider how various formats, such as XML (text) or PDF, resist migration needs [2]. Second, backups can easily become corrupted by large amounts of data [7]. From an SDG perspective, data are the tendo Achilles of DH projects. Both SDG 13 (climate action) and SDG 7 (arguing for affordable and clean energy) underpin the problematic aspects of producing, storing, and using data in DH activities. Energy consumption (SDG 7) is an emerging perspective on sustainability in digital humanities.

This is because storing and using data requires large amounts of power in data centers worldwide, but is invisible to users. For instance, images on websites, images, PDF, and other files are presented after a search, then may be downloaded locally, and finally stored in personal data houses (such as Google Drive, OneDrive, etc.), exponentially increasing the storage and transportation needs of the same data [8].

Another side effect of using energy and applicable to DH projects is called the 'rebound effect', where information of almost unlimited access to data storage increases the use, even the costs of consuming more energy affects the climate [9]. Finally, Digital Humanities projects should aim to reduce inequality (SDG 10). For instance, knowledge and data can be closed in a database and are often inaccessible to all. Therefore, researchers should be obliged to maximize the use of their data using the FAIR principles [10].

2.3. Sustainability and knowledge creation

The understanding of where and how new knowledge is created when the humanities interact with the digital world is understudied. When scholars started to develop DH services, they handed over the responsibilities to IT departments resulting in services "that have not adequately addressed the epistemological trajectories being designed into their technological infrastructures." [11]

Moreover, the making (practices) and thinking (theory) when developing DH services need to be addressed and designed accordingly [12]. Therefore, humanities scholars need to be aware of their role as developers and designers, as the knowledge inscribed in the interface and the service in general is relevant to how the content is used. From a sustainability perspective, the above addresses access to knowledge for all (SDG 4), where gender-based bias and other issues regarding inequalities based on race, ethnicity, and income (SDG 9) need to be considered.

In addition, the "outputs" of digital research projects in the humanities have specific values [13]. First, value is given by the understanding that users give to the output. Second, the preservation of the outputs over time includes, as mentioned, the design of the interface, especially how a search is developed. The latter is unfortunately volatile and undervalued.
3. ENCODE

In recent years, digitization of ancient written objects has become increasingly possible and sustainable because of the rapid development of e-infrastructures, digital infrastructures that provide services and tools in virtual (and collaborative) environments. In the framework of ENCODE, a three-year (September 1, 2020 - August 31, 2023) Erasmus+ Strategic partnership for higher education, partners from 6 European universities (Alma Mater Studiorum Università di Bologna, Julius Maximilian Universität Würzburg, KU Leuven, Università degli studi di Parma, Universität Hamburg, and Universitetet i Oslo) joined forces. They have - and are still - undertaking several steps towards bridging the existing gap in the teaching/learning domain of ancient writing cultures between the peculiar humanistic training and the now essential digital competences required for study, research, and employment.

3.1. ENCODE and Sustainability

The project’s sustainability is guaranteed by compliance with several best practices. In the conception phase, the project was described in detail, using the template of the Erasmus+ funding application. In this application, 6 Intellectual Outputs were outlined, including the organization of multiplier events, conferences, and training events, hosted by each of the partner institutions. Although sustainability - a word that occurs 19 times in the proposal - was not directly linked to the SDGs, several outputs tried to overcome some of the goals and challenges presented in the document from the United Nations. SDG 4 (Quality Education) and SDG 9 (Industry, Innovation and Infrastructure) are implicitly included in some of the project’s outputs, such as the framework of digital competences for students and teachers dealing with written cultural heritage or the design of innovative and customizable teaching modules for an ENCODE database, guidelines, and an online course. Moreover, the teaching methodology focuses on mutual learning among trainees and trainers, allowing replicability of activities by producing models of training sessions with different ENCODE modules and making self-training materials accessible to academics and other researchers [14].

Additionally, the project aims to adhere to the FAIR principles [10]. The ENCODE outputs such as the database and the online course should be findable via different platforms. Both outputs ensure the accessibility of the teaching modules and training materials by offering them freely available in open access and with an appropriate license. Such a license (CC BY or CC0) also guarantees the uptake and recycling of the modules, especially the training materials, making it possible for anyone to modify them according to their own needs [15]. Finally, by complying with the main standards in the field of ancient written materials (EpiDoc, Linked Open Data, etc.), it is possible to use all the ENCODE outputs in a collaborative manner and link them to most existing projects dealing with the study of ancient texts.

3.2. The ENCODE online course on #dariahTeach

One of the main outputs consists of the creation of a MOOC or online course, a collaborative platform on which introductions for teaching staff, researchers and other users will be brought together with the training materials that are produced for the different ENCODE training
events. A key concern of the partners was the long-term sustainability of such a digital platform. Whereas in a first phase, the partners explored the possibilities of setting up their own platform (e.g., Blackboard, Coursera, etc.) or choose an existing commercial initiative (such as Udacity, edX or Google Classroom), regarding the long-term preservation of the materials, it was decided to team up with a partner from another European consortium, DARIAH.

The #dariahTeach platform, developed using the open-source software Moodle, provides academics and other partners from the GLAM sector with the opportunity to create a free course. In contrast to an actual MOOC, these online courses are more modular and do not have a fixed start date. Modular course design is an approach to course development in which course materials are organized into smaller, self-contained units that can be rearranged, updated, or replaced, as needed.

This approach can facilitate updates and revisions of the content over time, allowing for the incorporation of new research findings and ensuring that the course remains current and relevant. This can help ensure that the ENCODE course remains up-to-date, and provides learners with the most current and accurate information. Partnerships and collaborations with other organizations, such as DARIAH, but also domain-specific partners, such as the digital classicists community, can help promote the sustainability of such online courses by ensuring their long-term impact and availability. By exploring such partnerships and collaborations with other organizations, course creators can identify new opportunities for sustainability and impact, ensuring that the course remains relevant and valuable over the long term [16].

3.3. Challenges and issues

During the project’s lifecycle, the ENCODE partners faced several challenges and issues regarding the sustainability of the different outputs. When, at the beginning, opting to host dislocated digital resources, the ENCODE team considered many different solutions. Regarding a sustainable approach, especially in trying to comply with the FAIR principles and best practices from Open Science, we decided to:

- Reduce the risk of obsolescence: the online course is hosted with an open-source technology such as Moodle and maintained by the #dariahTeach project. Partners develop course materials, in close collaboration with the research community. The ENCODE guidelines on different digital competences in the field of ancient writing cultures will be published in an open GitHub repository and in collaboration with the digital classicists community.
- Improve interoperability: ENCODE uses open-source technologies (e.g., eXist-db for the database platform) and standards compatible with other existing platforms and projects. By using and stimulating Linked Open Data, it is possible to re-use and modify the ENCODE training materials according to each user’s needs.

However, some challenges have not been sufficiently addressed or have occurred during the project’s lifecycle and are still being investigated:

- Reduce environmental impact: in addition to the use of open-source technologies and standards, ENCODE was and still is dependent on the support of the technological services from each of the partner’s institutions. In many of the partner universities, reducing the
environmental impact of digital research infrastructures is on the long-term agenda, but not a top priority for the short term.

- Hosting and the long-term preservation of dislocated digital resources: one of the partners (Hamburg University) engaged from the beginning of the project to self-host the ENCODE database with the teaching modules, but due to a career change from the local expert, Hamburg University even decided to completely opt-out of the Erasmus+ project, leaving us with the challenge to find a new, long-term hosting plan for the ENCODE database. For the other outputs, it was decided to collaborate with and trust existing platforms, such as GitHub as a repository for the guidelines, institutional repositories for the project’s publications, and #dariahTeach for the online course.

Most of these challenges and issues had already been considered at the start of the project and described in the project proposal, although some decisions had to be reconsidered when facing the issues described above. This proves again how, in the lifecycle of (academic) research projects also sustainability is an aspect that requires sufficient consideration beforehand, but also needs to be re-investigated along the way. The challenges of creating dislocated digital resources in the ENCODE project were (and are still) addressed during the monthly transnational partner meetings and in close collaboration with external partners.

4. Mapping Digital Humanities and sustainability

Understanding sustainability in Digital Humanities is complex. Researchers, students, IT and university staff, and the public are all stakeholders, as well as the IT infrastructure, including the software, the project with data, and how it is organized. One way to approach all complex connections and dependencies is to adopt a holistic approach rather than focusing on each separate part. Systems Thinking is one way to create an overview as it addresses systems as “interrelated, or interdependent parts that form a complex and unified whole that has a specific purpose” [17]. In addition, Systems Thinking emphasizes an understanding of systems without the aim of solving inherent problems [18]. Therefore, the integrative properties of systems thinking fit to approach complex problems, as is often the case in DH projects, with multiple stakeholders and divergent perspectives, without complicating the process. In addition, the richness of stakeholders in DH projects provides an arena where reorganization and conservation of tensions between the activities is the rule. For example, the ENCODE project encountered a situation where the IT staff in charge of the database service changed workplaces, creating an unstable situation with several downtime periods. In Systems Thinking, these tensions are called self-organization [19]. Self-organization is when a system acquires its own “identity”, where “coherent patterns of relationships are internally structured and develop over time.” [19]

4.1. Iterative Inquiry

As a method for mapping out and analyzing the structure and processes of complex systems, the authors opted to use the design-based method “iterative inquiry” [20, 18, 21]. This method explains the different levels in a social system and how all the components are connected and interrelated at the micro - meso - macro level. For instance, for the ENCODE project, the micro
level includes the local staff of the project, while the meso level consists of the university and the national epigraphic (or other domain-specific) community. Finally, at the macro level are the international communities and the public at large. However, for instance, at the meso level, the ENCODE project also has a role of sharing and disseminating ancient writing cultures. To map out the entire system, the approach uses several nested undersystems, such as Function, Structure, Process, and Structure/Purpose. More specific:

- **Function** = the action activating the sequences in the system
- **Structure** = stakeholders and their relationship
- **Process** = activities
- **Structure/Purpose** = unique environment

### 4.2. Using “Iterative Inquiry” to analyse the ENCODE project

Using the web-based visual platform Miro to visualize the output of the “iterative Inquiry” process, the entire system design of the ENCODE project, it was possible to discover various invisible aspects. Figure 1 shows how the linearity of the ENCODE project, typical for a DH project, has changed. The circular understanding of the different levels of the project and how they are connected and interrelated visualizes the dependencies. What the iterative inquiry shows is an awareness of each component at the different levels. Moreover, when observing how the project functions between the micro and the macro level, part of the less relevant technical aspects disappear.

Another aspect emerging from the iterative inquiry emphasizes the difficulty of materializing the SDGs in DH projects. One experience of the ENCODE project is the rationale for addressing these perspectives from the very beginning. As mentioned earlier, the ENCODE project considers SDG 4 (digital illiteracy) and SDG 13 (climate action) regarding open access, as they were well addressed in the original proposal. In addition, several UX tests are planned to guarantee SDG 10 (inequality) and SDG 4 (digital illiteracy).

On the other hand, the energy crisis, SDG 7, was not easy to react to, as the project does not have a direct connection to service providers, for instance, where and how videos are stored and accessed. This section outlines an iterative inquiry that has not yet been used in the ENCODE project. However, it has the potential to guide the project in assessing the SDG related to energy use, as it involves actions on multiple levels.

### 5. Discussion

Digital sustainability has become a crucial element in the development of new Digital Humanities projects. The FAIR and Open Access movement does help address some of the issues of long term access to all users, and directly underpins the necessity to react to inequality (SDG 10) and digital illiteracy (SDG 4). Both digital sustainability and the role DH plays in supporting the goal of a sustainable world need to be seen and used simultaneously in DH projects. A call to rethink DH approaches used when starting a project was chosen for the ENCODE project, resulting in a dislocated placement of resources. The use of the iterative inquiry tool to create an overview and understanding of the ENCODE project (see Figure 1) is valuable. The Systems
Thinking approach and the tool can be used to examine tensions, such as the goal of the project versus sustainability or long-term preservation of data versus presenting earlier results of a project. These new scopes have changed the linearity of the ENCODE project toward a circular understanding of the different levels.

The emerging dependencies at the higher level (macro) of the iterative inquiry of the technical aspects of the project have mostly disappeared. This is an aspect that the ENCODE project is currently working on. In other DH projects, the focus is often on specific phases, such as results and findings and writing articles. While other aspects are often left behind, one example is the lifecycle of DH projects. This highlights the complexity of DH projects, from an advanced network of online services and infrastructures to proprietary project metadata. Therefore, the ENCODE project has opted for dislocated platforms such as GitHub for the Guidelines, and #dariahTeach for the MOOC online course.
Finally, energy consumption and other material usage when storing data or running servers for applications are affected by DH and need to be scrutinized with the SDGs in mind. For instance, SDG 7, affordable and clean energy, is based on a forecast that ICT will demand 21% of all energy produced in the world by 2030 (sdgs.un.org). When planning DH projects, the use of energy in data centers, which includes cloud computing, big data analysis, and Artificial Intelligence, also needs to consider the effects on resource extraction, manufacturing, transport, use, and end-of-life of ICTs. As mentioned before, how images, pdf, and other documents are stored and presented online may have an impact on energy consumption and indirectly on CO2 emissions (SDG13: Climate Action).

6. Guidelines

The following guidelines, supporting sustainable digital research resources, are the results of the ENCODE project. However, the last three emerged during the final phase of the project.

- Reuse of data (create a Data Management Plan)
- Public and easy access for researchers (Open Science)
- Support Linked Open Data
- Use the User experience (UX) approach so your data is user-friendly, and new knowledge is created and stored in the interface
- Use UD – Universal Design (e.g., for the visually impaired) access to all
- Decrease power usage by using links to images. Don’t copy them!
- Be aware that we must accept the scientific value of creating an open-access database!

7. Conclusion

This paper presents how the European project ENCODE, under the ERASMUS+ umbrella, has focused on sustainable research infrastructure. Approaching a project with a sustainable, dislocated digital resources mindset is possible. From the beginning, the sustainable aspect and the 17 Sustainable Development Goals (defined by the United Nations) have been addressed when planning the activities and as part of workshops and seminars that are part of the project’s outputs. These outcomes include guidelines to support sustainable digital research resources.

Acknowledgments

The work presented here is part of the Erasmus+ ENCODE project. The authors wish to thank all the project partners in ENCODE and are also grateful to the sustainabilitylab.uio.no for their support.
References


