

Best Practices in the Implementation of Research Infrastructure in the Academic Environment: Shortcomings and Revisions

Michal Lorenz*, Ema Juranová**, Michal Konečný***,
Hana Kubelková****, Veronika Wolfelová*****

ARTICLE INFO

Article history:

Received 28 July 2022

Revised 27 January 2023

Accepted 07 February 2023

Keywords:

Research Infrastructure,
Digital Platforms,
Curation Workflow,
Research Data,
Research Environment,
Best Practice

ABSTRACT

The Digitalia MUNI ARTS - a local node of the LINDAT/CLARIAH-CZ research infrastructure at the Faculty of Arts of Masaryk University constitutes a repository in the Islandora system. It is used for long-term preservation of research data together with their research environment in the form of digital platforms. We transfer the digital outputs of humanities scholars research to the repository according to a set plan, which is based on best practice recommendations for project management and digital curation. In this paper, we present how the results of interviews with platform developers and infrastructure stakeholders translate into the curation workflow, and a resulting model for migrating digital platforms to the repository. Reflecting on three types of problems we encountered during the implementation of platforms into the repository - communication problems, problems of external dependence, and management problems - we describe a modification of the migration process. We present six recommendations for repository administrators and curators in an academic setting - holding an introductory meeting with developers, researching significant and relevant theories of knowledge domain, consulting license experts, prioritizing requirements, and preparing handover protocol and progress reports.

-
- * Assistant professor, Department of Information and Library Studies, Faculty of Arts, Masaryk University, Czech Republic (lorenz@mail.muni.cz) (First author)
** Digital Curator, Centre for Information Technologies, Faculty of Arts, Masaryk University, Czech Republic (119876@mail.muni.cz) (Co-author)
*** Digital curation consultant, Centre for Information Technologies, Faculty of Arts, Masaryk University, Czech Republic (michal@michalkonecny.com) (Co-author)
**** Project manager, Centre for Information Technologies, Faculty of Arts, Masaryk University, Czech Republic (hkubelkova@mail.muni.cz) (Co-author)
***** Digital humanities specialist, Centre for Information Technologies, Faculty of Arts, Masaryk University, Czech Republic (475579@muni.cz) (Author of visualizations)
- International Journal of Knowledge Content Development & Technology, 13(4): 095-117, 2023.
<http://dx.doi.org/10.5865/IJKCT.2023.13.4.095>
-

1. Introduction: Digital Projects at the Faculty of Arts MUNI

The first digital projects at the Faculty of Arts, Masaryk University (MUNI) were created thirty years ago and the increase in their numbers can be observed even before the Velvet Revolution in 1989. The interest of experts focused on various applications of machine text recognition, computer typesetting, statistical tools, and, in particular, tools usable in computational linguistics. A database with reports on research projects using computing technology in humanities was created as the first digital platform, along with a platform for regular meetings of experts interested in the use of computing technology in the field of humanities (Rambousek, 2000). Over a period of 30 years, the world of computing technology has changed significantly, the use of computers has expanded and it has transformed not only education, but also research including the research topics. During this period, a number of digital projects financed by grant agencies or individual departments and service units of MUNI were created at the Faculty of Arts. Despite the initial cooperation of experts, who were using the capabilities of computing technology to approach the issues in humanities, no cohesive community has been established at the Faculty of Arts MUNI. The ubiquitous use of ICT in teaching, administration and research has led to divergent developments. Although an IT support workplace has become established, the common identity among the humanities scholars is disappearing, being primarily determined by the scientific discipline instead of the application of computing methods to resolving the scientific issues of humanities. Digital platforms - the results of research projects, such as databases, geographical information systems, and other digital outputs - are not registered anywhere. They often serve to only partial goals of studies in their individual disciplines, regardless of their long-term sustainability and reuse. Awareness of the work of other experts using computing technologies as well as of the standards of the web and information environment, the role of metadata and the identity of the digital humanities in general, is disappearing. In this situation, many digital products of research activities remain unmaintained and unupdated, hidden or poorly available to a wider range of potential interested parties, unused and obsolete, despite the efforts of experts and considerable financial investments. We consider such development to be undesirable. The possibility of change has come with the engagement in the construction of a large research infrastructure LINDAT/CLARIAH-CZ.

The following paper summarizes our previous experience with building a local node of the large national research infrastructure over the 3 years of the project at the Faculty of Arts MUNI, which we would like to share with other experts dealing with or planning to approach similar tasks. In the paper, we first present in more detail the environment which the repository was implemented in, the building block of our local infrastructure node, and our plans and ideas about the implementation of the research infrastructure, which were based on the available manuals of digital curation and good practice in building repositories. Then we confront them with the obstacles and problems that we have encountered in practice. The conclusion of the paper is devoted to the chosen method of resolving problems, procedures applied to prevent them and recommendations extending the already codified best practice in the implementation of similar projects, which offers lessons learnt for all future creators of digital repositories and infrastructures in humanities.

2. LINDAT/CLARIAH-CZ – large research infrastructure

There are a wide range of literature on the implementation of institutional repositories. Asadi, Abdullah, Yah, & Nazir divide their research topics into six categories: deployment, implementation, adoption; benefits and challenges of institutional repositories; development, content management, and policy; user behavior; research frameworks and conceptual models; and integration (2019). Most studies, over 27%, addressed the first category, issues related to repository design. The experience of implementing a repository in DSpace is described by J. Barwick (2007). The main problem in implementing the repository at the university was the restrictive copyright of publishers and the identification of versions of articles suitable for publication in the repository, which eventually led to the recommendation not to grant copyright to publishers. A. Miller (2017) discusses a partnership approach to academics that not only helps to increase the amount of content made available in the repository, but also to build a community of practice. He recommends that legal and ethical issues with copyright should be prevented at the evaluation stage of projects. Coughlin (2022) describes the process of implementing and maintaining a university institutional repository in the Samvera system, focusing particularly on the transition to a new solution based on user needs. He highlights how difficult it is to select the appropriate infrastructure to support the platforms as it is difficult to predict in advance how they will be used. Sweeper and Ramsden (2020) in their study emphasize the role of stakeholders in helping to extend the reach of the repository as well as working with faculty to engage them in using the repository, as this supports the integration of the role of the repository in the university's strategic plans. Stein (2021) also emphasizes the need to reflect the views and needs of faculty and students who will use the repository in decision-making processes. Li and Ronghui recommend building a service-oriented repository by linking the repository to the scientific research management (2019). The COAR working group, formed in 2016 to explore the next generation of repositories, presents a vision of "repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication, on which value-added service layers will be deployed, transforming the system to be more research-oriented, open to and supportive of innovation, and collectively managed by the scholarly community" (COAR, 2017). Our approach to implementing an institutional repository meets this vision. Repositories in most institutions are used to make the scholarly output of faculty members available in the form of articles and educational materials. A few repositories provide space for research data. In our paper, we describe a repository that provides access to platforms used by scientists to apply data in the research process. The resulting architecture is not oriented towards homogeneous groups of institutional employees but is tailored to different communities and scientific domains in order to assist scientific research and to foster collaborations with other scientists outside the institution. Our goal is to partner with scholars and colleagues at the university in research, offering them a repository for their structured research data, digitized cultural heritage objects, and the tools and processes they use to analyse them. Rather than focusing on the general characteristics of repository implementation, we focus in detail on decision-making process in every step involved in converting digital platforms into a repository.

The primary objective of the LINDAT/CLARIAH-CZ infrastructure is to make resources and

data in the fields of arts and humanities available to researchers, students, and, thanks to the emphasis on open access, also to the public or companies from the industry. As one of the partners in these efforts, the Faculty of Arts MUNI contributes in several areas (see Fig. 1).

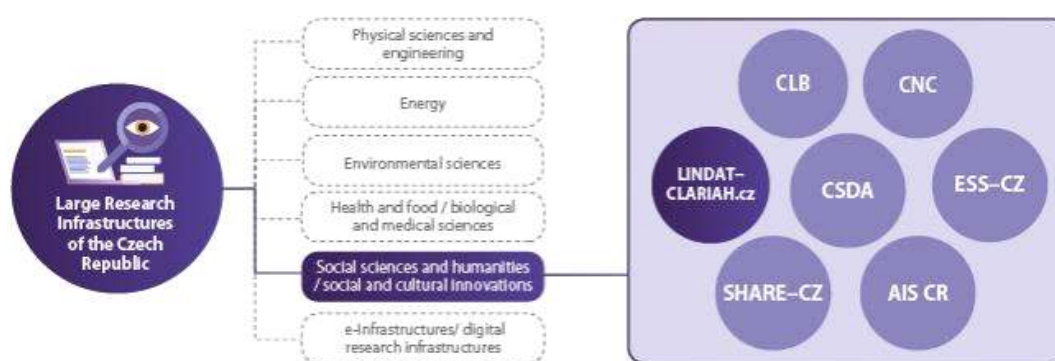


Fig. 1. LINDAT/CLARIAH-CZ among large research infrastructures

The most important part is building a trustworthy repository for the long-term storage of digital platforms created at the Faculty of Arts MUNI. It is called Digitalia MUNI ARTS (<https://digitalia.phil.muni.cz/en>). We transfer metadata of the stored platforms to the central repository of the LINDAT/CLARIAH-CZ research infrastructure (<https://lindat.cz>) in accordance with the adopted metadata schema, which also supports the visibility of implemented and emerging research projects, data reuse and cooperation with partners at home and abroad. We also provide support and consulting for academics who develop their own digital platforms. For our purposes, we understand a digital platform as an application framework that specifies the basic structure and functions of the system by both hardware and software means in such a way that the resulting architecture is modular, programmable, and interoperable with other applications, technologies, and services (Bogost - Montfort, 2007; Baldwin - Woodard, 2009). The digital platform serves as a virtual research environment for a user community in a given domain. When moving digital platforms and databases to the Digitalia MUNI ARTS infrastructure repository, we adapt the individual interface of each platform to the requirements of their creators and the needs of researchers, focusing on usefulness and usability with the secondary goal of enabling needs and infrastructure users modeling. In the transfer, metadata is also enriched in order to make the platforms more effective in research. The unrealized goal so far is data enrichment, supporting the use of digital research methods to find answers to long-term as well as current questions in humanities and social sciences. For this purpose, we plan to use tools supporting digital research, e.g. using the NameTag tool (Straka & Straková, 2014) for the recognition of named entities in selected text platforms. By combining services and data sources, infrastructure becomes an integral part of research. Technology and its configuration delimit the questions that can be subsequently asked by means of them. Thus, the infrastructure becomes the way how to ask questions and its structure identifies what is [storable], while the identification determines its structure at the same time (Derrida, 1996). However, formulating meaningful questions depends on experts, as does the resulting interpretation of the obtained data. Another goal is to support the consolidation and growth

of the community of digital humanities scholars at the Faculty of Arts MUNI and the promotion of digital humanities. In the following chapters, we will focus on the building of the local infrastructure for digital platforms at the Faculty of Arts - Digitalia MUNI ARTS (see Fig. 2).

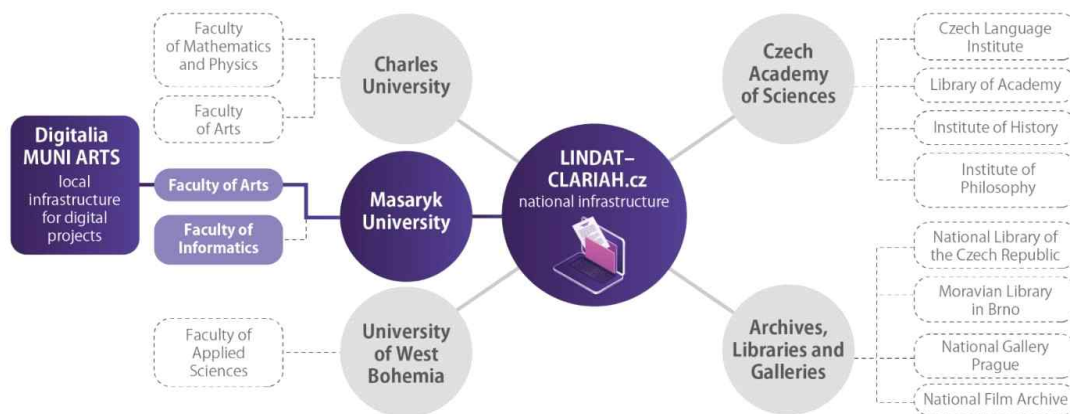


Fig. 2. Situating Digitalia MUNI ARTS in LINDAT/CLARIAH-CZ

3. Building the infrastructure

3.1 *The infrastructure environment*

With its 23 departments, the Faculty of Arts MUNI is one of the largest faculties in the Czech Republic. It is a complex and diverse environment which provides a variety of interesting projects including digital products created in some of them. Some of these products are created by academics on their own, literally left to their own devices, others are created in cooperation with a programmer, often only outsourced externally, or recently more often cooperating as a member of the grant team. The information quality of platforms also varies depending on the knowledge or availability of expertise in developing digital platforms. Some sets of research data are not sufficiently described using metadata, sometimes errors and poor-quality processing accompanied by the unavailability of a database specialist lead to the creation of duplicate and redundant data, while their users need to invent complicated procedures to extract information applicable to their research questions from the data. The authors of a previous survey of platforms interviewed academics from departments involved in building research digital platforms (Lorenz & Martínková, 2021). This allows us to reflect on the state of digital research and the availability of digital platforms with research data at the Faculty of Arts at Masaryk University.

3.2 *Platforms and their values*

The initial survey showed that the research infrastructure, built in such a highly heterogeneous

environment as the Faculty of Arts, must necessarily be highly complex. The synthesis of platforms in one infrastructure creates a layered structure, the components of which overlap with each other, with a common basis for interoperability, but also differences due to different technologies, formats, standards, and cultures. Infrastructure is built on an installed base and embedded „inside of other structures, social arrangements and technologies” (Star & Ruhleder, 1996, p. 113). Connecting platforms to a single network increases the availability of data and offers an opportunity for digital research. However, the local use of data depends on the data culture of both the institution and individual academic disciplines. Data culture means “the different cultural norms, value systems and beliefs that inform, frame and justify people’s practices of data production, processing, distribution or use” (Bates, 2018, p. 191). During interviews with representatives of university workplaces that develop research digital platforms, we encountered attitudes that complicate the analysis, re-use, and accessibility of data. One of the participants understands data as investments, which he seeks to profit on and is not interested in providing them freely to others, that is why the workplace keeps its digital platforms isolated and inaccessible. A participant from another department uses the digital platform providing access to the work of Arne Novák, a major Czech literary critic, to search for relevant texts. However, the participant compared the quantitative analysis of the texts to the “study of a beetle, whose legs and elytra have been torn off”. Such an epistemological stance sees data as a passive and dormant information, not as a resource of new relational perspectives and a material of knowledge actively shaping the researcher’s perspective as a consequence of its own existence.

The technological solutions for data storage and presentation involve ethical values that are implemented in the design of digital platforms. The values enter infrastructures in several ways. Either they are implemented by the creators themselves, or by users involved in the design of the digital platform. The design replicates the social arrangement of the group of creators and developers, in particular, the structure of their communication, thus copying the values expressed in the administrative protocols and instructions of their organization into the system (Conway, 1968). The platform developers themselves project the expected ways of using the platform on to technology, documentation, and educational materials, thus inscribing their values into the design of the platform, the organization of the information space, and into the code of technologies and software. Designers must also translate the interests and demands of users into the specific and consequently general needs, which they want to satisfy (Hanseth & Monteiro, 1998). This process, referred to as translation, implements user values into the design of the platform. For example, if they build the platform on the basis of a participatory design, they must not only negotiate the requirements and preferences of users, but also seek solutions to their conflicting interests. The very process of negotiating with users implements the value of democracy and openness into the platform. Another path of values into the system is described by Tera McPherson in her study of the Unix operating system (2012). She has shown how the characteristics and rationality of a certain period (*Zeitgeist*) enter into the design of technological systems through cultural production and organization of knowledge. If a “ghost in the digital machine” is hidden in computing systems (McPherson, 2012, p. 34), more of these ghosts can be hidden in the layered infrastructure at the same time. Their identification awaits yet analyses in the fields of code study, software, and information or media archaeology.

The development of a research infrastructure in a highly heterogeneous environment is not only a matter of making digital resources and information or computing tools available, but also of targeted cultivation of data culture and support for the community of digital humanities scholars.

3.3 Survey of platforms

The first step in building the local infrastructure was a survey of the environment, which included a questionnaire survey and interviews with creators. The questionnaires and interviews were structured in such a way as to gather as much information as possible about the functions of the platforms, their designated communities, and data practices related to data production, data care, and its long-term preservation. All institutes and departments at the Faculty of Arts were included in the questionnaire survey. Subsequently, we conducted interviews at those units that operated or planned to produce digital platforms in the following three years.

Using the questionnaire survey, described above, we identified a total of 50 platforms from various fields of humanities, incl. archaeology, ethnology, film science, music science, philosophy, etc. Those who were willing to provide access to their data via the public interface (35) were included in the online catalogue of platforms (LINDAT/CLARIAH-CZ, 2021a), one of the project outputs. We have divided the platforms into six categories (their definition is available on the online catalogue page): bibliographic databases, digital libraries, factographic databases, dictionaries and encyclopaedias, geographical information systems, and language corpora. From the identified platforms, we have selected four that will be moved to the forthcoming Digitalia MUNI ARTS infrastructure in the first stage of the project.

3.4 Selection of platforms

The selection of platforms was made on the basis of defined criteria (for more details see LINDAT/CLARIAH-CZ, 2021b). When selecting the platforms for migration to the local Digitalia MUNI ARTS infrastructure, we tried to represent various categories of platforms in order to gain experience in transferring diverse content and to be able to plan the future strategy better. At the same time, we had to reflect the request for conversion for the Digital Library of the Faculty of Arts MUNI to support one of the faculty's services. Based on the project commitment, in 2020, we finally selected four platforms to be transferred to the infrastructure:

- 1) Digital Library of Arne Novák - providing access to the extensive work by Arne Novák
- 2) Digital Library of the Faculty of Arts, Masaryk University - digitized publications from the production of the Faculty of Arts
- 3) Cinematic Brno - factographic database that documents the history of film screening and viewer preferences in Brno in 1918-1945
- 4) Projectiles - factographic database with 3D objects of prehistoric projectiles

The following sections of the paper describe the process of migrating selected platforms to the

infrastructure. Firstly, we will focus on the formal requirements and the ideal form of transfer, then we will describe the course of the process in practice, with emphasis on the differences between the ideal and actual forms of transfer and the practical experience implemented into the formalization of the platform migration process based on these differences.

4. Preparation and planning of the repository

4.1 Curation workflow

When planning a new digital repository, or, as in the case of Digitalia Muni ARTS, designing an entire infrastructure covering multiple repositories, a properly designed curation workflow is the basis for future sustainability. Therefore, from the very beginning, our team's intention was to build a repository based on standards and to base the workflow design on best and recommended practices.

The reference model of the Open Archival Information System (OAIS) has been chosen as the logical formal framework that largely influenced the planning and implementation of the infrastructure. This model, adopted as the international standard ISO 14721, does not provide the specific technical design of the archive. It is a conceptual model that defines the general architecture of the repository. The model describes the internal arrangement of the archive in six parts, referred to as functional units, and the external environment composed of management, creators, and end users (Lavoie, 2014).

Another important basis, especially in the identification of activities related to the migration of the content of the included platforms and its subsequent maintenance, is the life cycle model of digital curation developed at the British Digital Curation Centre. This model lists and describes the activities that are carried out, intermittently and periodically, in connection with digital objects that are subject to long-term preservation (Higgins, 2008). Structured distribution of activities makes it possible to better estimate the time, personnel and technical demands. The nature of the two mentioned models shows that the design of the curation workflow focuses on the definition and description of the activities that should be repeated with minimal deviations. From the point of view of management, these are processes with precisely defined roles, responsibilities, and dependencies. The processes described in this way then become an important artefact that serves, for instance, to set up contractual relations with partners and data creators.

4.2 Institutions and stakeholders

An important aspect is also the context that the research infrastructure is created in. The various platforms involved have been created under different conditions, which means that not only their technical implementation differs, but also the structure and expectations of stakeholders and, last but not least, the needs and capabilities of end users. Useful tools that facilitated our understanding of the role of the repository in the wider environment comprised the concepts of vision, mission,

and strategy, which are usually encountered in the field of strategic corporate management (Williams, 2009). These three concepts make it possible to briefly formulate the basic starting points and can be applied both to the repository itself and to the institution or organization responsible for its operation:

- The vision describes the state in the future that we are trying to achieve.
- The mission is an expression of an intention or purpose.
- The strategy is a set of specific steps that lead to the achievement of long-term goals.

In the context of long-term preservation, our mission is the conservation of unique research data created by the research activities of the employees of the Faculty of Arts of Masaryk University, which could be irretrievably lost. The vision is the creation of a trusted repository that will ensure the preservation of data together with its virtual research environment and its accessibility to end users. The strategy describes specific steps for the technical implementation of such a repository and its further development.

Stakeholders associated with the repository are a diverse set of individuals (but also other institutions), representing “interested groups” with very diverse kinds of relationship to the repository. Understanding their requirements and capabilities plays a key role in defining long-term preservation strategies. Communication between repository administrators and stakeholders is necessary in order to meet the expectations of stakeholders (Lavoie, 2014). A specific group of stakeholders defined in the OAIS model is the so-called designated community. It is a subset of end users whose members should be able to understand the archived information in the form in which it is archived and made available (Lavoie, 2014).

4.3 Planning the establishment of the repository

Digital Preservation Coalition in its Digital Preservation Handbook (2015) provides an overview of steps that should be part of the planning of the creation of a modern digital repository. The Digital Preservation Handbook recommends evaluating the readiness of the institution to take care of the data in the repository over the long term. This is done using the Levels of Digital Preservation matrix created by the US National Digital Stewardship Alliance, which distinguishes 4 levels of preservation in 5 key areas: storage, data integrity, data control, metadata, and content (Alliance, 2021). In the event that the current level of preservation is not sufficient, data security is the primary task.

Another key task in planning the establishment of the repository is a thorough documentation of processes associated with long-term preservation. Documentation becomes the part of the technical and descriptive metadata and is necessary for the long-term sustainability of the archive. Based on the collected information about the stored data and the institution in charge of its preservation long-term repository strategies and rules are defined. These include a precise definition of its purpose, plans for further development, communication with stakeholders, regular revisions, and a technical implementation plan. Such a plan includes specifications of requirements for the software solution,

decisions on the hardware or cloud infrastructure used, file formats supported, metadata definitions, and other details. The strategy also includes the choice of appropriate long-term preservation methods - the decision on whether the data will be migrated to other file formats, standardized to ensure its wider usability, or in the case of more complex systems, maintained in its current form, with the means necessary to make them available, for example, in the form of emulation.

The long-term sustainability of the repository is integrally linked to the continuous development of the knowledge and competences of those who are responsible for its operation. Only in this way can we respond both to inevitable changes in the technological field (for example, obsolescence of technologies or formats), as well as to changes in the needs of end-users and, in particular, the designated community. It is equally important to maintain contact with the community that is involved in the development of long-term data preservation and our own contribution to the development of the field, for example, via publishing activities and sharing our experience.

4.4 Application of the principles of project management to the emerging repository

The management of the implementation of the digital repository can be based on the procedures and methods of project management as defined by the standards commonly used in the corporate environment - for example, the IPMA, PMI or Prince2 methodologies.

In addition to the aforementioned general project management methodologies, there is also a dedicated Planning Tool for Trusted Electronic Repositories - PLATTER, an output of the Digital Preservation Europe project, which places a significant emphasis on credibility. According to PLATTER, a repository is considered trustworthy if its ability to perform certain functions can be demonstrated and if these functions meet the minimum agreed criteria applicable to all "trustworthy repositories" (Rosenthal, Blekinge-Rasmussen, & Hutař, 2009). Similar to general methodologies, PLATTER uses a planning cycle divided into individual steps and focuses on the area of planning of the strategic objectives of the repository. Originally, it was created as a supplement to the DRAMBORA tool, which served for self-audit of repositories (DRAMBORA Consortium, 2015). The latter is no longer being developed (although its online version is still available), however, the use of PLATTER in archive planning increases the chances of success in auditing the repository using other tools and certifications (Rosenthal, Blekinge-Rasmussen, & Hutař, 2009).

5. Implementation of the research infrastructure

One of the first steps of the project was setting up the team and preparing the environment for building the infrastructure. The core of the team consists of the main project researcher (at the same time, the coordinator with the role of information requirement analyst), project manager, metadata librarian, and two programmers. This core became the basis of a wider project team, which has changed over time. Since its inception, the project has involved 14 people in a total of 2 FTE/year. Other roles represented in the project team: digital curator, digital humanities specialist or discipline-specific consultant.

5.1 Plan for building the research infrastructure

When building the infrastructure at the Faculty of Arts MUNI, we divided the whole process into the preparatory and construction stages. Those stages were further divided into several parts. The preparatory stage involved the development of a plan for building the research infrastructure, the identification of stakeholders, and the selection and implementation of a technical solution. The construction stage consists of two complementary processes: migration of individual platforms and working with the academic community.

5.1.1 Preparing the plan

Planning followed a pre-defined strategy. It included the building of an infrastructure for digital platforms (later called Digitalia MUNI ARTS), subsequently populating it with existing local platforms, as well as ensuring the transfer of metadata to the LINDAT/CLARIAH-CZ national node. The importance and benefits of the planned infrastructure were described in the explanatory memorandum, drawn up as part of the effort to ensure the long-term support of the parent institution. The research infrastructure is also part of the Strategy of the Faculty of Arts MUNI for 2021-2028 (Horáková, 2021), which also mentions infrastructure in its objectives. At the same time, one of our key objectives is to obtain a certificate of credibility for the repository where research data platforms are stored. When deciding what certification to apply for, we considered two options - Nestor Seal for Trustworthy Digital Archives and Core Trust Seal. Finally, we decided on the latter option, due to the financial demand, the course of the evaluation process, and the available practical experience from colleagues operating the central repository of LINDAT/CLARIAH-CZ, which is already certified.

5.1.2 Identifying the stakeholders

During the planning stage we identified nine groups that would be affected by the prepared infrastructure. We analysed their respective needs and expectations. The first three groups are involved in building research platforms, the other three groups provide the management for the home institution, and the last three groups form an external network of users.

- Platform creators - platform creators expect seamless input of new data into the infrastructure while ensuring long-term accessibility of the content. It is evident from these requirements that the infrastructure must ensure the long-term preservation of live data.
 - Platform programmers and engineers - this group needs clearly described conditions for migrating data and metadata as well as requirements for data formats.
 - Platform administrators - administrators of individual platforms assume that their participation in the project will bring them benefits, especially in terms of ensuring the long-term sustainability and functionality of the platforms as well as maintaining their content. Platform managers need to enhance the prestige of the workplaces the platforms have originated at and to report the results that the platforms are producing. Therefore, when migrating platforms, it is necessary
-

to take care to maintain the identity of the original platform (logos, association with the workplace, relevance to the designated community).

- Faculty management - stakeholders from the faculty management expect that the project will support and facilitate the implementation of top-level research at the faculty. The faculty management is regularly informed about the work procedures and the results achieved at specified intervals.
- Centre for Information Technologies at the Faculty of Arts MUNI - this special purpose department of the faculty is responsible for the use of the allocated funds for the project. The management of the centre is informed about the progress of work and troubleshooting at regular meetings; its representatives also participate in regular summary meetings of project teams every 3 months.
- Open Science support group - expectations of stakeholders in this group include adherence to the institution's information policies in licensing and making research data available as well as identifying and communicating issues and conflicts in implementing information policies into scientific practice. Common goals in data openness facilitate collaboration in the selection and licensing of data across platforms.
- LINDAT/CLARIAH-CZ large research infrastructure - coordinators and partners in the LINDAT/CLARIAH-CZ large research infrastructure expect meeting the set objectives, delivering outputs, transmitting predefined metadata, and regular reporting on the responsible use of allocated funds.
- DARIAH-EU international European network - DARIAH-EU partners need to be provided with metadata in an agreed form, together with statistics on the use of the infrastructure and a list of outputs for the year.
- Digitalia MUNI ARTS users - this group of stakeholders, including researchers, students, general public, and commercial companies, needs access to the content in a convenient form, full documentation, and support in the event of difficulties with access.

5.1.3 Selection and implementation of the technical solution

The selection of the system for creating repository involved several steps. First, we created a list with system requirements. It was important to enable the storage and display of different types of data with different granularity of individual objects. We also wanted to preserve the heterogeneity of the research environment of the Faculty of Arts MUNI as much as possible, while enabling batch administration of several dozen platforms at the same time. Subsequently, a search focused on repository systems and their properties was carried out, followed by testing and evaluation of the systems. The evaluation included also sustainability and size of the support community of developers. Based on the evaluation, the Islandora system was selected as the most suitable one due to its high flexibility in the area of creating structures, typing of stored objects and, in particular, the possibility of user input of data (configurability of forms, etc.).

5.1.4 Project management

The implementation of the project plan is the responsibility of the manager, who is the primary

coordinator of activities and is responsible for compliance with the defined schedule. We use Microsoft 365 tools and applications for project management, namely team communication, task planning, scheduling and document sharing, and GitHub to manage technical documentation. Coordination of the project team takes place through joint personal meetings or online meetings at MS Teams twice a week: one with the curatorial part of the team, the other with all members of the core team.

5.1.5 Transferring individual platforms

After the selection and implementation of the technical solution, we started the migration and integration of individual platforms selected on the basis of predefined categories into the research infrastructure. We have prepared a checklist with a transfer scenario for the implementation of the transfer. The transfer process itself is described in more detail below.

5.1.6 Working with the scientific community

We support the increased use of infrastructure by active cooperation with digital humanities scholars and interested members of the academic community. We support the cooperation in several ways. We organize thematic meetings, focused on, for example, citing data or working with the LINDAT/CLARIAH-CZ central repository. Each platform transferred to Digitalia MUNI ARTS is presented to the public at a joint meeting, where, in addition to the presentation of the platform itself, training in using the platform also takes place. Once a year, we also organize an online conference called Digital Data from the Perspective of a Humanities Scientist, including a virtual poster section too. Both the lecture and the poster sections let researchers, academics, and students from various institutions share their experience from research projects in digital humanities. We also support the training of students in methods and procedures using infrastructure for research in their disciplines. Within the faculty, we organized a faculty-wide propaedeutic course ARTS020 Digital Humanities, which brings students closer to the topics of digital humanities and demonstrates various approaches and tools for working with data in the context of humanities. Inter alia, we cooperate with other scientific institutions, projects and research infrastructures.

6. Procedure for migrating platforms to the infrastructure

To transfer the already existing platforms to the Digitalia MUNI ARTS infrastructure, we have set up a specific procedure in the form of a checklist, modified for individual platforms. In general, it contains four stages divided into several parallel sub-processes: domain model - data migration - interface creation - user testing. The division into four phases is the result of a combination of several sources (Miller, 2015; Van Tuyl, Gum, Mellinger, Ramirez, Straley, Wick, & Zhang, 2018; Hardesty & Homenda, 2019) and empirical expertise of individuals within the team (digital curation, programming, project management). Each successfully published platform was a source of learning and improvement of applied practices.

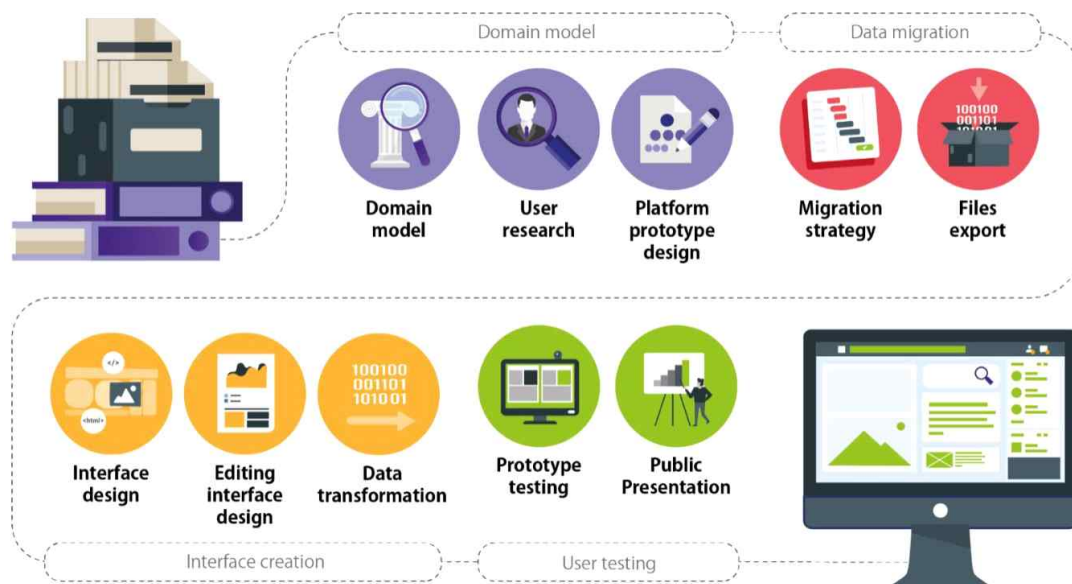


Fig. 3. Digitalia MUNI ARTS workflow

• *Domain model*

1) *Domain model creation*

The first step in creating a new platform or transferring an existing one is to explore the domain which the platform is embedded in. For this purpose, discussions with experts and platform developers take place in order to find out what theories they are based on, what processes they examine and how they name them; we collect information about the history of the application and the community which the platform is designated for. This helps us understand the terminology used in the field. All of these steps help us better understand the environment which the platform is located in and prepare the architecture of the information space.

2) *User research*

In order to better understand not only the scientific domain, but also the content and purpose of the platform, we conduct interviews with users and creators of the platform. The interview is prepared according to the sense-making methodology, which helps us uncover not only the functioning of the platform, but also how users use the platform to give meaning to the data made available in their research, thus also identifying problems and barriers that they are trying to overcome. For a more detailed understanding of research practices from a pragmatic perspective, the interview is conducted in the form of a timeline interview. The researcher illustrates the solution of a typical or last performed task in the platform and each step is analysed using a set of iterative questions. Based on the analysis of research practices, we identify places that need additional design improvements with the potential to facilitate the navigation of users in the information space of the platform while conducting their research.

3) Designing the platform prototype

At this stage, a Dev server is created and based on information obtained from domain modelling and conversations with users, a platform prototype is designed on it. We compile the functions that the platform offers before the transfer as well as the desirable functions and interactions that the users and creators require or need.

• Data migration

1) Compiling the migration strategy

At this stage, it is necessary to compile a migration strategy that describes the list of all files, the transfer process for each file and metadata type, and the target structure of the repository system. According to the file formats that the platform contains, we create a metadata schema with the contained metadata mapped to the Dublin Core. This metadata format is designated for submitting metadata to the LINDAT/CLARIAH-CZ national node. The strategy must take into account the planned integration and interoperability of the platform with other systems, e.g. with the Citace.com system for generating citations of digital objects and data.

2) Exporting files and metadata

The export of files and metadata to the new system follows up. Automatic, but also random manual quality control of data and metadata is performed. When deficiencies are found, corrections and additions are made.

• Interface creation

1) Designing the presentation interface

Just as a high quality editing interface is necessary for the creator, it is necessary to create a user-friendly and useful interface for platform users who will be accessing the content. The creation of the front-end includes the selection of appropriate functional elements, creation of information architecture and navigation elements, and the implementation of a uniform visual style, defined for the Masaryk University websites.

2) Designing the editing interface

One of the key functions of the new system which the platform is transferred to is the creation of a high-quality editing interface. In this step, close cooperation with the creators and programmers of the platform is needed, as their experience and established practices in uploading content must be taken into account. The editing interface is improved by the functions identified in interviews with creators and experts.

3) Data transformation

Transferred data must be handled in such a way as to ensure its integrity, authenticity, secure preservation, easy presentation, and accessibility to users. The necessary functions, such as check-

sums, are configured in the system.

• *User testing*

1) *User testing of the prototype*

In the last stage, user testing is performed using the created prototype of the transferred platform. The aim is to find out how effective and intuitive the platform interface is. We prepare a list of tasks for the testers to complete in the database. We design tasks in such a way that they require using various functions of the platform, such as filtering the content, downloading data, and comparing search results. Then, we observe the testers during the performance of tasks and conduct a short interview with them to explain some of the steps. Based on the results of testing and feedback from users, modifications of the platform are made and the accompanying technical and curation documentation of the platform is completed.

2) *Presentation to the public*

After the modifications are finished, the platform is transferred from the Dev server to a live version intended for publication. The final form of the platform is publicly presented to the academic community, students, project partners, and other interested experts at the announced meeting. Primarily, new functionalities and possibilities of working with the accessed research data and digital objects are presented.

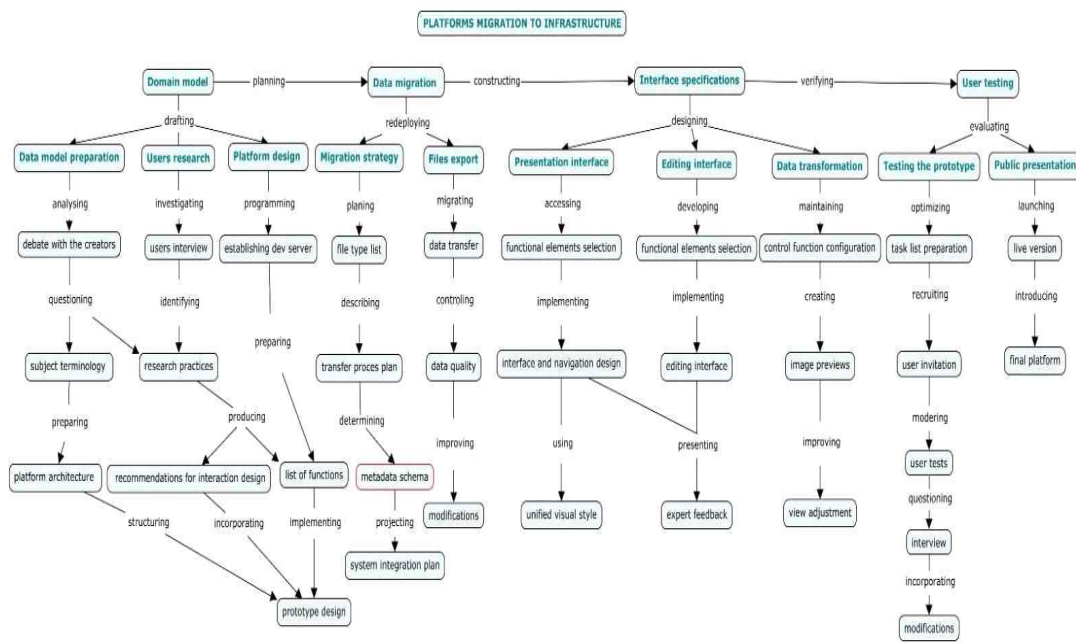


Fig. 4. Platforms migration to infrastructure

7. Plans meet reality

During the migration of the platforms, which was implemented according to the procedure proposed on the basis of the above-described general principles, we encountered several problems that hindered or otherwise complicated the completion of the entire transfer process. Each iteration of the procedure revealed different problems. After the completion of the transfer of each platform, the entire core team assembled at a joint meeting to analyse the identified problems and reflect them in a modified procedure. The problems that we had to deal with can be classified as communication problems, problems of external dependence, and management problems.

Communication problems primarily concerned misunderstandings between our team and the team of creators and administrators of the converted platform. In the course of cooperation, it was necessary to clarify the concepts that were used with a different meaning or were not understood by the humanities experts. Concepts such as platform, research infrastructure, digital library or metadata were particularly difficult to understand. The result was a series of concerns that sometimes led to refusal. The scholars feared that their platform would become buried and lost among other platforms in the infrastructure, the visibility of the platform would decrease, the access of the designated community to the results of their work would become more difficult, or there would be a loss of control over the data or the entire platform. We also encountered a refusal to process metadata, because it did not contribute to the research itself and its creation was seen as unnecessarily time-consuming. Often, these misunderstandings and concerns can be overcome by explaining the state of affairs and the usefulness of the procedure, but even this is not necessarily the rule. We also experienced a misunderstanding with the partner who eventually withdrew from the platform migration; this is also the reason why one of the planned platforms was not transferred to the infrastructure and was replaced by another platform according to the order. It was necessary to formalize the transfer protocol and also to clearly define the competences and obligations of the parties, including the assurances on the copyright of the creators of the platform.

Also, it is necessary to clearly formulate what input data and information will be needed from the administrators to move the platform and regularly inform them about the course of work on the transfer. For more effective communication, it is also necessary to get better acquainted with the terminology and theories of the modelled knowledge domain, which requires a deeper immersion in the topic than just relying on interviews with experts and creators. The concept of open science and open data, including FAIR principles, also needs to be communicated clearly. Although the resistance and distrust of scholars to the concept of open science has been relatively rare, the desire to monetize the outputs of their research work or to make it somehow exclusive, as well as the pressure resulting from the evaluation of research work, still block the way to some interesting platforms that have a high potential for research in digital humanities. However, a more frequent problem concerns licences to the contents of the platforms. For some of the sources, it is not clearly determined who the intellectual property belongs to, especially if some of the authors are already dead or where the digitized source was created as a result of cooperation of a number of authors from several institutions. When communicating and tackling these problems, we also used consultations and cooperation with a lawyer and the MUNI Open Science support group to assist us in the selection

of appropriate licences.

This brings us to the problems of external dependence. The pace of progress on the project must be harmonized and often adapted to the pace of the various organizational units of the university. Continuity with the faculty or university environment is also in the adopted information policy of the institution, which, in addition to licences, determines also the treatment of the work made for hire, the policy of self-archiving, reporting the results of the research work, including reporting the use of the infrastructure itself. We also had to resolve a serious security problem associated with a hacker attack on one of the infrastructure servers. The programmers faced a DDoS attack and were assisted by the MUNI Cybersecurity Team, which performed a forensic analysis and helped with providing server security.

Management problems mainly concerned the timing of processes and procedures. Interviews with creators and experts as well as the examination of users and their research practices led to the identification of a large number of requirements and needs, the satisfaction of which would require considerable work commitment and time capacities. Due to the large number of platforms and the pace at which more are created (around ten per year) and the limited number of people working on the infrastructure, only limited attention can be paid to each platform. This leads to the necessary prioritization of the identified requirements; the part of the requirements categorized as “nice to have” is moved to the documentation and postponed for an unspecified future time. The problem occurs also if the platform does not have sufficiently cleaned or metadata-described data. The involvement of our team in data cleaning is not desired, due to limited time capacities; we refer the platform developers to data analysts that they can hire for data cleaning. We also entrust the metadata description to the original team. Nevertheless, we offer at least a little help with fine-tuning the results, still, only to a limited extent, so that it does not interfere with our ability to ensure the transfer of data to a high-quality platform architecture with a user-friendly interface.

The last major problem that we had to address so far consists in the very nature of research data and platforms. The infrastructure is used for research and, by its nature, requires not only accessibility, but also modification and expansion of data sets. At the same time, research data in individual platforms needs to be stored in the long term with all appropriate archival procedures provided. The digital library and digital archive meet in the infrastructure with their presentation and archiving functions: the infrastructure ensures the archiving of live data. This raises the problem of clearly defining the difference between small and large changes with consequences for subsequent archival and also citation practices.

8. Final recommendations

After identifying the problems, we complemented the platform migration procedure in order to prevent possible problems and speed up the transfer process. These modifications, complementing the general principles of the procedure, are the result of practices that have proven their value and have helped streamline the whole process of transferring platforms to the infrastructure.

- A joint introductory meeting - at the beginning of the transfer of the specific platform, before the commencement of discussions with experts, we organize a clarification meeting of the creators, programmers and administrators of the platform with the entire core team of the infrastructure. At the meeting, we explain the selected concepts, which will help us not only streamline communication and prevent misunderstandings, but also realistically define the expectations of what we can provide to the target group. We pay particular attention to the concepts of large research infrastructure, repository or digital library and metadata. We also pay attention to the unification of terminology, as the same terms may have different meanings in other domains.
- A handover protocol - in the preparation stage, we set up a handover protocol on the transfer of the platform to the infrastructure, to be signed by both parties, i.e., the creators/administrators and the representatives of the infrastructure team. The protocol clearly sets out the responsibilities and obligations of the parties involved.
- Expert consultation on licences - still in the first stage, it is also necessary to set up the rights and licenses for the entire content of the platform and decide on the degree of data openness. We will invite a lawyer and the Open Science MUNI group to assist in the selection of appropriate licenses.
- Identification of theories - during domain modelling, we study expert articles by researchers in order to understand the theoretical background and how to use the collected data to produce knowledge in a given domain, which is beneficial mainly in the stage of designing the platform functions.
- Progress report - we try to prevent communication misunderstandings by regularly reporting on the progress of work. In the course of the platform transfer process, platform developers/managers are informed at fortnightly intervals about the progress and results of the platform transfer to the infrastructure.
- Requirement prioritization - we have included two core team meetings in the process of transfer of each platform, to go through and evaluate the requirements and identified needs of users and creators. The first prioritization meeting takes place after the creation of the list of functions and recommendations based on the analysis of interviews with creators and users, the second one after user testing and demonstration of the platform prototype to creators. Each requirement or need is evaluated on a must-have, nice to have, long term and not needed scale with an estimate of the time demand of its implementation into the platform. Requirements that will not be incorporated into the final version of the prototype are recorded together with an estimate of time demand as an addendum to the handover protocol.

Cultivation of data culture is a long-term strategic goal, which is systematically addressed by delegated groups at the university. We try to contribute to its development by educating students in a faculty-wide course, annually organizing several thematic workshops for students from the academic community, inviting also colleagues from partner institutions or other large research infrastructures, as well as meetings with the professional public when presenting the transferred platforms. Despite the increasing availability of digitized sources and data, research practices of experts do not undergo significant changes.

Easier accessibility and searching in digital sources are used, using the potential of computational methods for analysis is less frequent. Therefore, in the future, we plan to focus more on the implementation of language technologies provided by the LINDAT/CLARIAH-CZ repository into individual platforms and semantization of the administered data using crowdsourcing workshops, where we expect to create a virtual research environment supporting data analysis by non-technological users and form groups of enthusiasts around platforms who will learn methods of more advanced work with the research data accessed in this way. The creators of new platforms also often ask us for recommendations of programmers to help them implement their plans. Therefore, we are going to create a tool for sharing time and expertise that would connect both expert groups.

Acknowledgements

This paper was written under the LINDAT/CLARIAH-CZ (LM2023062) project, fully supported by the Ministry of Education, Youth and Sports of the Czech Republic under the Large Infrastructures for Research, Development and Innovation programme.

References

- Alliance, D. S. (2021). 2019 Levels of Digital Preservation. doi: 10.17605/OSF.IO/QGZ98
- Asadi, S., Abdullah, R., Yah, Y., & Nazir, S. (2019). Understanding Institutional Repository in Higher Learning Institutions: A Systematic Literature Review and Directions for Future Research. *IEEE Access*, 7, 35242-35263. doi: 10.1109/ACCESS.2019.2897729
- Baldwin, C. Y., & Woodard, J. (2009). The Architecture of Platforms: A Unified View. In Gawer, A. (Ed.). *Platforms, Markets and Innovation* (pp. 19-44). Cheltenham: Edward Elgar Publishing.
- Barwick, J. (2007). Building an Institutional Repository at Loughborough University: some experiences. *Program: electronic library and information systems*, 41(2), 113-123. doi 10.1108/00330330710742890
- Bates, J. (2018). Data cultures, power and the city. In Kitchin, R., Lauriault, T. P. & G. McArdle (Eds.). *Data and the City* (pp. 189-200). Abingdon, Oxon: Routledge.
- Bogost, I., & Montfort, N. (2008). New Media as Material Constraint: An Introduction to Platform Studies. In Ennis, E. at al. *Electronic Techtonics: Thinking at the Interface* (pp. 174-191). Proceedings of the First International HASTAC Conference Duke University, Durham, North Carolina, 2007. Lulu Press.
- COAR (2017). Next Generation Repositories: Vision and Objectives. Confederation of Open Access Repositories. <http://ngr.coar-repositories.org/>
- Conway, M. E. (1968). How do committees invent? *Datamation*, 14(4), 28-31.
- Coughlin, D. (2022). Balancing Community and Local Needs. Releasing, Maintaining, and Rearchitecting the Institutional Repository. *Information Technology and Libraries*, 41(1). doi 10.6017/ital.v41i1.14073
-

- Derrida, J. (1996). *Archive Fever: A Freudian Impression*. Chicago: Chicago University Press.
- Digital Preservation Handbook. (2015). Digital Preservation Coalition.
<https://www.dpconline.org/handbook>
- DRAMBORA Consortium. (2015). DRAMBORA Interactive: Digital Repository Audit Method Based on Risk Assessment. <https://www.repositoryaudit.eu/>
- Hanseth, O., & Monteiro E. (1998). Understanding information infrastructures. Unpublished Manuscript.
https://www.researchgate.net/publication/265066841_Understanding_Information_Infrastructure
- Hardesty, J., & Homenda, N. (2019). The Ecosystem of Repository Migration. *Publications*, 7(1), 16. doi: 10.3390/publications7010016
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 3(1), 134-140. doi: 10.2218/ijdc.v3i1.48
- Horáková, J. (2021). Strategický plán Filozofické fakulty Masarykovy univerzity 2021-2028 [Strategic Plan of the Faculty of Arts of Masaryk University 2021-2028]. Brno. Masarykova univerzita.
https://is.muni.cz/do/phil/uredni_deska/Ostatni_dokumenty/dlouhodob_e_zamery/dlouhodob_y_zamer_ff_mu_2021-2028/DZ_FF MU_2021_everze.pdf
- Lagzian, F. Abrizah, A. & Wee, M. Ch. (2015). Critical success factors for institutional repositories implementation. *The Electronic Library*, 33(2), 196-209. doi: 10.1108/EL-04-2013-0058
- Lavoie, B. (2014). The Open Archival Information System (OAIS) Reference Model: Introductory Guide. DPC Technology Watch Report 14-02 October 2014. Digital Preservation Coalition. doi: 10.7207/twr14-02
- Li, W. & Ronghui, Z. (2019). Construction of Service-oriented University Institutional Repository Integrating Scientific Research Management. In 5th International Conference on Information Management (ICIM). doi 10.1109/INFOMAN.2019.8714668
- LINDAT/CLARIAH-CZ. (2021a). Catalogue of platforms at FA MU. Digital Humanities MUNI ARTS.
<https://digital-humanities.phil.muni.cz/en/research-and-projects/catalogue-of-platforms-at-muni-arts>
- LINDAT/CLARIAH-CZ. (2021b). Survey and selection of digital platforms. Digital Humanities MUNI ARTS.
<https://digital-humanities.phil.muni.cz/en/articles-3/survey-and-selection-of-digital-platforms>
- Lorenz, M., & Martinková, P. (2021). Towards research infrastructure: digital platforms and data cultures. In Steinerová, J. and M. Pastierová (Eds.). *Knižničná a informačná veda XXIX: Library and Information Science XXIX* (pp. 60-75). Bratislava: Univerzita Komenského.
- McPherson, T. (2012). U.S. Operating Systems at Mid-Century: The Intertwining of Race and UNIX. In Nakamura, Lisa a Peter Chow-White. *Race after the Internet* (pp. 21-37). New York: Routledge.
- Miller, A. (2017). A case study in institutional repository content curation: A collaborative partner approach to preserving and sustaining digital scholarship. *Digital Library Perspectives*, 33(1), 63-76. doi 10.1108/DLP-07-2016-0026
- Miller, S. J. (2015). *Metadata for digital collections: a how-to-do-it manual*. New York: Neal-Schuman.
-

- Rambousek, J. (2000). Filbit: sdružení uživatelů výpočetní techniky v humanitních vědách [Filbit: an association of computer users in the humanities]. <https://www.phil.muni.cz/digit/filbit>
- Rosenthal, C., Blekinge-Rasmussen, A., & Hutař, J. (2009). Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER) [Planning Tool for Trusted Electronic Repositories (PLATTER)]. Praha: Národní knihovna České republiky. <http://www.ndk.cz/platter-cz>
- Stein, Z. G. (2021). Shopping for an IR: The Search, Adoption, and Implementation of the University of Louisiana at Lafayette's Institutional Repository Platform. *Codex*, 6(2), 93-113.
- Straka, M., & Straková, J. (2014). NameTag. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11858/00-097C-0000-0023-43CE-E>
- Sweeper, D. & Ramsden, K. (2020). Establishing and promoting an institutional repository and research information management system. *Library Hi Tech News*, 37(7), 9-12. doi 10.1108/LHTN-04-2020-0031
- Van Tuyl, S., Gum, J., Mellinger, M., Ramirez, G. L., Straley, B., Wick, R., & Zhang, H. (2018). Are we still working on this? A meta-retrospective of a digital repository migration in the form of a classic Greek Tragedy (in extreme violation of Aristotelian Unity of Time). *Code4Lib Journal*, 41. <https://journal.code4lib.org/articles/13581>
- Williams, K. (2009). *Strategic Management*. New York: DK Publishing.

[About the authors]

Michal Lorenz is assistant professor of Information and Library Studies at Masaryk University in Brno, Czech Republic. He holds a PhD in Information Science from Charles University in Prague. His research interests include annotation behaviour, data culture, LIS curriculum studies. From 2012 to 2015 he served as a member of the Scientific Board of the Moravian Library. In 2014-2018 he was a member of the EUCLID committee and participated in the organization of the 23rd International Conference BOBCATSSS in Brno. Since 2019, he has been working on building the LINDAT/CLARIAH-CZ research infrastructure at Masaryk University.

Ema Juranová has an MA in Library and Information Science from Masaryk University, in the Czech Republic. In 2020 she took part in the LINDAT/CLARIAH-CZ infrastructure as a digital curator.

Michal Konečný is software architect, analyst and consultant, specialist in digital curation. he works at 24i Media focused on distribution of audio- and video content and other media over the internet. He freelances to help companies and organizations implement their projects in the field of internet and education. He organizes digital curation workshops in memory institutions across the Czech Republic, since spring 2019 he also lectures this topic at Masaryk University.

Hana Kubelková has an MA in Prehistoric Archaeology of the Near East from Masaryk University,

in the Czech Republic. She is currently pursuing a PhD in Archaeology at the Masaryk University. From 2019 to 2022 she took part in the LINDAT/CLARIAH-CZ infrastructure as a project manager. In 2021 and 2022 she co-organized workshops on data management for archaeologists and spoken on multiple conferences on digital archaeology or data archiving.

Veronika Wolfelová has an MA in Library and Information Science from Masaryk University, in the Czech Republic. In 2020 she took part in the LINDAT/CLARIAH-CZ infrastructure as a digital humanities specialist. Currently she works as a graphic designer at the Czech Statistical Office.
