RECOMMENDATIONS FOR THE CREATORS OF PLATFORMS

The following are recommendations on topics associated with the design of online platforms for research and they are not linked to a specific technical solution. Online platforms for research include various digital libraries, databases, dictionaries or geographic information systems.

It is not an exhaustive list nor generally applicable rules; rather, these are topics worth considering.

1) Specify the purpose and the intended function.................................................................2
2) Describe the content ............................................................................................................2
3) Document the procedure and the result ..............................................................................3
4) Ensure copyright for the provided content .........................................................................3
5) Identify the licences the data is subject to and how it should be cited .........................4
6) With users in mind ...............................................................................................................4
7) Share information and (meta)data .....................................................................................4

Loosely based on

1) Specify the purpose and the intended function

The basic questions can be summarized as follows: Why should the platform exist? What should its content be? Who will be working with the platform and how?

The reason for the existence of the platform is probably given by the research project and the lack of an appropriate resource. The platform is usually used by its administrator (research team) when adding and editing its content. This research team also works with the platform after the content is provided. If the platform is publicly accessible, the group of potential users and their needs extends. The needs of users and the required functions can be described in more ways, from simple text descriptions to various types of models. There are also a number of methods for identifying these needs (more under point 6). The decisions concerning the description of the content are also important for the functionality (more under point 2).

Other requirements may pertain to the data about the use of the platform. For example, it is appropriate to determine whether we want to record the numbers of file downloads or individual record views and whether this data should be displayed publicly, directly in the system.

2) Describe the content

In order to be able to work with the objects in the collection (scanned materials, data, photos, text, ...), these objects should be appropriately described.

In a dummy collection of letters, their description may concern the letters alone (names of individuals who exchanged the communication, description of the content of the letters, dates of dispatch, paper size, colour of the ink, ...) or their senders and recipients (country of origin, field of expertise, ...). How comprehensive the description should be, it depends on the needs of the users, first of all, the research team that creates the collection. The decisive factor is also the amount of time available to create the collection.

When formulating the rules for describing the content, it is necessary to decide what data to include, where it will be obtained from and how it will be entered.

The following are examples of potentially problematic areas:

- Objects in the collection may have authors or other associated individuals that should be clearly identifiable. The problem may be aliases, different variants of writing the name (J. R. R. Tolkien vs John Ronald Reuel Tolkien) or different individuals of the same name.
- Similar problems may occur when identifying geographic areas. For example, if the dummy collection of letters contains letters sent from the town of Zlín from the period when it was called Gottwaldov, it is necessary to consider which name should be used.
- The description of the content can be very subjective and for one collection it is often prepared by more people. Therefore, it is appropriate to consider drafting allowed terms and phrases and instructions for their use, including the required
number of entries. The same object can be described from different aspects, therefore, the choice of terms should be adapted to the purpose of the collection.

- Problems can be caused by various ways of writing dates (4. February, 2020 vs 2020-02-04 vs February 4, 2020), ranges (1800-1900 vs 19th century), text description of periods (Late Middle Ages, Jura), and estimated dates (circa 1800 vs 1800?). Whatever the chosen format, it is imperative to be consistent in its use.
- There are freely available lists of persons, places and topics which may also include synonyms and relations between the contained entries. For local names, Getty Thesaurus of Geographic Names (TGN) can be used; the entry for Zlín also lists Gottwaldov as an alternative name. The use of these dictionaries, when implemented appropriately, can ensure unambiguous identification outside the respective platform.

All of the above decisions on how to describe objects should be listed and available to everyone who may work with the collection. These decisions will be affected, inter alia, by the practices established in the respective discipline and should be made with regard to the future usability and sustainability of the platform and its content.

The method of storing the descriptive information will also vary, depending on the chosen technical solution.

3) Document the procedure and the result

There is often little time for documentation, still, it is necessary in order to use the full potential of the laboriously built collection. The technical documentation is essential for any machine processing of the contained data or its transfer to another platform (for example, if the existing solution becomes obsolete and does not work properly). For inspiration, you can use the available Documentation Template (PDF), listing the basic domains to be addressed.

For better use of the collection, it is also important to document the decision which led to its creation and form. General information about the content and objectives of the project should be available also on the website of the platform, if it has a public address.

4) Ensure copyright for the provided content

If the collection contains digitized content, it is necessary to know whether and how it can be used. It is also necessary to identify the copyright holders. Special attention should be given to personal and sensitive data, if any such data is included in the collection.

According to Czech law, where more than 70 years have passed since the death of the author of the digitized content, the work can be automatically treated as copyright-free. Otherwise, stipulations of §27b(3) of the Copyright Act should be observed. In the event that the copyright holder cannot be found in those resources, a licence for certain uses of orphan works should be requested pursuant to §37a.
5) Identify the licences the data is subject to and how it should be cited

The question of further use of contained data does not pertain only to freely accessible collections. If you are the creators of the content or its parts, it is appropriate to specify, which licence this content is provided under. You can use the Creative Commons licence, for instance. The CC BY variant is recommended as the least restrictive option.

The database itself as a whole may also be licensed. It may be protected by copyright, if is the so-called original database, or by the database right. For more information and a schematic, visit Otevřená data (Open Data; Czech version only).

For further work with the content, it is important to define clearly its creator and time the content was created and published. Furthermore, it is customary to indicate how the collection should be properly cited, either as a whole or as individual objects. Citing was addressed in an article (from June 2020, Czech version only).

6) With users in mind

The users of the research platforms are usually the creators themselves. If the platform has a potential for further use, it is appropriate to have more people involved in its creation. An important part of preparing the web content and services is also accessibility to users with various types of disabilities.

Prototypes (interim, incomplete versions) can be used to test the platform and find out whether it is easy to use. For user testing, it is recommended to assign tasks the testers will be asked to perform: for example, view all letters exchanged between person A and any person from Zlín. General information about user testing can be found, for example, on 100 metod (100 Methods; Czech version only), along with other techniques for identifying user needs.

After putting into operation, it can be monitored how users use the platform (e.g., using Google Analytics or HotJar). It is possible to use data generated by the system itself or methods such as user observation or shadowing.

7) Share information and (meta)data

To make potential users aware of the platform, it is worth trying to find out whether there is a register of platforms with similar focus where the platform could be listed. Platforms created at the Faculty of Arts, Masaryk University, are entered in the catalogue. There are also common search services that collect metadata* from multiple platforms. In this case, it is required that the metadata meet certain requirements.

* The US National Information Standards Organization (NISO) defines metadata as structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.