National Czech Programme | eosc

Recommendations for Creating Repositories in the NRP Environment

M. Janíček, I. Brejchová, D. Mikšík and the EOSC-CZ and NRP Projects Collective

Version 1.0 from September 8, 2025

This document is primarily intended for representatives of user communities considering the establishment of a new repository instance in the **National Repository Platform (NRP)** environment**.** It will help you navigate the whole process of setting up and running a repository and not forget any crucial aspect. The aim is to provide an overview of the various activities that are integral to the repository creation process, including those that are the responsibility of a representative of the commissioning user group rather than the NRP. The document will help you make an informed decision on whether a repository in the NRP environment is the right solution for you. It can also serve as a guide to more detailed materials on individual activities and as an aid in scheduling work on a new repository if the user group decides to build their repository in the NRP environment.

The document has been produced as an output of Key Activity 7.1 *Methodological Materials for NRP Software Services* of the NRP project and will be continuously updated.

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

National Repository Platform for Research Data (NRP)
CESNET, an interest association of legal entities
Correspondence address: Generála Píky 430/26, 160 00 Prague

Version history

Version	Date	Description
0.9	23. 4. 2025	First public version [Czech only]
0.9.1	19. 5. 2025	Correction of text readability [Czech only]
0.9.2	21. 7. 2025	Addition of references to repository licenses and Metadata Standards Catalogue, clarification of the section Clarify the purpose of the proposed repository exclusive to the form, and clarification of some wording.
0.9.3	14. 8. 2025	Removal of incorrect link to OS II and new version of Conditions (3.2)
1.0	8. 9. 2025	Link to the document describing AAI for NRP and proofreading of the English text.

Table of Contents

Introduction

The document Conditions for Creating New and Modifying Existing Domain Repositories[1] (hereafter referred to as *Conditions*) describes three basic use cases for NRP:

I. Building a repository using one of the core repository systems, i.e., as an instance of CESNET Invenio, CLARIN DSpace or ASEP/ARL (point B of the OS II call, i.e., building new repositories).

II. Building a repository using alternative repository systems in justified cases (point B of the call).

III. Connecting an existing repository that is still operated outside NRP systems (point A of the call, which concerns only existing repositories).

For each of the use cases, it is specified what will be provided by the infrastructure as a service and what is the responsibility of the repository administrator. In the document you are reading right now, we discuss the first use case in more detail. Parts of this document may also be useful for those building repositories using another repository system, but in that case, we cannot provide any methodological support on how you can implement each step. The administrator of an existing repository may choose to either integrate an existing solution (see the section *Integrating an existing stand-alone repository into the NRP/NDI environment*) or migrate the data to a new instance of the core repository system in the NRP environment (in which case you can skip chapter 1. Before Establishing the Repository in this document and continue to chapter 2. Before Implementing a Repository Instance).

We also take most of the definitions of basic terms from the document *Conditions*, so we recommend that you familiarize yourself with them before reading this document.

[1] Matyska, L. et al. (2024). *Conditions for Creating New and Modifying Existing Domain Repositories. version 3.2* https://www.eosc.cz/media/3884116/conditions-for-creating-new-and-modifying-existing-domain-repositories.pdf

1. Before Establishing the Repository

Before deciding to establish a repository in the NRP environment, representatives of the scientific community that the repository is intended to serve should try to briefly answer the questions in this template.[2] If you still decide to set up a repository after answering these questions, the repository system specialists will help you select a repository system (from the core systems offered by the NRP) most suitable to your needs and guide you through its subsequent implementation. Your answers to the questions in the form will serve to clarify your requirements and help us to choose a suitable solution. It's okay if you can't answer all the questions completely and accurately yet.

We assume that you will want to fund the repository development through a grant. The recommendations in this chapter overlap and complement the requirements for project applications. You can also consult with us before submitting a project application.

1.1 Selecting a Suitable Repository System

Carefully reflect on the answers with your community. Is the establishment of a new domain-specific repository an adequate solution to your problem? Will your community have sufficient institutional support and staffing (see Conditions for more details) to create, set up and operate its new repository instance within the National Repository Platform?

If so, the next step is to contact the NRP repository system specialists,[3] who will help you fine-tune your requirements and select and then implement a suitable solution (Invenio, DSpace, ASEP/ARL). The different repository systems have essentially the same core functionality but may differ in the details of how specific features are implemented and the extent to which you can customize the features. The ease of integration with other systems, for example, may also be a deciding factor for you. We will therefore recommend a system only after we have discussed your requirements. You can learn more about the systems from their website and documentation, and you can also try out the user interface of existing instances.

Invenio

- Invenio website[4]
- Invenio RDM documentation[5]
- National Data Repository[6]
- National Metadata Directory[7]

DSpace

[2] Purpose and Basic Features of the Proposed Repository. https://www.eosc.cz/media/3886528/repository_purpose_form.docx
[3] Contact us. http://www.eosc.cz/sluzby/ukladani/repozitare-v-nrp/kontaktujte-nas
[4] Invenio. https://inveniosoftware.org/
[5] Invenio RDM: The turn-key research data management repository. https://inveniordm.docs.cern.ch/
[6] National Data Repository. http://datarepo.eosc.cz
[7] National Metadata Directory. https://nma.eosc.cz/

- DSpace website[8]
- Overview of DSpace features[9]
- LINDAT/CLARIAH-CZ repository[10]

ASEP/ARL

- Documentation[11]
- Repository of the CAS[12]

If you are not satisfied with any of the NRP core repository systems, you can also choose to build an instance of your own repository system on NRP resources, but in that case, you will also bear the other costs of this process (see Chapter 5: Building a Repository Running on NRP Resources Without Using the Core Repository Systems in the Conditions).

[8] DSpace: Build an Open Digital Repository. https://dspace.org/

[9] DSpace 7.x Documentation: Functional Overview. https://wiki.lyrasis.org/display/DSDOC7x/Functional+Overview

[10] LINDAT/CLARIAH-CZ Repository. https://lindat.mff.cuni.cz/repository/

[11] ASEP: Data Repository. https://asep-portal.lib.cas.cz/uvod-do-asep/datovy-repozitar/

[12] ASEP: CAS Repository. https://asep.lib.cas.cz/arl-cav/en/search/

2. Before Implementing a Repository Instance

Once you have decided to build a repository instance in the NRP environment and have chosen a specific repository system, you will need to develop a more detailed specification of the repository to be implemented. This chapter will help you not forget anything crucial.

2.1 Data and Metadata

Metadata allows us to find and describe data in a structured way for other users who need to contextualize and interpret the data correctly. However, there is not always a clear boundary between data and metadata, nor between different types of metadata. In this document, we will consider metadata to be a structured description of a record, and we will use data to refer to files that can be uploaded to a record. The (metadata) data model will form the foundation of your repository, so it is important to carefully consider its selection based on a deep understanding of the data to be stored in it.

There are basically two types of repositories in terms of their data and metadata model:

(a) General repositories with heterogeneous data

The metadata model serves as a wrapper over any dataset that falls within the scope of the repository as you define it. An example of such a model is CCMM (Czech Core Metadata Model) for describing datasets, which is based on DataCite and DCAT-AP. It will be used in the National Metadata Directory (NMA) and the Catch-all repository. Its specification can be found on GitHub[13] or ccmm.cz.[14]

If your expert community needs to primarily store datasets or needs to store very heterogeneous data, they may choose to reuse the Invenio implementation already being built over CCMM in NRP, or just extend it with domain-specific fields (e.g., domain-controlled vocabularies), which we as an infrastructure can help with, as long as it is not a large-scale change.

b) Dedicated repositories with homogeneous data

A repository contains data objects of predefined types and uses specialized (domain-specific) metadata models to describe them. If there are more than one such object type, it is necessary to describe the relationships between them.

This option places significantly higher demands on the repository administrator, who must be able to implement the proposed metadata model with their own resources. The infrastructure will provide a technical specification on how to do this for each of the underlying repository systems. It is strongly recommended not to invent the metadata model yourself, but to adopt an already existing metadata schema if it exists. For example, the Metadata Standards Catalogue[15] can be used to select a schema. NRP

[13] Med, M. et al. (2024). Czech Core Metadata Model specification. GitHub. https://github.com/EOSC-CZ/CCMM
[14] Czech Core Metadata Model. https://www.ccmm.cz/en/
[15] Metadata Standards Catalogue. https://rdamsc.bath.ac.uk/

provides the DataSpecer tool[16] for specifying metadata models. It will allow you to generate a machine-readable schema in JSON and XSD based on your specification. The repository systems also allow you to define multiple metadata models in a single repository instance. At the same time, data represented in one schema can be transformed into as many others as you like, if you specify how. At a minimum, conversion to CCMM will be required, allowing the harvesting of metadata into NMA. Your proposed metadata model should aim to meet the FAIR principles, which are discussed in more detail in 2.5 Ensuring FAIR P.

The full specification of the (meta)data model will at least include machine and user interface field names, the data type of the fields, and information about the obligation and repeatability of the fields. The specific form in which the model will be specified is subject to agreement between you and the specialist of the selected repository system.

In addition to the metadata model itself, we recommend that you address the following aspects of the data and metadata:

Collections

- The data and metadata in the repository can be grouped into collections. Consider whether you want to use them and if so, describe how they will be created and what they will be used for.

Identifiers

- Metadata records in the repository must be tagged with persistent identifiers (PIDs). It is up to you to choose the appropriate type of PID (e.g. DOI, handle, ...) and determine how and when they are assigned. You can learn more about support for persistent identifiers at the Czech national level at identifikatory.cz,[17] which is maintained by the National Center for Persistent Identifiers at the National Library of Technology. The selection of appropriate PIDs can be consulted with the Centre (identifikatory@techlib.cz). Any changes to the data or metadata should be recorded, typically in the form of the creation of a new version or revision. Different repository systems may take slightly different approaches to versioning and PID allocation; the repository specialists of each system will explain in detail the system's default versioning settings and what can be customized.

Backup

- Regular backups of data and metadata to secure locations are provided by the infrastructure.

Formats

- Next, we recommend doing an analysis of the existing file formats in which the data is stored to see if they are suitable for long-term archiving. If not, we recommend converting it to a more suitable format before archiving. This will then form the basis of

[16] Dataspecer: A tool for management of semantic data specifications. https://dataspecer.com/
[17] Identifikatory.cz: Persistent Identifiers website. https://identifikatory.cz/en/

the list of formats that are accepted by the repository and that you commit to supporting. Having and enforcing such a file policy will make the long-term management of the repository much easier. You can refer to lists of formats suitable for long-term preservation, such as the one from the Library of Congress.[18] We recommend using open data formats as much as possible.

Embedding data and metadata

- Data and metadata can be uploaded to the repository via a form in the user interface or via the API. We discuss the management of the data storage process in more detail in chapter 2.3 Selected Aspects of Administration. We recommend using the API at least at the beginning to initialize the repository prototype with parts of your data along with metadata, which will allow testing the UI with real data. However, the data and metadata must be structured and cleaned before importing. In the same way, the data can be imported into the finished repository at startup so that the data does not have to be manually entered by the user community.

Interfacing with other systems

- Repositories can function as one element in a broader ecosystem of disciplinary tools and practices for working with data, so it is also important to know which external systems the repository will be linked to and for what purpose. Examples of such external systems include electronic laboratory notebooks (ELNs), the use of repositories to automatically store data from instruments, or linking a repository as a data source for other tools (e.g. Jupyter notebooks). Repository APIs allow not only the entry of data but also access to it. However, each access from outside systems should always be individually assessed and the technical and administrative (e.g. legal) context verified.

2.2 Legal and Ethical Aspects

In the preparation of the repository, some legal aspects need to be considered. The repository service provider must be authorized to store and access the data, and the contributors must be able to determine how other users can further handle their data (specific licenses, access restrictions and embargoes, etc.). Specific requirements for data access restriction options may be reflected in the overall setup of the repository and its workflows. In the context of this issue, we refer to the following terms:

Deposit licenses

- When depositing content (datasets) into a repository, the contributor is required to grant the repository service provider permission to use the dataset, specifically to store it and possibly make it available to the public. If the contributor represents the creator of the dataset, they must be authorised to be able to grant a licence to use the dataset. This

[18] Library of Congress. *Recommended Formats Statement.* https://www.loc.gov/preservation/resources/rfs/index.html

permission is granted by the contributor through a so-called deposit licence. Sample license texts are provided by the infrastructure.[19]

User licence

- The repository operator should define a set of so-called user licenses from which contributors can choose when depositing a dataset. The specific conditions for the use of the dataset by public users (how users can/cannot handle the dataset) will be governed by the user licence chosen and attached by the contributor. As a rule, these will be variants of the Creative Commons public licenses,[20] but the range of licenses offered may be wider – depending on the needs of the specific institution/repository service provider. The possibility of depositing a dataset to the repository without a user licence is also an option.

Dataset without user licence attached

- If no user license is attached to the dataset (if the repository allows such an option), the dataset will be in "all rights reserved" mode. In practice, this means that the dataset is available to users on a "read-only" basis and can only be used beyond reading under exceptions to copyright provided by law (e.g. to quote excerpts, to use for personal use). Such a dataset will not be further shared unless the user obtains explicit permission from the person who exercises copyright ownership of the dataset (typically the author or employer).

Other possible restrictions on user access to the dataset

- In addition to the terms of the chosen user licence, public access to the dataset may be restricted by:
    - a delay of access (embargo);
    - the choice of the accessibility level of the dataset in the repository, e.g. by limiting the range of users to whom the dataset is available only to persons from a specific institution;
- granting the access to the dataset based on a licence agreement/NDA;
    - limiting the access to the dataset by the need of password/contributor approval.

Data stored in the repository may require different levels of protection in terms of security and access, depending on its sensitivity. This ranges from data that can be made freely available to anyone, data that is not intended for publication, to data that can cause significant damage if made available to unauthorised persons. In the process of developing a repository, careful consideration must be given to the sensitivity of the data to be stored and the security arrangements for handling it.

Another important issue is the preservation and possible disclosure of data that is personal data. The processing of this type of data is governed by data protection legislation, notably the

[19] Deposit license template. Outputs of the EOSC initiative. https://www.eosc.cz/projekty/vystupy-iniciativy-eosc-cz#depo_licence
[20] Creative Commons. *About CC Licenses.* https://creativecommons.org/share-your-work/cclicenses/

General Data Protection Regulation (GDPR),[21] which implies specific principles and requirements that must be met without fail.

Carefully consider whether the data you plan to store in the repository includes sensitive data and/or personal data and what technical and organisational measures you will need to take when storing and disclosing it.

2.3 Selected Aspects of Administration

In this section, we will focus only on selected aspects of repository administration, namely roles, workflows, long-term funding of the repository operation and management of users and user groups. Other basic repository concepts – record, community, etc. – are described in the *General Principles of NRP* chapter of the *Conditions* document.

Role

- Users accessing the repository are assigned **roles** that co-determine which features and functionalities of the repository will be available to them.
- As an example, we list the usual roles, which can be extended if necessary
  - anonymous reader – not logged in user;
  - community member – logged in user, can e.g. read records that are not accessible to non-logged in users;
  - submitter – a logged-in user with permission to create (in a defined context) records in the repository (either the dataset creator or an authorised representative of the dataset creator);
  - curator – a user checking the accuracy and completeness of records before their publishing (accepting them into the repository);
  - administrator – a user with almost unlimited permission responsible for the repository from a technical point of view, typically delegated to the programmers of programmers.
- The roles determining the level of permission in a repository instance are not necessarily the same as the roles of the people who run the repository. The responsibilities of the repository administrator and data curator roles are described in the *Conditions* document. The role of repository administrator is typically performed by individuals, whereas institutions typically act as partners or contractors of the infrastructure provider (CESNET/e-INFRA.cz). The repository administrator represents the repository in organisational matters and, among other things, sets the repository policy. However, they do not necessarily have to have administrative rights to the repository; these are usually delegated to programmers on the side of the repository system provider.

Workflows

[21] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

- A workflow is an algorithmized procedure for creating a new record or for accessing an existing record in a repository.
- A typical simple workflow for creating a new record (or "data deposition") includes:
  - Creating a draft record (draft);
  - Submitting the draft for review and approval by the data curator;
  - Approval of the draft – the record is published;
  - Rejection of the draft – the record is returned to the contributor for correction.
- Workflows can be added to and modified as needed – extensive modifications may require development capacity on the user group side.
- If the deposition is to be initiated not only by the user-individual but also by the user-machine, create a mock-up of the entire scenario and consult with the appropriate repository system specialist.

Funding, sustainability

- How is/will the repository be financed? As the text of the *Conditions* document indicates, the repository administrator must be available throughout the repository lifecycle, and the roles of data curator and user support "must exist and be staffed throughout the repository's production". There are costs associated with this that need to be accounted for in the long term. It is natural that funding sources will change over time and that at the time of the repository's establishment, the user group does not have precise information on what resources it will use to cover these costs in the long term. We recommend keeping this aspect of running a repository as an ongoing theme in the strategic and operational decisions of the user group behind the repository – stable staffing for these roles is not just a question of funding, but also of the functional embedding of the people looking after the repository into the structures of the institutions that employ them.
- The involvement of institutions in the operation of repositories, and thus their funding, contributes to the long-term stability of repositories. It can therefore be recommended that the user group should start as soon as possible to look for ways to involve the institutions that have an actual or potential interest in the repository in its maintenance. To this end, the NRP project will provide basic models for the long-term sustainable operation of repositories (expected date Q4/2025).

Users, accounts, groups

- The NRP project envisages that the EOSC CZ Authentication and Authorisation Infrastructure (e-INFRA AAI)[22] will be the initial (not necessarily the only) source of information on users and groups. It is therefore expected that the repository administrator (or their designee) will become familiar with the EOSC CZ AAI architecture and its interfacing with the repository, which will be described in the forthcoming AAI Initial Architecture for an NRP document, and that they will manage groups and users using the Perun AAI environment.

[22] Lényi, P., Vyskočil, P., Vojáček, L. (2025). *Iniciální architektura AAI pro NRP*. https://www.eosc.cz/media/3854788/inicialni-architektura-aai-pro-nrp.pdf

2.4 User Interface

Repositories supported by the NRP funding have mandatory publicity. Logos and rules regarding their location will be made available to you. How to use them correctly will be specified in the forthcoming design manual. In addition, it will also include the terms and conditions for running the repository on the eosc.cz subdomain (should you choose this option) and basic accessibility recommendations.

For the Catch-all data repository, which is implemented in Invenio, we are working on a UI visual, which can then serve as a template that you can subsequently customize. This will be particularly useful if your metadata model is an extension of CCMM. It will then be up to you to just determine the colour scheme of the repository, the repository logo and the accompanying text.

We will handle minor customizations for you, but you will need to communicate exactly what to modify and how. For major customizations, you will need to pay a developer from your own funds to implement the requirements you specify. What is already a big customization and what is not yet, we will discuss with you in detail beforehand.

The more your chosen metadata model differs from existing templates and the more extensive and complicated your metadata model is, the more challenging the UI development process will be and the more in-house capacity you will need to dedicate to this activity.

The user interface of any repository will typically consist of at least the following pages, the form of which will need to be agreed upon:

- Home page
- A page with a metadata record and the ability to download data
- Browse and search:
    - By what metadata attributes should the records be filtered?
    - What metadata attributes should the records be sorted by?
- What metadata attributes should be displayed on the record detail and preview?
- Editing the interface and deposit form:
    - What widgets will be used when entering metadata? (drop down/autocomplete etc.);
    - Customization according to special requirements based on the deposition workflow.

If you want to provide the users with other ways to view the data, you must check with the repository system specialist to see if the repository system supports such methods by default. If not, you will have to program such extension from your own resources as an additional application layer on top of the data. Computations over data typically already go beyond the repository systems. For this purpose, tools for transferring data to the e-INFRA CZ environment will be available, where dedicated resources are available for computational operations over such prepared data.

We strongly recommend that the repository is primarily in English. However, you need to think about what other language versions the repository will need to have and to what extent the content will be translated:

a) The homepage and platform information will be multilingual;

b) The field labels and taxonomy will also be multilingual;

c) The text fields in the metadata records will also be multilingual.

The repository system specialist will provide you with details of how the translation of content is specifically handled in their repository system. Different translation options are differently complicated to implement, so you must consult your choice with the specialist of the respective system. If you choose multiple language versions, you must supply all translations and are responsible for ensuring that they are updated as changes are made. In the case of option c), you must design how the translations will be delivered to users and under what circumstances the translated content will be displayed.

2.5 Ensuring FAIR Principles

The repository's support of FAIR principles intersects with its core functions. However, given the importance of FAIR principles in the current research data ecosystem, we recommend looking at supporting their implementation separately.

The NRP also provides other tools to support the production of research data that fulfils the FAIR principles, details of which can be found in the document *Analysis and design of procedures to fulfil the FAIR principles within the NRP*,[23] where the individual FAIR principles are also explained in detail. Before you start planning your repository implementation, we recommend that you check whether there are suitable, or at least close, domain-specific FIPs (FAIR integration profiles). You will also have the FIP wizard, a tool for declaring the specific solutions with which you have decided to implement each aspect of the FAIR principles. Each repository should create and update a FIP record on behalf of its user community, against which the degree of fulfilment of the FAIR principles will be assessed.

Findability

- Identifiers
  - What persistent identifier will your repository use?
- Metadata schema
  - Do you need to use a domain-specific metadata schema and if so, does it already exist, or will you need to create one? How will it map to CCMM (for linking to NMA) or to the DataCite schema (required when using a DOI identifier)?
- Will controlled vocabularies or ontologies be used to describe data objects? Which ones? Will it be possible to describe them with (freely chosen) keywords?
- To which systems (other than NMA) should metadata be further provided?

Accessibility

[23] Trtíková, I., Bártková, P., Biedermannová, L., Jindřich Fejfar, Robert Pergl, David Novák, & Tomáš Čížek. (2025). *Analysis and proposal of procedures to implement FAIR principles in the framework of NDI/NRP*. https://doi.org/10.71495/5dv6-v116. AT THIS MOMENT IN CZECH ONLY

- Protocols
    - What protocols will be used to access the data? Will there be a need to use other protocols beyond the usual https?
- Authentication and authorization
    - How will authentication and authorization be handled? Will only e-INFRA AAI be used, or will there be additional authentication and authorization systems?
    - How will access to restricted data be handled, or the mechanism for granting access?

Interoperability

- Data formats
    - What will be the used/preferred data formats? Are they all open?
    - Data structure
    - Is there an industry-defined data/metadata structure that could be used?
    - Are the data objects that will be stored in the repository already created using this structure?
- Links to other data and metadata
    - What types of links between data need to be considered ("derived from", "cites", "is part of", etc.)?
- Documentation
    - What is an appropriate documentation format that can be stored with the data object? Is the data structure sufficiently described?

Reusability

- License
    - What licenses are appropriate for your data objects? Do you need to work with special (as defined) licenses?
- Provenance
    - What is the appropriate way to capture provenance information (history of creation and manipulation of the data object, instruments and tools used, etc.)? How is detailed provenance information relevant to your data?

3. During the Implementation of the Prototype

3.1 Documentation and User Support

The repository administrator creates user documentation, instructional materials for contributors (e.g. how to insert data, how to fill in metadata) – NRP provides documentation templates and technical content, as well as a tool for creating and maintaining documentation.

A complementary part of the documentation can be a FAIR implementation profile, recorded in the FIP Wizard, which helps you to document in a structured and machine-readable way how you have decided to implement the various FAIR principles in the repository.

The repository administrator is also obliged to fill in metadata about the repository in the National Repository Catalogue.

Users involved in validating repository prototypes should be told how to provide feedback to the repository administrator. The repository administrator should therefore identify both the communication channels and the responsibility for handling user requests – this activity can be used as a first step to set up routine user support that the repository manager will provide throughout the lifetime of the repository (see the *Conditions* document). The repository documentation and the email for contacting the repository should be forwarded to the NRP service desk so that it can forward any queries to you that it is unable to answer based on the documentation alone.

3.2 Training

The repository system provider will give the repository administrator technical documentation of the repository sufficient for the normal work of the repository administrator, data curator, and its developer. In case of questions, you can contact the specialist of the respective system.

3.3 Repository Functionality Testing

Before deployment to production, it is necessary to test that the resulting repository satisfactorily meets the specification and needs of the commissioning user group. The repository administrator (or someone delegated by the administrator) should review the metadata previews, metadata entry forms, search, and any other features customized for the repository. Any comments should be forwarded to the repository system specialist for evaluation and possible incorporation. Expect that testing will likely take place in multiple iterations. You may decide that it would be beneficial to have end-users' comments on some parts of the system to uncover aspects of the repository instance that are unintuitive or incomprehensible to them. It is then possible to contact ux@eosc.cz for help with setting up user testing or with deploying a tool to monitor user behaviour on the site.

4. Once the Site is Up and Running

4.1 Monitoring and Evaluation

The relevance and usefulness of the repository can be measured, among other things, by machine-processed statistics. Each repository system allows some data to be tracked at the level of the entire repository, communities or collections, down to individual records, while other data on repository usage can be obtained from external analytical tools. Basic parameters tracked directly by repository systems include, but are not limited to:

- collections/communities – number of registered users, distribution of registered users by their affiliation, number of records in the collection, new records since a defined date, amount of user data, number of metadata views from the collection, number of dataset downloads from the collection, distribution of license types;
- individual records – number of metadata views, number of dataset downloads from a record, data volume.

If other analytics data are relevant for your repository, it is advisable to define and describe them in detail when specifying the needs of the repository, so that the tools for tracking them are included in the development of the repository from the beginning.

If it will be possible to deposit data to the repository without detailed individual control of each record (e.g. using automatic checking of input fields), it is recommended to perform regular quality control of the data and metadata.

For repositories that are linked to external systems (e.g. metadata models, controlled vocabularies), it is essential to continuously monitor and evaluate changes of their source data.

Repositories provide different services to different user groups, and as the social, technical and disciplinary context is constantly changing, it is also appropriate to continuously collect direct feedback from different user groups, evaluate it, and then use it for iterative repository adjustments.

4.2 Communication with the Public

Once the repository has been handed over to the user group, the documentation has been completed and communication channels for user support have been set up, it is appropriate to inform the professional and lay public about the newly available repository. It is recommended to use the services of the public relations departments at institutions with links to the repository – these departments usually have a network of formal and informal relationships and procedures in place for effective dissemination of information about scientific outputs. You can also contact the EOSC CZ External Relations Department (pr@eosc.cz). Members of the user community can then, of course, disseminate jointly prepared messages through their personal channels.