

Conditions for Creating New and Modifying Existing Domain Repositories

L. Matyska and the EOSC-CZ, CARDS, and NRP Projects
version 3.2

1 Terminology

The *National Repository Platform (NRP)* is a distributed system for creating repository instances, developed primarily within the NRP and IPs EOSC-CZ projects with the support of the IPs CARDS.

A *repository instance* (or simply *repository*, where no ambiguity arises) refers to a specific repository of a scientific group or institution. Examples of repository instances include a catch-all data repository or the “repository of fish genomic data”.

A *repository software system* refers to the software package on which repository instances are built. Within the NRP, three supported core repository systems are available: CESNET Invenio, CLARIN DSpace, and ASEP/ARL. (Do not confuse the terms *repository instance* and *repository system*.) Repositories may also be built on other, alternative repository systems. Examples of alternative repository systems include Islandora (used in the pilot HUME Lab repository), or a custom system on which a repository instance is built for the specific needs of a particular scientific group (such as the pilot repository for biodiversity herbaria).

In addition to repository systems, the NRP also provides the infrastructure for running them – covering layers from hardware to data storage (via the S3 protocol) and application runtime environments in containers (Kubernetes). For the purposes of this document, it is not necessary to further distinguish these layers; practically, they represent the availability of *S3 and Kubernetes as services* for operators of individual repository systems as well as the repositories themselves within the NDI. We will refer to this layer as the *data storage and application runtime environment*, or, in a more technical context, as *S3 + Kubernetes*.

A specific role is held by the *National Metadata Directory (NMA)*, which is a repository instance designated exclusively for the automated aggregation of metadata from all repository instances. The NMA is developed and operated under the IPs EOSC-CZ project. Metadata exported to the NMA must have a permissive license that allows unrestricted harvesting and distribution (effectively equivalent to CC0²); note that stating this requirement in terms of licensing does not have the goal to trigger discussion whether metadata is a specific object with a license²).

The *National Catalogue of Repositories (NKR)* has a specific role in the infrastructure; it will register repositories and their characteristics and will be established under the IPs CARDS project.

2 High-Level Architecture of the NRP

For the purposes of this document, the architecture of the NRP can be simplified into four layers that provide services one to another:

1. At the highest level are the end users of the repositories. These users are typically only interested in using specific repositories (for uploading or downloading data), or in

searching for relevant records within the repositories or the National Metadata Directory.

2. The end users are served by repository administrators, who are primarily responsible for the operation and configuration of the repositories. Repository administrators either use a standard implementation of a repository system (and thus rely on the services of the third layer), or operate a custom system, in which case they typically access the services of the fourth layer directly.
3. The operation and development of standard repository systems is ensured by the infrastructure, which provides these systems to repository administrators as a service (within the limits described in this document).
4. At the lowest level, the infrastructure provides the operation of the technical foundation of the NRP, specifically storage and the environment for running applications/containers (S3 and Kubernetes).

3 General Principles of the NRP

Within the NRP, we understand a *repository* (repository instance) according to the following definition: “A repository is the technical, personal, and procedural provision of a long-term storage for the preservation and publication of citable digital objects.”

All repositories within the NRP meet this definition. In particular, they must contain citable records and provide both a web interface and an API for machine access.

A citable record refers to the reliable preservation of a clearly identified digital object (e.g., a dataset). This involves two fundamental components: object identification and a guarantee of immutability.

Object identification typically means assigning a persistent identifier (commonly a DOI). In case of repositories that behave more like databases, an appropriate system for unambiguous identification of individual records with equivalent properties is used.

The guarantee of immutability is a precise description of which modifications are allowed once a record has been finalized. Generally, the data in a finalized record should not be altered. In the case of metadata, it may be appropriate to allow the addition of references to related items, such as “UsedBy” or “ObsoletedBy” relationships. Corrections should primarily be implemented through versioning. The purpose of making the record citable is to ensure reproducibility of scientific results, providing guarantees that the data is still the same as, e.g. referenced in the literature. This prevents situations in which different users rely on the assumption that a given dataset is still the same, while it has in fact been modified—even if only by a minor correction.

This does not imply that the NRP must enforce a single, uniform policy on corrections and modifications across all repositories. Such policies should be set according to the needs and customs of each user community. Nevertheless, every repository *must* define its own policy aiming to minimizing allowed changes, preferring mechanisms like versioning or separately identified “patch” bundles for larger datasets. The repository’s policy must explicitly delimit and precisely describe the scope of permitted changes. A system that allows arbitrary changes to closed records without any of the guarantees described above *cannot* be considered a repository.

As a side note, it is worth mentioning that a technical repository instance can, in some repository systems, be divided into logical components (usually called *communities*), which have independently configured user access controls and workflows over a shared set of

metadata models. A *community* within a repository therefore behaves like an independent repository (in the sense used in the project call) and should be regarded as such in all respects.

From an organizational perspective, it may make sense to aggregate thematically related repositories into shared technical instances. We recommend consulting with the repository system specialists—whose contact information is available on the EOSC website—to determine appropriate granularity.

Wherever this document refers to a *repository*, it naturally also applies to a *logical repository* within a shared technical instance.

For completeness, the term *collection* is also noted. In Invenio, a *collection* refers to a “named search result,” i.e. a selection of records based on predefined criteria (e.g. “all records of authors from a certain group”).

Any additional systems operated within the NRP must be directly related to the purpose of the repository platform (for example, the Data Stewardship Wizard – DSW – supported within the NRP). Hardware and personnel capacity for such tools is planned within the NRP.

Conversely, the NRP does not provide capacity for general storage of unannotated data for end users, nor does it offer computing resources for end users. However, the NRP will ensure basic connectivity to the computing infrastructure of e-INFRA CZ (primarily MetaCentrum and IT4Innovations). The environment for data storage and running applications within the NRP must serve exclusively the purposes of the NRP.

In addition to the repository software systems themselves and the pilot repository instances implemented on them, the NRP also includes auxiliary services as an integral part of its infrastructure—most notably authentication and authorization infrastructure (AAI), data transfer tools, and similar components.

The second fundamental principle of operating NRP systems is that all repositories being developed must, to the greatest extent possible, use the supported standard systems operated within the platform. This is the only way to ensure proper support from the EOSC-CZ, NRP, and CARDS projects. If, for any reason, these repository systems are not suitable—for example, the pilot herbarium repository, which is an established tool used by a broad professional community across Europe and operates on its own repository infrastructure while utilizing the NRP's S3+Kubernetes environment—the NRP allows for the development of custom solutions. In such cases, however, the creator and administrator of the repository must possess sufficient technical expertise and personnel capacity to operate and maintain such a solution and cannot expect additional support from the NRP.

Furthermore, even in such scenarios, the repository (or custom repository system) must be connected to the standard AAI interface, provide an appropriate API, and establish integration with the National Metadata Directory (NMA), data transfer services, and any other tools gradually deployed within the NRP—tools that will be automatically integrated with officially supported repository systems within the NRP project.

In all cases, repositories should be established in consultation with the relevant domain-specific expert working group of EOSC, which should shape recommendations regarding appropriate structure, granularity, and domain-specific metadata models for repositories in different fields.

Since operating a repository inevitably creates requirements for setup and ongoing maintenance, repositories should not be overly narrowly specialized (for example, “a repository for a faculty department” or “a repository for photographs from the Dolní Dvorská

archaeological site, 1960–1980”), as such narrowly scoped repositories would be very difficult or even impossible to ensure their long-term sustainability.

4 Basic Use Cases of the NRP

To provide a basic overview of the service levels available within the NRP and IPs EOSC-CZ projects, we distinguish three basic use cases of the NRP:

1. Establishment of a repository using standard repository systems, i.e., as an instance of CESNET Invenio, CLARIN DSpace, or ASEP/ARL (corresponding to point B of the OS II call—i.e., the creation of new repositories).
2. Establishment of a repository using alternative repository systems in justified cases (also corresponding to point B of the call).
3. Integration of an existing repository currently operated outside of the NRP systems (corresponding to point A of the call, which applies exclusively to existing repositories).

General notes on role assignment in repository administration and operation:

Whenever this document refers to the assignment of a certain role, it does not imply any assumptions about required staffing capacity or number of individuals. However, for the sake of redundancy and continuity, it is strongly recommended that each role be covered by more than one person. At the same time, it is assumed that one person may perform multiple roles in relation to the repository and the NRP. Parts of the associated responsibilities can, of course, be delegated to other individuals—for example, the repository administrator may delegate the roles of curator, reviewer, etc.

It is expected that newly created repositories will undergo a pilot phase to set up their configuration and produce the necessary supporting documentation and then enter production.

Unless otherwise specified, all described roles must exist and be staffed for the entire duration of production operation of the given repository. The repository administrator must be available throughout the entire lifecycle of the repository.

Likewise, all required documentation must be maintained throughout the production operation of the repository.

4.1 Establishing a Repository Using Core Repository Systems

The NRP provides a repository for a scientific community or an institution as a service. In general, this is covered by following steps:

1. Consultations related to the selection of an appropriate repository system (from the supported core systems: CESNET Invenio, CLARIN DSpace, ASEP/ARL), including methodological and analytical support to define the needs and expectations of the user community.
2. Consultations regarding the selection (or development) of a discipline-specific metadata model, with consideration of its feasibility in the respective repository systems.
3. Consultations concerning interoperability/mapping of the discipline-specific profile to the core metadata model for the National Metadata Directory (NMA).

4. Creation of the repository instance based on the aforementioned analysis (including functionality testing in collaboration with the repository administrator). In cases where custom components are developed by the user group, support is provided in the form of documentation and consultations.
5. Operation of all NRP layers from hardware (including reliable data storage) to the setup of the relevant repository instance. For repositories with custom user-developed components, the NRP provides an environment for their operation, along with documentation and consultation support for such an operation mode.
6. This includes full integration into all required systems, in particular (for these activities, no special additional capacity is expected from the repository administrator, unless explicitly stated otherwise):

- a. Deployment of standard metadata profiles and their registration (excluding specific schemas beyond the capabilities of the standard repository systems). Standard metadata profiles include the CCMM profile (Czech Core Metadata Model, <https://github.com/techlib/CCMM>) and those derived from it by adding domain-specific elements (typically under a few dozen elements and without complex dependencies). For larger metadata models, involvement of the user group is expected.

Note: each repository system has its own limitations in terms of complexity of model definition. Repository administrators must select a model and provide a standardized description that meets the specific requirements of the repository system (e.g. JSON in YAML for Invenio). The NRP will then set up the repository with this model and requires cooperation from the repository administrator/technical contact for testing and fine-tuning. For more complex models that require additional modifications, repository administrators must have sufficient technical and personnel resources to implement the necessary adjustments to the repository system or its interface. These activities and associated costs are not covered by the NRP, which only provides capacity for using pre-prepared standardized procedures.

- b. Technical setup of metadata harvesting into the NMA, in accordance with the core metadata model.
- c. Technical configuration of the data deposition workflow.
- d. Integrated persistent identifier (PID) assignment functionality, along with available methodological and administrative support (see identifikatory.cz).
- e. Technical integration with the e-INFRA CZ AAI systems, including the configuration of user roles and groups within the repository instance.
- f. Prefabricated materials for user documentation creation that cover common parts provided by the infrastructure. These also serve as a template for creating documentation for a specific system. To create a documentation instance, the existing website <https://docs.nrp.eosc.cz/> and the e-INFRA CZ documentation system (source documents in Markdown managed in Git) can be used.
- g. User support for repository administrators, L2 and L3 support for escalated requests from end users of the repositories (but not L1 end-user support). For user support by the repository administrator, a queue in the RT system can be provided, along with the creation of an appropriate email alias.
- h. Integration into the national e-infrastructure environment, specifically the availability of tools for transferring data between the repository, general storage, and computational resources within the e-INFRA CZ infrastructure (MetaCentrum, IT4I, CESNET storage, etc.).
- i. Configuration of repository system logging into the central logging system.

- j. Inclusion in cybersecurity monitoring (CESNET-CERTS, FTAS), performing security and particularly penetration tests of systems. Incident handling and collaboration with the cybersecurity team.
 - k. Collection of statistical data on the operation and usage of the system.
 - l. Operational monitoring.
7. We expect that repositories created as instances of the core NRP repository systems will meet the technical and organizational requirements for the infrastructure of trusted repositories from the perspective of grant providers in the Czech Republic. Once the qualification of a trusted repository for these purposes is obtained, meeting the conditions for the repositories themselves (data stewardship, etc.; these rules are currently being prepared) will suffice, with the entire technological foundation and guarantees provided by the NRP.
 8. General service conditions will be prepared for the operation of the NRP. Repositories can be established by individuals demonstrably connected to the academic community in the Czech Republic. However, for trusted repositories, it is expected that there will be contractual arrangements with the relevant user institutions, i.e. institutions related to the user community that established the repository.

The primary contact for requests to establish repositories is the specialist for the respective system. The specialist acts as a system integrator, conducting a basic analysis with the repository administrator and data curator and arranging consultations with specialists in cases where the user's request is more complex.

In such a case, the user community is expected to:

1. Establish the role of the repository administrator. The repository administrator is the partner of the infrastructure operator (primarily the NRP project) for agreement on the configuration of the repository in all the points described below. The repository administrator also has primary responsibility for the data stored in the repository and all the settings described below (which can, of course, be delegated to other individuals as needed). The repository administrator is also informed about operational events of the repository and repository systems or the entire NRP (updates, outages, etc.). The repository administrator is also responsible for collaborating with the cybersecurity team and reporting cybersecurity incidents if they occur at the level of the repository they manage.
2. Establish the role of a data curator, who sets general rules for the data stored in the repository (e.g. regarding retention periods based on record type) and makes decisions on specific datasets (e.g. addressing deletion requests). The curator also acts as a metadata specialist, ensuring metadata harmonization within the repository in accordance with established domain-specific metadata profiles and ensuring metadata interoperability with other systems (especially NMA). They participate in setting the domain-specific metadata profile. These roles can be separated if necessary.
3. Define the domain-specific metadata profiles available in the repository (in collaboration with IPs CARDS and specialists for the respective systems), utilizing tools for metadata profile management. However, implementing a metadata profile in the repository requires serialization of the profile into the repository configuration and the creation of a user interface for metadata entry and search. The infrastructure will handle this for standard metadata models. For larger and more complex metadata models, the user community/repository administrator must ensure the capacity for implementation

on their side (the infrastructure will provide documentation and potential consultations). Define the metadata schema elements that will be exported to the NMA (mapping to the core metadata model). For metadata profiles that exceed the capabilities of core repository systems, collaborate on implementing their support.

4. Determine the list of licenses available in the data deposition process.
5. Determine the list of supported data formats that can be stored in the repository (this can even be “any”).
6. Define the workflow for data deposition (e.g. the record approval process).
7. Define the workflow for data access (from open access to a process involving approval by an ethics committee or an access control committee).
8. Define the roles of user groups in the repository, such as regular submitter, curator, approver in different workflow parts. Link these roles to groups of individuals in the EOSC CZ AAI.
9. Create user documentation for the repository. NRP will provide prefabricated components describing the basic functioning of individual repository systems within NRP and recommended documentation templates. However, the user documentation must describe the metadata models used, the repository's deposition workflow, the search interface, descriptions of group roles within the repository, and similar aspects for specific repository instances. Tools from e-INFRA CZ/NRP can be used for documentation creation, which is done in Markdown format and maintained in GIT.
10. Describe the repository policy, especially when a record is considered closed—this policy should clearly define when a stored record in the repository is considered finalized and what changes are permissible in finalized records (e.g. “adding a metadata item with a link to a correction or usage of the record, but nothing else”). The repository policy description can also be prepared as certification documentation (e.g. for Core Trust Seal).
11. Provide user support (first level) for end-users of the repository. The NRP infrastructure will provide optional tools for tracking user requests. Other levels of user support, which include escalated requests requiring intervention from infrastructure administrators and support for the repository administrators themselves, are covered by NRP resources.
12. Submit information to the National Repository Catalogue—register the repository and its parameters in the National Repository Catalogue (NKR) and send updates in case of changes. The records include metadata profiles (schemas) and the controlled vocabularies and ontologies used. Submission of information about the repository to NKR should be automated via OAI-PMH or API.

4.2 Establishing a Repository Operated on NRP Resources Without Using Core Repository Systems

In necessary cases (where, even after consultation, it is not possible to utilize any of the three provided and supported repository systems), it is possible to operate repositories built on alternative implementations of repository systems directly within the NRP environment.

In such cases, the user community managing such a repository will essentially receive access from NRP to an environment for data storage and application operation (specifically S3 and Kubernetes). However, they will be responsible for all activities and bear all associated costs related to the installation and integration of the alternative repository system and the specific instance of the repository or repositories (if they wish to operate more than one) within the NRP environment and its operation.

The repository administrator must also take responsibility for:

1. All items described as the administrator's responsibility in the case of using core repository systems.
2. Ensuring the installation and operation of the repository and the appropriate software infrastructure (typically an alternative repository system) within the application environment, utilizing the NRP's storage layers.
3. Deploying domain-specific metadata profiles and registering them (integrated into the metadata profile registration system), harmonizing metadata within the given repository, and ensuring metadata interoperability with other systems, especially NMA (this can also be handled by the repository curator).
4. Selecting and implementing the assignment of persistent identifiers from the set of standardly supported ones, configuring the assigned ranges.
5. Setting up the technical configuration for metadata harvesting into the NMA according to NMA's requirements, in compliance with the core metadata model (CCMM).
6. Technically configuring the workflow for data deposition and access control to the data.
7. Configuring technical integrations with the e-INFRA CZ AAI systems and defining the roles of user groups in the repository instance.
8. Creating user documentation.
9. Creating documentation for system administration and operation.
10. Providing user support for end-users at all levels, from L1 to L3, except for requests directly related to the operation and configuration of the application environment and data storage (S3 + Kubernetes).
11. Providing tools for integration into the national e-infrastructure environment, especially for transferring data between the repository, data storage, and computational resources in the e-INFRA CZ infrastructure.
12. Configuring the logging of repository systems into the NRP's central logging system.
13. Establishing cybersecurity oversight (CESNET-CERTS, FTAS). Collaboration in conducting security and especially penetration tests for the repository and systems directly related to it. Handling incidents and collaborating with the cybersecurity team. Mandatory reporting of cybersecurity incidents.
14. Compliance with standard service conditions and defining additional conditions in cooperation with NRP compliance.
15. Collecting statistical data on the system's operation and usage.
16. Operational monitoring.

The repository administrator must also ensure sufficient capacity of system administrators and other personnel for the stable operation of the system.

4.3 Integration of an Existing Independently Operated Repository into the NRP/National Data Infrastructure (NDI)

This use case covers situations where an existing repository, operated as an independent entity, is to be connected to the NRP/NDI environment.

The administrator of a repository operated outside of NRP has full responsibility for its operation, from hardware to the repository service itself. To consider such a repository as "connected to NRP/NDI," it must generally satisfy the same conditions as repositories in alternative implementations directly operated within NRP. However, the repository administrator is also responsible for operating hardware resources, system management, and complete user support. This concerns not only the repository system itself and its data storage but also other components necessary for its operation. The administrator of such a repository

must ensure functionality equivalent to that provided by NRP. There are no restrictions on how this is achieved and implemented, but all functionalities must be provided in an adequate form (adequacy is primarily determined by the administrator but the administrator may be asked to provide justification of the measures for the NRP administrators).

The basic minimum required connection to the NDI environment consists of:

- Connection to NMA and providing metadata in accordance with the core metadata model
- Submitting information to the National Repository Catalogue
- Connection to the AAI operated by NRP
- Defined APIs for data transfer
- Assigning PIDs (not necessarily DOIs)

These repositories are also subject to all provisions concerning administrators, roles, licensing settings, and other policies and conditions related to the operation of the repository. The repository must also have appropriate cybersecurity infrastructure and at least a basic level of monitoring to ensure operational quality and data collection for statistical purposes. The repository must be logged, which is essential for analysing cybersecurity incidents. The repository administrator is also responsible for sufficient compliance with legal and other regulations, depending on the nature of the data and operation.

Within technical possibilities and capacities, NRP services are available to the administrator of the existing repository, and it is strongly recommended to make as wide use of standard services as possible. Specific configurations will need to be addressed for each specific repository. Depending on the specific technical situation, a combination of an independent repository solution as described in this section with the use of certain NRP services can be considered (for example, a model where such a repository uses S3 in NRP as one of its data storage solutions).

5 Final Remarks

The primary goal of this document is to provide a tangible understanding of the level of service that NRP offers for individual use cases. Given its structure, we did not consider it useful to integrate a temporal perspective into the document; the timeline for the availability of individual services can be found in the project's timeline.

6 Annex: Structured Aggregation of Requirements into Tables

6.1 Conditions for Establishing a Repository Using Core Repository Systems

Condition	Establishing the role of repository administrator
Consultant	tech-support@eosc.cz
Expected Responsibilities of the Repository Administrator	<p>The repository administrator is a partner of the infrastructure operator (i.e., primarily the NRP project) for the agreement on the configuration of the repository in all the points described below. The repository administrator also bears primary responsibility for the data stored in the repository and for all the settings described below (which, of course, can be delegated to other individuals as needed).</p> <p>The repository administrator collaborates with the infrastructure operator in deploying and testing the repository instance. The repository administrator is also informed about operational events related to the repository and repository systems or the entire NRP (such as updates, outages, etc.). The repository administrator is also responsible for collaborating with the cybersecurity team and for reporting cybersecurity incidents if they occur at the level of the repository they manage.</p>
Description/Comments	A mailing list of a team that handles these matters is listed as “consultant”.

Condition	Establishing the role of data curator
Consultant	Hana Vyčítalová metadata@techlib.cz Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz
Expected Responsibilities of the Repository Administrator	<p>Establishment of the role of a data curator, who defines general rules for the data stored in the repository (e.g. in terms of retention periods based on record type) and makes decisions about specific datasets (e.g. data deletion requests handling). The curator also acts as a metadata specialist, meaning they are responsible for the harmonization of metadata within the given repository in accordance with established domain-specific metadata profiles, and for ensuring metadata interoperability with other systems (particularly NMA). The curator also contributes to the definition of the domain-specific metadata profile.</p>
Description/Comments	<p>The data curator may be the same person as the repository administrator, but it is recommended to separate these roles and assign them to different individuals (the workload will depend on the size and complexity of the repository, the data it contains, and the level of services the repository is expected to provide). It is also possible to separate the roles of curator and metadata specialist.</p>

Condition	Establishing a thematic metadata profile
Consultant	Jakub Klímek jakub.klimek@matfyz.cuni.cz David Antoš david.antos@cesnet.cz (technical implementation of export to NMA) Hana Vyčítalová metadata@techlib.cz (interoperability with the core metadata model)
Expected Responsibilities of the Repository Administrator	Definition of domain-specific metadata profiles that will be available in the repository (in cooperation with IPs CARDS and the relevant system specialists) using the metadata profile management tool. Definition of metadata schema elements that will be exported to the NMA (mapping to the core metadata model). In the case of metadata profiles that exceed the capabilities of the core repository systems, collaboration on the implementation of their support is also required.
Description/ Comments	

Condition	PID configuration
Consultant	Hana Heringová identifikatory@techlib.cz
Expected Responsibilities of the Repository Administrator	Selection of assigned persistent identifiers from the set of standardly supported ones, setting of assigned ranges. Responsibility for the used PIDs, PID integration and adherence to best practice in working with PIDs in accordance with National PID Centre (NTK) and PID providers.
Description/ Comments	This mainly involves membership in international organizations and consortia that set the rules for handling PIDs – for example, assigning DOIs: existing projects will provide both methodological and technical assistance, but the formal responsibility lies with the operator of the given repository. The repository operator must become a member of the DataCite consortium, not the providers of the repository systems (CESNET/CUNI/KNAB).

Condition	Licences
Consultant	Pavel Straňák stranak@ufal.mff.cuni.cz
Expected Responsibilities of the Repository Administrator	Ensuring that each dataset gets a license during the publication process.
Description/ Comments	Each dataset available in the repository will have a license specified in the metadata.

Condition	Data deposition workflow
Consultant	Pavel Straňák stranak@ufal.mff.cuni.cz
Expected Responsibilities of the Repository Administrator	Establishment of a workflow for data deposition (e.g., record approval process). Establishment of a data access workflow (from open access to a process involving approval by an ethics or access control committee).
Description/ Comments	

Condition	User groups roles
Consultant	Peter Lényi lenyi@ics.muni.cz
Expected Responsibilities of the Repository Administrator	Establishing and implementing roles and permissions within user communities in the repository instance, e.g. common depositor, curator, approver. Linking these roles to groups of people in e-INFRA CZ AAI. Setting and implementing additional rules to control access to data for user groups, e.g. sharing specific data with selected users only.
Description/ Comments	

Condition	Repository user documentation
Consultant	Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz Jan Kolouch jan.kolouch@cesnet.cz Pavčina Špringerová springerova@ics.muni.cz / ux@eossc.cz
Expected Responsibilities of the Repository Administrator	Creating user documentation for the repository. Prefabricated components describing the basic operation of core repository systems in the NRP and recommended documentation templates will be available for this purpose by NRP. However, the user documentation must describe, for specific instances of the repository, the metadata models used, the submission workflows, the search interfaces, the roles of different user groups within the repository, and similar aspects.
Description/ Comments	

Condition	Repository policies
Consultant	Jan Kolouch jan.kolouch@cesnet.cz Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz
Expected Responsibilities of the Repository Administrator	Description of the repository policy, especially regarding when a record is considered closed. This policy must clearly define when a saved record in the repository is deemed finalized, and what changes are permitted in finalized records (e.g., "adding a metadata field linking to a correction or use of the record, but nothing else").
Description/ Comments	

Condition	Establishing first level (L1) user support for the repository
Consultant	Jan Kolouch jan.kolouch@cesnet.cz
Expected Responsibilities of the Repository Administrator	Providing user support (first level) for end users of the repository. The NRP infrastructure will offer (optional) tools for user requests tracking. NRP funding also covers higher levels of user support, which include escalated issues requiring intervention from infrastructure administrators, as well as support for the repository administrators themselves.
Description/ Comments	

Condition	Data submission to the National Catalogue of Repositories
Consultant	Petra Černohlávková petra.cernohlavkova@techlib.cz
Expected Responsibilities of the Repository Administrator	Registration of the repository and its parameters in the National Catalogue of Repositories (NKR), along with continuous updates in case of any changes. The registration also includes metadata profiles (schemas) and the controlled vocabularies and ontologies used.
Description/ Comments	Sending repository information to the NKR should be automated via OAI-PMH or API.

6.2 Conditions for Establishing a Repository Based on NRP Resources Without Using Core Repository Systems

A repository administrator using a custom implementation of repository software must meet all the conditions outlined in the previous section, as well as the following requirements related to the corresponding repository system:

Condition	System installation and setup
Consultant	tech-support@eosc.cz (David Antoš david.antos@cesnet.cz)
Expected Responsibilities of the Repository Administrator	Ensuring the installation and operation of the repository and the corresponding software background (usually an alternative repository system) in the application runtime environment and using NRP storage layers.
Description/ Comments	

Condition	Implementation of domain metadata profiles
Consultant	Jakub Klímek jakub.klimek@matfyz.cuni.cz
Expected Responsibilities of the Repository Administrator	Deployment of domain metadata profiles and their registration (in integration with the metadata profile registration system).
Description/ Comments	Domain models should be mappable to the core metadata model.

Condition	Harvesting metadata to NMA
Consultant	David Antoš david.antos@cesnet.cz Hana Vyčítalová metadata@techlib.cz (interoperability with the core metadata model)
Expected Responsibilities of the Repository Administrator	Technical setup of metadata harvesting into NMA according to the requirements of NMA using the OAI-PMH protocol (by selecting from supported serializations of metadata format) in accordance with the core metadata model.
Description/ Comments	

Condition	Implementation of a data deposition workflow in a repository
Consultant	Pavel Straňák stranak@ufal.mff.cuni.cz
Expected Responsibilities of the Repository Administrator	Technical setup of data deposition workflow and data access control.
Description/ Comments	

Condition	PID implementation and configuration
Consultant	Hana Heringová identifikatory@techlib.cz
Expected Responsibilities of the Repository Administrator	The selection and implementation of persistent identifier (PID) assignment from the officially supported set, and the configuration of assigned PID ranges. Responsibility for the PIDs used, their integration, and adherence to best practices in working with PIDs, in accordance with the guidelines of the National PID Center (NTK) and PID providers.
Description/ Comments	In the case of alternative repository implementations, the administrator is responsible for the technical implementation of PID assignment within the operated system. The operator of the given repository also bears responsibility for the correct use of PIDs, their integration, and adherence to best practices in working with PIDs, in line with the guidelines of the National PID Center (NTK) and PID providers. If an organization wishes to assign DOIs within the allocation for Czechia, it must become a member of the DataCite consortium.

Condition	Licenses
Consultant	Pavel Straňák stranak@ufal.mff.cuni.cz
Expected Responsibilities of the Repository Administrator	Each dataset accessible in the repository will have a specified license in its metadata. (This will probably be a requirement in NMA, but we are not there yet).
Description/ Comments	Ensuring that each dataset gets a license during the publication process.

Condition	Integration of e-INFRA CZ AAI
Consultant	Peter Lényi lenyi@ics.muni.cz
Expected Responsibilities of the Repository Administrator	Integration of the repository instance with user authentication via e-INFRA CZ AAI, preferably using the OIDC protocol. Definition and implementation of roles and permissions within user communities in the repository instance, e.g., regular contributor, curator, approver. Linking these roles to user groups in e-INFRA CZ AAI. Configuration and implementation of additional access control rules for user groups, e.g. sharing specific data only with selected users. Definition and implementation of the user lifecycle in the repository, including the removal of inactive users.
Description/ Comments	Connection is via OIDC, or SAML can be used in justified cases.

Condition	User documentation
Consultant	Jan Kolouch jan.kolouch@cesnet.cz Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz Pavčina Špringerová springerova@ics.muni.cz / ux@eossc.cz
Expected Responsibilities of the Repository Administrator	Creation of user documentation for end users of the repository, including a basic description of the chosen implementation.
Description/ Comments	

Condition	Internal documentation
Consultant	tech-support@eossc.cz Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz
Expected Responsibilities of the Repository Administrator	Creation of documentation for system administration and operation.
Description/ Comments	We anticipate that the documentation may be needed by the KA2 NRP when collaborating on integration into S3 and Kubernetes environments, as well as when resolving potential operational issues.

Condition	User support setup and operation
Consultant	Jan Kolouch jan.kolouch@cesnet.cz
Expected Responsibilities of the Repository Administrator	Setup of support system for end users at all levels, i.e. L1 to L3, except for requests directly related to the operation and configuration of the environment for running applications and storing data (S3 + Kubernetes). User support must be available throughout the lifetime of the repository.
Description/ Comments	

Condition	Tools for data transfer to and from the national e-infrastructure environment
Consultant	Jiří Sitera sitera@civ.zcu.cz
Expected Responsibilities of the Repository Administrator	Provision and maintenance of tools for integration into the national e-infrastructure environment, especially for data transfers between the repository and data repositories and also computing resources in the e-INFRA CZ e-infrastructure.
Description/ Comments	Depending on the API provided by the repository implementation and the needs of the user community, this may include, for example, integration <ul style="list-style-type: none"> • for CLI tools in computing environments (MetaCentrum, IT4I) • of portal systems, e.g. Galaxy • with Jupyter notebooks • staging to a computing environment running from the repository ("on the button")

Condition	Operational logging
Consultant	Radko Krkoš Krkos@cesnet.cz Andrea Kropáčová andrea.kropacova@cesnet.cz
Expected Responsibilities of the Repository Administrator	Configuration of the repository systems logging to the central NRP logging system, with respect to the setting of logging parameters of the operated system and application services to support machine processing for analytical processing of operational records Application of configuration changes according to the CESNET SOC recommendations. The transport protocol (API) is Syslog (RFC 5424). Monitoring of completeness and consistency of logging.
Description/ Comments	Application of configuration changes during the lifetime of the repository (not expected often) and monitoring are continuous activities.

Condition	Cybersecurity
Consultant	Andrea Kropáčová andrea.kropacova@cesnet.cz Radko Krkoš Krkos@cesnet.cz
Expected Responsibilities of the Repository Administrator	Notification of the IP addresses on which the repository is operated, contact details of responsible persons, and reporting of planned atypical activities (e.g., large-scale data transfers, high connection volumes, internal penetration or other testing, etc.). Integration of the repository's access component into the FTAS network traffic monitoring infrastructure (configuration of netflow data export). Incident response and cooperation with the CESNET-CERTS cybersecurity team in handling incidents. Mandatory reporting of cybersecurity incidents to CESNET-CERTS. Systematic monitoring of technical vulnerabilities and response to them (vulnerability management).
Description/ Comments	IH collaboration, CERTS reporting and vulnerability monitoring are ongoing activities throughout the life of the repository. It is essential that the repository is run by an administrator who is familiar with the basic principles of OS management and logging.

Condition	Compliance
Consultant	legal@eosc.cz Marcela Pospíšilová pospisilova@cesnet.cz
Expected Responsibilities of the Repository Administrator	Setting compliance with standard terms of service, defining additional conditions in coordination with NRP compliance.
Description/ Comments	These are primarily procedurally legal activities.

Condition	Traffic statistics
Consultant	Jan Kolouch jan.kolouch@cesnet.cz Pavčina Špringerová < springerova@ics.muni.cz / ux@eosc.cz >
Expected Responsibilities of the Repository Administrator	Setting up a system to collect statistical data about system operation and usage. Collection and provision of this data throughout the lifetime of the repository.

Description/ Comments	We do not prescribe a specific API for this system, rather we assume an agreement with the alternative repository administrator on what is included in their chosen repository variant.
--------------------------	---

Condition	Operational monitoring
Consultant	Jan Kolouch jan.kolouch@cesnet.cz Pavčina Špringerová < springerova@ics.muni.cz / ux@eossc.cz >
Expected Responsibilities of the Repository Administrator	Setup of repository services monitoring.
Description/ Comments	This is a requirement on the repository, but the complexity is related to the fact whether the alternative repository system chosen by the repository administrator contains a monitoring system or not.

Condition	Data submission to the National Catalogue of Repositories
Consultant	Petra Černohlávková petra.cernohlavkova@techlib.cz
Expected Responsibilities of the Repository Administrator	Registration of the repository and its parameters in the National Catalogue of Repositories (NKR), continuous sending of updates in case of changes. The registration also includes metadata profiles (schemas) and used controlled vocabularies and ontologies.
Description/ Comments	Sending repository information to the NKR should be automated via OAI-PMH or API.

The list above, considering the comments in its text version, can also be used for cases of connecting an independently operated repository to the NDI environment.

The rest of the document consists of the initial assignment from the Charles University team preparing the OS II, for reference:

Conditions for creating new and modifying existing subject repositories in the Open Science II project

Meeting the connection requirements to the NRP and NMA and their integration into the jointly built NDI

Charles University asks the IPs CARDS and NRP projects investigators to complete the basic parameters for the preparation and definition of outputs prepared in the Open Science II (OS II) project.

The document will be made available to all potential beneficiaries (regardless of the intended status of "partner" or "mini-project beneficiary") in the structure of the EOSC thematic (disciplinary) WGs.

Delivery date: 11 October 2024, consultation for filling up: irena.velebilova@ruk.cuni.cz

6.3 Introduction

The OS II call imposes an obligation on the applicant (p. Open Science II, https://opjak.cz/wp-content/uploads/2024/06/Vyzva_Open_Science_II_web.pdf, page 3 onwards):

Activity 2. Development of existing repositories, building new disciplinary/interdisciplinary repositories and data consolidation

*The aim of the activity is the development, consolidation, or creation of thematic (disciplinary or interdisciplinary) research data repositories, **their mandatory connection to the National Repository Platform (hereinafter referred to as "NRP") and the National Metadata Directory (hereinafter referred to as "NMA")** and their integration into the jointly built NDI in order to consolidate and make available research data in the Czech Republic. Specifically, the subject of the activity is*

A. development of repositories existing outside the NRP at the time of the call: expansion of the communities using these repositories; visibility of the repositories beyond the primary target group (within and outside the discipline/topic cluster)

B. the creation of new disciplinary and cross-disciplinary repositories: consolidation of user communities that have been operating without properly managed repositories; visibility of repositories beyond the primary target group (within and outside the disciplinary/topical cluster). The creation of new disciplinary/interdisciplinary repositories is expected primarily in the form of repositories in the NRP; the creation of a new self-managed repository must be explicitly justified

6.4 Purpose of data collection

- Charles University assumes that the standardisation of NRP, NDI and NMA parameters is one of the goals/objectives of the CARDS and NRP project and are not available now.
- For the preparation of the OS II project, it is necessary to define the framework and basic conditions now, because they influence the number and content of the outputs to be prepared in OS II.
- If the CARDS and NRP investigators assume that the OS II project investigators will also be involved in the standardisation process (e.g. by providing a participant in the working

group preparing the standard, by opposing proposals, by verifying the prototype, ...), then such an assumption should be included in the description of the condition as one of the activities that the output "subject repository" (hereinafter referred to as the subject repository) should take into account.

- The goal of the data collection is to create a "checklist" of conditions for the OS II subject repository to be ready to meet the NRP and NMA linkage requirement and to be included in the NDI.
- Please elaborate each condition separately in the structure below (ideally in the attached table).
- Examples of "conditions" for OS II branch repositories: e.g. authentication; securing the position of repository curator, system administrator, ...; securing the operation of helpdesk L1, L2; L3, verification of the design of a standard for a service or microservice, acquisition of infrastructure - HW, SW, elaboration of methodology and instructions for the helpdesk; use of persistent identifiers: DOI, Handle, ORCID, cybersecurity standards, metadata schema format...

6.5 Structure of a document template (explanatory notes):

- condition: explicit short title of the condition.
- consultant of the condition: name, email, we assume that it will be the investigator of the key activities of the NRP; CARDS.
- description of a condition:
- factual description: structured, factually specific text that clearly and concretely describes the subject matter of the condition. Including a proposal, if any, for how to meet the condition (regardless of whether these conditions are provided by the field repository investigator from OS II project funds or from other funds).
- existing standard: if there is a reference to a published standard for the condition (e.g. AAI, then provide the reference).
- what will be ready to meet the condition from the NRP, CARDS side: what specifically can the subject repository expect to be ready to meet the condition in the NRP; CARDS output, i.e. what it does not need to plan resources for.
- repository type: specify whether the condition applies*:
 - A = development of repositories existing outside the NRP;
 - B = creation of a new repository within NRP instances**;
 - C = creation of a new repository outside the NRP instance**.

*All types of repositories can be listed.

**Note: Invenio; Clarin DSpace, ASEP and Biodiversity Data Repository are considered as NRP instances.

- project: from which project; CARDS, NRP; other (fill in).
- project output: name of the related project output: output of CARDS/NRP related to the defined condition (see OS project outputs and linkages table: <https://www.eosc.cz/media/3729597/vystupy-a-provazby-projektu-open-science.xlsx>). The row number from this output table can also be provided.
- deadline for OS II: date when CARDS/NRP output will be ready for OS II investigators (if known).
- Note: a space that must never be missed

condition			
consultant	name	email	
factual description of the condition			
existing standard			
what will be prepared by NRP/ CARDS to meet the condition		approximate date of preparation completion to meet the condition	
repository type	A = development of repositories existing outside the NRP	B = creation of a new repository within NRP instances	C = creation of a new repository outside the NRP instance
project	CARDS	NRP	other:
- project output			
- OS II deadline			
note			