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

L. Matyska and EOSC-CZ, CARDS and NRP project teams

(version 3.0)

Terminology

The National Repository Platform (NRP) is a distributed system for repository instances creation; built mainly within the NRP and IPs EOSC-CZ projects with support from IPs CARDS.

By *a repository instance* (or for short, a repository, unless there is a risk of confusion) we mean a specific repository of a scientific group or institution. An example of a repository instance is the catch-all repository for data: https://data.narodni-repozitar.cz/ or the "vertebrate genomic data repository".

By *a software repository system,* we mean a software package on which repository instances are built. There are three *basic* repository systems supported by the NRP: CESNET Invenio, CLARIN DSpace, and ASEP/ARL. (Do not mix the terms “repository instance” and “repository system”.) Repositories may be built on other, *alternative* repository systems. An example of an alternative repository system is Islandora (used in the pilot repository HUME Lab), or a custom system on which a repository instance is built to meet the specific needs of a particular scientific group (the pilot Biodiversity Herbarium repository is an example).

In addition to repository systems, NRP also provides the environment their operation, i.e. layers from hardware to the data storage environment (by S3 protocol) and the environment for running applications in containers (Kubernetes). For the purposes of this document, there is no need to differentiate this layer in further detail; in practice, they imply the availability of S3 and Kubernetes as a service to the operators of the individual repository systems and the repositories themselves in the NDI. Further, we will refer to this layer as the *environment for data storage and running applications,* or at a more technical level as *S3 + Kubernetes*.

The National Metadata Directory (NMA), which is a repository instance exclusively for the automated aggregation of metadata from all repository instances, has a specific status. The NMA is built and operated within the IPs EOSC-CZ project.

The National Catalogue of Repositories (NKR), which will record repositories and their properties and will be created in the IPs CARDS project, also a specific status too.

General principles of the NRP

In the context of NRP, we understand a repository (repository instance) according to the following definition:

"A repository is a technical, personal and processing support of a long-term storage facility for deposition and publication of citable digital objects."

In general, all repositories within the NRP must fulfil the definition above, i.e. they must contain citable records, have a web interface and an API for machine access.

A citable record is a reliable storage of a clearly identified digital object (e.g. a dataset). This contains two basic components: object identification and the guarantee of the extent of the changes. Object identification is typically the assignment of a persistent identifier (typically e.g. DOI; if the repository is more in the nature of database records, then a suitable system for unambiguous identification of a specific record with similar properties). A change extent guarantee is a precise description of what changes are allowed to be made to the finalized record. Typically, the record data should not be changed after finalization; for metadata, it usually makes sense to allow the addition of references to related items, such as Used-By or Obsoleted-By relationships. Corrections to records should be made primarily by versioning. The aim of record citation is to ensure the replicability of scientific results. In particular, this is to avoid the situation where different users use a particular dataset expecting that it is still an identical dataset, when the dataset has been changed, even by a marginal correction.

This does not mean that the NRP should have a uniform policy for corrections and changes of repository records. Such a policy must be set according to the needs and practices of the user community. However, the repository must set its policy to minimize the changes allowed, preferring mechanisms such as versioning or storing separately identified patch packages to larger datasets, and so on. The repository policy must explicitly define and describe the scope of allowed changes. A system that allows arbitrary changes to closed records and makes no guarantees in the sense described above cannot be considered a repository.

As an aside, it is worth noting that in some repository systems, the technical instance of a repository can be divided into logical components (usually called communities), which particularly have separately configured user access control and workflow over a shared set of metadata models. A community within a repository therefore acts as a separate repository (in the sense of the project call text) and should be understood as such in all aspects. From the point of view of repository organisation, it may make sense to aggregate discipline-related repositories into joint technical instances; we recommend consulting the appropriate granularity with the operators of the respective system. Wherever we refer to a repository in this document, this includes a logical repository within a joint technical instance.

Any other systems operated in the NRP must be directly related to the purpose of the repository platform (e.g. the NRP-supported Data Stewardship Wizard, DSW). Hardware and staff capacity is planned for such tools in the NRP.

On the contrary, the NRP does not have the capacity of a general storage of unannotated data for end users, nor computing capacity for end users. The NRP will provide a basic connection to the computing resources of e-INFRA CZ (especially MetaCentrum and IT4Innovations), though. The environment for data storage and running applications in the NRP must serve exclusively the purposes of the NRP.

Apart from the repository system software themselves and the pilot repository instances implemented on top of them, ancillary services such as authentication and authorization infrastructure (AAI), data transfer tools, etc. are an integral part of the NRP.

The second basic principle for operating NRP systems is that all repositories that are built must use supported standard systems operating within the platform to the maximum extent possible. This is the only way to ensure proper support from the EOSC-CZ, NRP and CARDS projects. If these repository systems are not suitable for any reason - an example is the pilot repository for herbaria, which is an established tool for a wide expert community in Europe, uses its own repository facilities and uses directly the S3+Kubernetes environment from the NRP – the NRP allows to build its own solution. In this case, the repository creator and administrator must have sufficient expertise and staff capacity to provide a similar solution and cannot expect additional support from the NRP. And, of course, even in such a case, they are obliged to connect the created repository (or the new repository system) to the standard AAI interface, to provide appropriate APIs and interfaces to NMA, data transfer services and possibly other tools that will be gradually deployed in the NRP and where, for directly supported repository systems, this connection will be provided within the NRP project.

In all cases, repositories should be established in consultation with the discipline-specific EOSC Expert Working Group, which should form ideas on the appropriate structure, granularity and disciplinary metadata models for repositories in each discipline. As the operation of a repository inevitably creates demands for both set-up and maintenance, repositories should not be too highly specialised (especially *not a* "departmental repository" or an "Institute photo repository of Dolní Dvorska Archaeological Site 1960-1980", as in such a case it will be very difficult or impossible to ensure their long-term sustainability).

Basic NRP use-cases

For a general overview of the level of services available under NRP projects and IPs EOSC-CZ, which can be followed up by OS II call projects, we distinguish three basic NRP use cases:

I. Building a repository using basic 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 where justified (point B. of the call).
III. Connecting an existing repository that is already operating outside the NRP systems (point A. of the call, which applies only to existing repositories).

General remarks:

- Whenever we talk about the establishment of a role in this material, it does not make any assumptions about the necessary capacity or number of people. However, for the sake of substitutability, it is strongly recommended that each role shall be covered by several individuals. On the other hand, it is assumed that one individual can act in multiple roles in relation to the repository and the NRP. Parts of the respective responsibilities can of course be delegated to other individuals, typically the repository administrator delegates the role of curators, approvers and so on.
- It is expected that the newly created repositories will have a pilot phase for their setup and the generation of the necessary documentation, and then they will go into production phase.
- Unless otherwise specified, all roles described must exist and be staffed for the entire production lifetime of the repository. The repository administrator must be available throughout the whole repository's life cycle.
- Similarly, all required documentation must exist for the entire production lifetime of the repository.

Building a repository using NRP basic repository systems

The NRP provides the establishment of a repository as a service to the scientific community or institution. **In this case, the NRP, IPs EOSC-CZ and IPs CARDS projects provide, and the repository administrator effectively "receives as a service", the following:**

1. Suitability consultations related to the selection of a repository system (from the basic repository systems, i.e. CESNET Invenio, CLARIN DSpace, ASEP/ARL), methodological and analytical support to determine the needs and expectations of the user group.
2. Consultations related to the selection (or creation) of a domain-specific metadata model with regard to its feasibility in individual repository systems.
3. Consultations related to the interoperability/mapping of the domain-specific profile to the core metadata model for NMA.
4. Creating a repository instance based on this analysis (including functionality testing in collaboration with the repository administrator).
5. Operation of all layers of NRP from hardware (including reliable data storage) to the creation and setup of the respective instance.

That includes full integration into all necessary systems, in more detail (these are activities for which we do not expect, unless explicitly stated otherwise, additional capacity on the part of the repository administrator):

a) Deployment of standard metadata profiles [1] and their registration (except for specific schemas beyond the capabilities of the basic repository systems).
b) Technical settings for harvesting metadata into NMA in accordance with the core metadata model.
c) Technical settings of the data deposition workflow.
d) Directly available implementation of the allocation of persistent identifiers, available methodological and administrative support (see identifikatory.cz).
e) Technical setup of links to e-INFRA CZ AAI systems and user group roles in the repository instance.
f) Ready-made materials for creating user documentation.
g) User support for repository administrators, L2 and L3 support for escalated repository end-user requests (but not L1 end-users).
h) Integration into the national e-infrastructure environment, especially the availability of tools for data transfer between the repository, general repositories and computing resources in the e-INFRA CZ e-infrastructure (MetaCentrum, IT4I, CESNET repository, ...).
i) Configure logging of repository systems to a central logging system.
j) Inclusion in cybersecurity surveillance (CESNET-CERTS, FTAS), performing security and especially penetration tests of systems. Incident handling and cooperation with the cybersecurity team.
k) Collection of statistical data on the operation and use of the system.
l) Operational monitoring.

7. We expect that the repositories created as instances of the NRP basic repository systems will meet the technical and organisational requirements for trusted repository facilities from the point of view of grant providers in the Czech Republic. When obtaining the qualification of a trusted repository for these purposes, it will be sufficient to meet the conditions for the repository itself (data care, ...; these rules are still being prepared), the entire technological background and guarantee will be provided by the NRP.

In such a case, the user community is expected (and these activities should be supported by the OS II) to:

1. Establish the role of repository administrator. The repository administrator is a partner of the infrastructure operator (i.e. the NRP project in particular) for the repository

configuration agreement in all the points described below. The repository administrator also has primary responsibility for the data stored in the repository and for all the settings described below (which, of course, they delegate to others 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 cooperating with the cybersecurity team and for reporting cybersecurity incidents if they occur at the level of the repository they manage.

2. Establish of the role of a data curator, who forms general rules for data stored in the repository (e.g. in terms of retention length according to record type) and decides on specific datasets (e.g. resolves deletion requests). The curator also acts as a metadata specialist, i.e. They are a metadata specialist, who is responsible for the harmonisation of metadata within the repository in accordance with the established disciplinary metadata profiles and the interoperability of metadata with other systems (especially NMA). These roles may be separated from each other as appropriate.
3. Determine the disciplinary metadata profiles that will be available in the repository (in collaboration with CARDS IPs and methodologies for the respective systems) in the metadata profile management tool. Determine the metadata schema entities 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 basic repository systems, collaboration on the implementation of their support.
4. Determine the list of licenses available in the data deposition process.
5. Establish a data deposition workflow (e.g., record approval process).
6. Establish a data access workflow (from open access to a process involving approval by an ethics committee or access management committee).
7. Define the roles of user groups in the repository, e.g. regular depositor, curator, approver in individual parts of the workflow. Linking these roles to the groups of people in e-INFRA CZ AAI.
8. Create user documentation for the repository. For this purpose, ready-made components describing the basic functioning of the individual repository systems in the NRP and recommended documentation templates will be available from the NRP. However, the user documentation for the specific repository instances must describe the used metadata models, the repository deposition workflow, the search interface, the description of the user groups roles in the repository, etc.
9. Describe the repository policy, in particular when a record is considered closed; this policy should clearly indicate when a record stored in the repository is considered finalised, and what changes are allowed to finalised records (e.g. "adding a metadata item with a link to correct or use the record, but nothing else").
10. Provide user support (first level) for end users of the repository. The NRP infrastructure will provide (optional) tools to record user requests. The NRP funds cover the next levels of user support, which includes escalated requests that require intervention by infrastructure administrators, as well as support for the repository administrators themselves.
11. Submit data to the National Catalogue of Repositories - register the repository and its parameters in the National Catalogue of Repositories (NKR), continuous sending of updates in case of changes. The evidence also includes metadata profiles (schemas) and used controlled vocabularies and ontologies. Submission of repository information to the NKR should be automated via OAI-PMH or API.

[1] Each of the repository systems has specific constraints on the complexity and method of model definition. The repository administrator must choose a model and provide its description in a standardized form according to the specific requirements of the repository system (e.g. json in yaml for Invenio). The NRP will then provide a repository instance with this model and expects interaction with the repository administrator/technical contact) regarding testing and tuning of the model. More complex models, including those that cannot be prepared in this way without

further modifications, must have sufficient technical and staff capacity on the part of the repository administrator to implement any modifications to the repository system or its interface; these activities and associated costs are no longer borne by the NRP, which only has the capacity to use pre-prepared standardised procedures.

Building a repository running on NRP resources without using the basic repository systems

In necessary cases (where none of the three provided and supported repository systems can be used even after consultation), repositories built using alternative repository system implementations can be run directly in the NRP environment.

In such a case, the user community running such a repository will in principle get access to the data storage and application runtime environment (i.e. S3 and Kubernetes) from the NRP, but in such a case it must be responsible for all activities and bear all associated costs related to the installation and integration of the alternative repository system and the specific repository instance(s) (if it wants to run more than one) into the NRP environment and its operation.

The repository administrator must also take responsibility for:

1. All items described as the administrator's responsibility when using basic repository systems.
2. 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.
3. Deployment of disciplinary metadata profiles and their registration (in integration with the metadata profile registration system), harmonization of metadata within the repository and interoperability of metadata with other systems, especially NMA (can also be solved by the repository curator).
4. Selection and implementation of assignment of persistent identifiers from the set of standard supported identifiers, setting of assigned ranges.
5. Technical setup of metadata harvesting to NMA according to NMA requirements, including in accordance with the basic metadata model.
6. Technical setup of data deposition workflow and data access control.
7. Technical setup of links to e-INFRA CZ AAI systems and user group roles in the repository instance.
8. Creating user documentation.
9. Creation of documentation for system administration and operation.
10. User support for end-users at all levels, i.e. L1 to L3, except for requirements directly related to the operation and setup of the environment for running applications and storing data (S3 + Kubernetes).
11. Providing tools for integration into the national e-infrastructure environment, especially for data transfers between the repository and data repositories and computing resources in the e-INFRA CZ e-infrastructure.
12. Configure logging of repository systems to the NRP central logging system.
13. Inclusion in cybersecurity monitoring (CESNET-CERTS, FTAS). Cooperation in the implementation of security and especially penetration tests of the repository and directly related systems. Incident handling and cooperation with the cybersecurity team. Mandatory reporting of cyber security incidents.
14. Setting compliance with standard terms of service, defining additional conditions in coordination with NRP compliance.
15. Collection of statistical data on the operation and use of the system.
16. Operational monitoring.

17. Submission of data to the National Catalogue of Repositories - Registration of the repository and its parameters in the National Catalogue of Repositories (NKR), continuous sending of updates in case of changes. The evidence also includes metadata profiles (schemas) and used controlled vocabularies and ontologies. Submission of repository information to the NKR should be automated via OAI-PMH or API.

The repository administrator must also ensure that the system administrators and other staff have sufficient capacity to keep the system running stably.

Integration of the existing independently operated repository into the NRP/NDI environment

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

The administrator of a repository operated outside the NRP has full responsibility for its operation, from the hardware to the repository service itself. For such a repository to be considered "connected to the NRP/NDI", it must generally meet the same conditions as repositories in alternative implementations directly operated within NRP, except that the operation of the hardware resources, system administration and complete user support are also handled by the repository administrator. This applies not only to the operation of the repository system itself and its data repository, but also to other components that are necessary for its operation. The administrator of such a repository must provide functionality equivalent to that provided by the NRP; there is no restriction on how this will be achieved (how it will be implemented), but all functionality must be in a reasonable form (adequacy is primarily determined by the administrator, but may be requested by NRP administrators to be documented).

The minimum mandatory connection to the NDI environment consists of

- Connecting to the NMA and providing metadata in accordance with the core metadata model
- Submitting data to the National Catalogue of Repositories
- Connecting to the AAI operated by the NRP
- Defined API for data transfer
- Assigning PIDs (not necessarily DOIs)

All conditions relating to administrators, roles, licensing settings, and other policies and requirements related to the operation of the repository also apply to these repositories.

> The repository must also have adequate cybersecurity facilities and at least a basic level of monitoring to ensure the quality of operation and to collect data on the use of the system for statistical purposes. The repository must also have logging - essential for analysing cybersecurity incidents. The repository administrator is also responsible for ensuring a sufficient level of compliance with legal and other regulations, in relation to the nature of the data.

On the other hand, NRP services are available within the technical possibilities and capacities for the administrator of the existing repository, and we strongly recommend their widest possible use. The specific setup needs to be addressed specifically for each particular repository. Depending on the specific technical situation, a combination of a custom repository solution as defined in this section with the use of an NRP service can also be considered (e.g. a model where such a repository uses S3 in NRP as one of the data repositories is an example).

Closing remarks

The main objective of this document is to provide a graspable idea of the NRP service level for each use case. Given its structure, we have not found it useful to integrate a time perspective into it; the timing of the availability of each service can be traced in the project timeline.

Annex: Structured consolidation of requirements into tables

Conditions for building a repository using NRP basic repository systems

condition	Establishing the role of repository administrator
consultant	tech-support@eosc.cz
expected responsibilities of the repository administrator	The repository administrator is a partner of the infrastructure operator (i.e. the NRP project in particular) for the 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 for all the settings described below (which it delegates to others as necessary, of course). The repository administrator provides assistance to the infrastructure operator in deploying and testing the repository instance. The repository administrator is also informed about operational events of the repository and repository systems or the whole NRP (updates, outages, etc.). The repository administrator is also responsible for cooperating with the cybersecurity team and for reporting cybersecurity incidents if they occur at the level of the repository they manage.
description/comments	Please see the mailing list for the team that handles these matters.

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	Establishment of the role of a data curator, who formulates general rules for data stored in the repository (e.g. in terms of retention length by record type) and decides on specific datasets (e.g. resolves deletion requests). The curator also acts as a metadata specialist, i.e. they take care of the harmonisation of metadata within the repository in accordance with the established disciplinary metadata profiles and the interoperability of metadata with other systems (especially NMA). They participate in the establishment of the disciplinary metadata profile.
description/comments	The data curator can be the same person as the repository administrator, but we recommend that these roles are rather separated and delegated to different people (the size of the time commitment will depend on the size and complexity of the repository, the data contained in it and the level of services that the repository will provide). Alternatively, the role of curator and metadata specialist can be separated.

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 base metadata model)
expected responsibilities of the repository administrator	Determine the subject metadata profiles that will be available in the repository (in collaboration with CARDS IPs and methodologies for the respective systems) in the metadata profile management tool. Determine the metadata schema items that will be exported to the NMA (mapping to

	the base metadata model). In the case of metadata profiles that exceed the capabilities of the underlying repository systems, collaboration on the implementation of their support.
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 standard supported ones, setting of assigned ranges. Responsibility for the PIDs used, integration of PIDs and adherence to best practice in working with PIDs in accordance with National PID Centre (NTK) and PID provider guidelines.
description/comments	These are mainly memberships in international organisations and consortia that set rules on how to handle PIDs - e.g. DOI allocation: existing projects will provide methodological and technical assistance, but formally the responsibility lies with the operator of the repository, who must become a member of the DataCite consortium, not the repository system providers - CESNET/UK/KNAV.

condition	Licenses
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 workflow for data access (from open access to a process involving approval by an ethics or access control committee).
description/comments	

condition	Roles of user groups
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 groups of users, e.g. sharing specific data only with selected users.
description/comments	

condition	Repository user documentation
consultant	Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz Jan Kolouch jan.kolouch@cesnet.cz Pavlína Špringerová springerova@ics.muni.cz / ux@eosc.cz
expected responsibilities of the repository administrator	Creating user documentation for the repository. For this, ready-made components describing the basic functioning of the basic repository systems in the NRP and recommended documentation templates will be available from the NRP. However, the user documentation for specific repository instances must describe the used metadata models, the repository deposition workflow, the search interface, the description of the user group roles in the repository, etc.
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	A description of the repository policy, in particular when a record is considered closed; this policy should clearly indicate when a record stored in the repository is considered finalised, and what changes are allowed to finalised records (e.g. "adding a metadata item with a link to correct or use 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 provide (optional) tools for recording user requests. The NRP funds cover the next levels of user support, which includes escalated requests that require intervention by the 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), 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.

Conditions for building a repository running on NRP resources without using supported repository systems

A repository administrator with its own implementation of repository software must meet all the conditions of the previous section, plus the following conditions associated with the 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 thematic metadata profiles
consultant	Jakub Klímek jakub.klimek@matfyz.cuni.cz
expected responsibilities of the repository administrator	Deployment of disciplinary metadata profiles and their registration (in integration with the metadata profile registration system)
description/comments	Thematic 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 NMA requirements by OAI-PMH protocol (by selecting from supported metadata format serializations) 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	Selection and implementation of assignment of persistent identifiers from the set of standard supported identifiers, setting of assigned ranges. Responsibility for PIDs in use, integration of PIDs and adherence

	to best practice in working with PIDs in accordance with National PID Centre (NPC) and PID provider guidelines.
description/comments	In case of implementation of alternative repositories, the administrator is responsible for the technical implementation of PID allocation in the operating system. The operator of the repository is also responsible for the correct use of PIDs, the integration of PIDs and the adherence to best practice in working with PIDs in accordance with the guidelines of the National PID Centre (NTK) and PID providers. If it wants to allocate DOIs within the allocation for the Czech Republic, the organization responsible for the operation of the repository must become a member of the DataCite consortium.

condition	Licenses
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. (it might be a condition in NMK, but we are not there yet)

condition	Integration of e-INFRA CZ AAI
consultant	Peter Lényi lenyi@ics.muni.cz
expected responsibilities of the repository administrator	Connection of the repository instance to user authentication via e-INFRA CZ AAI, preferably via OIDC protocol. 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 groups of users, e.g. sharing specific data only with selected users. Determining and implementing the user lifecycle in the repository, including removing inactive users.
description/comments	Connection is via OIDC, or SAML can be used in justified cases.

condition	User documentation
consultant	Anastasia Avdeeva anastasia.avdeeva@ruk.cuni.cz Jan Kolouch jan.kolouch@cesnet.cz Pavlína Špringerová springerova@ics.muni.cz / ux@eosc.cz
expected responsibilities of the repository administrator	User documentation creation for end users of the repository, including basic description of the chosen implementation.
description/comments	

condition	Internal documentation
consultant	tech-support@eosc.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 end users support system for all levels, i.e. L1 to L3, except for requirements directly related to the operation; and setup 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 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 • For CLI tools in computing environments (MetaCentrum, IT4I) • Of portal systems, e.g. Galaxy • Staging to a computing environment running from the repository ("on the button") • Integration with Jupyter notebooks

condition	Operational logging
consultant	Radko Krkoš krkos@cesnet.cz Andrea Kropáčová andrea.kropacova@cesnet.cz
expected responsibilities of the repository administrator	Configuration of the logging of the repository systems 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). Logging completeness and consistency monitoring.
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	Inclusion in the CESNET-CERTS cybersecurity monitoring - notification of IP addresses where the repository is operated, contact details of responsible persons, reporting of planned atypical activities (large bursts of data, many connections, own penetration and other testing, etc.), Connecting the repository access element to the FTAS network flow monitoring infrastructure (configuring netflow data export) Responding to incidents and providing assistance to the CESNET-CERTS cybersecurity team in resolving incidents, Mandatory reporting of cybersecurity incidents to CESNET-CERTS. Systematic monitoring and response to technical vulnerabilities (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 be 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 Pavlína Špringerová springerova@ics.muni.cz / ux@eosc.cz
expected responsibilities of the repository administrator	Setting up a system to collect statistics/system operation and usage data. Collect and provide this data for 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 Pavlína Špringerová springerova@ics.muni.cz / ux@eosc.cz
expected responsibilities of the repository administrator	Setup of repository services monitoring.

This is not an authorised translation. The document "Splnění povinnosti napojení na NRP a NMA a jejich integrace do společně budované NDI" (version 3) was translated by DeepL and edited by the EOSC CZ secretariat team.

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 linking a separate 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

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

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...

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			

This is not an authorised translation. The document "Splnění povinnosti napojení na NRP a NMA a jejich integrace do společně budované NDI" (version 3) was translated by DeepL and edited by the EOSC CZ secretariat team.