

BOZZA

PNRR – M 4 - C2 - INVESTIMENTO 1.4

Decreto Direttoriale n. 3138 del 16 dicembre 2021 e s.m.i.

“Potenziamento di strutture di ricerca e creazione di *“campioni nazionali”* di R&S su alcune Key Enabling Technologies”

Numero del progetto: CN_00000013

Titolo del progetto: National Centre for HPC, Big Data and Quantum Computing

Area tematica: Simulazioni, calcolo e analisi dei dati ad alte prestazioni

Soggetto Proponente: Istituto Nazionale di Fisica Nucleare (INFN)

DATA MANAGEMENT PLAN

(DMP)

1. RIEPILOGO DEI DATI

- Qual è lo scopo della raccolta/generazione dei dati e la sua relazione con gli obiettivi del progetto?

The ICSC project Data Management Plan (DMP) aims to create a set of enabling technologies and solutions for scientific domains where the availability of large and reliable processing systems is a necessity not met by currently available systems. Data generated during the project can be divided into the following groups:

- *Data collected and/or generated by the Scientific spokes;*
- *Data generated by the project as a whole during continuous monitoring, analysis and documentation processes.*

The purpose of this data collection is to help in achieving the Objectives of the project.

The problem of data collection / generation is not trivial, since the Spokes deal with data of markedly heterogeneous nature and with different levels of sensitivity: from data which is intended to be made public after short embargoes (as typical in fundamental research and cosmo observation), to data containing personal information which needs to be treated in compliance with European and national legislations, to data which needs to be treated in compliance with IP regulations.

The CN has been designed to allow for data-intensive use cases, with data repositories distributed on the Italian territory either domain-specific or generalistic, well connected among themselves and with the major processing facilities.

Scientific domains whose research is boosted by the availability of large datasets are, just for example:

- *hard science domains, which analyze or simulate data from more and more complex instruments, with output rate which cannot be analyzed with standard procedures;*
- *medicine and drug design, which has been proven to profit from the availability of multi-instrument data;*
- *earth observation and climate research, which rely on the availability of sensor data;*
- *digital society and smart cities, which profit from multi-modal data collected from heterogeneous sensors.*

On top of that, the availability of data from multiple domains via an integrated HPC-Cloud-Big Data infrastructure, a key overall objective of the CN, can foster cross-fertilization and knowledge sharing.

Data handling, and in particular its aspects of sensitivity, rights and patterns of access and lifetime, is an essential part of the project, and needs to be covered by a specific Data Management Plan (DMP). As

explained in the project document, a DMP will be made available by the project at the end of the first year of activities (Milestone 6 – M9-M12), once the specific interventions and research directions are specified in the respective domains. It will be constantly updated throughout the project, following updates in the use cases, in the legislation, and in the understanding of data access and retention patterns. The DMP will include information on (at least):

- *the handling of research data during & after the end of the project;*
- *what data will be collected, processed and/or generated;*
- *which methodology & standards will be applied;*
- *whether data will be shared/made open access;*
- *how data will be curated & preserved (including after the end of the project).*

In general terms, the CN and its spokes will follow the “Guidelines on Data Management in H2020”, "Horizon Europe Data Management Plan Template", in addition to the EU General Data Protection Regulation (GDPR): these documents offer a good operational context and are already known and adopted by most of the CN partners. Recognizing the importance of ethical and privacy concerns, the adopted approach will address ethical and privacy requirements, including explainable and trustworthy AI. Additionally, the CN will follow the implementation guidelines of the National Plan for Open Science from PNR 2021-2027, which is currently still in draft status.

- Quali tipi e formati di dati verranno generati/raccolti dal progetto?

The type and source of data to be handled (either collected or generated) depends not only on the spoke and its domain, but also on the specific use cases considered.

In general, the spokes can treat among others:

- *scientific data:*
 - *raw data from scientific instruments;*
 - *data from industrial processes;*
 - *data from existing repositories, for example including historical trends or pre-digested data;*
 - *data from simulation activities, like in the realization of digital twins or in the simulation of hard science processes;*
 - *multi-modal data from heterogeneous sources.*
- *other products of research:*
 - *public data published on the project website;*
 - *private and public deliverables, as specified by the Description of Work;*
 - *reports on the Data Management Plan;*
 - *reports on dissemination workshops;*
 - *annual reports;*
 - *financial reports.*

Data formats are again spoke-dependent. In some domains (cosmo observation, materials & molecular science, medicine) standardized formats exist and are of common uses. In other domains, specific data formats are defined and used whenever needed.

In general, using specific data formats is not an obstacle to data reuse, if such formats are well documented and consistent. The DMP will hence need to address, per spoke and possible per use case, a clear definition of the formats used; such definition will need to be kept up to date in case of changes throughout the project lifetime.

- Riutilizzerete i dati esistenti e in che modo?

In some domains pre-existing data are of large value, and cannot be easily regenerated when a new research starts.

This includes, for example, data from large science experiments (colliders, gravitational wave detectors, satellites), repositories of medical or climate data, large databases on materials. These data will be made available to the CN users (possibly after proper embargo periods), either via direct access to source repositories or via the injection on the Centre storage, according to the policies defined by the data owners. Data on the CN will not be available to all the Centre affiliates, but it will retain its access policies and restrictions as agreed with the data owners and in accordance with the respective IP rights.

- Qual è l'origine dei dati?

Data accessed by CN affiliates vary in origin not only spoke per spoke, but typically use case per use case. Data from large scientific instruments comes from existing repositories handled by the experimental communities. Climate data can be injected or accessed from repositories at local and global level. Medical data can come from public or private repositories, scaling from small hospitals to large (super)national entities.

Finally, data originated by monitoring and assessment activities will be generated according to the Continuous Risk Monitoring and Assessment procedures detailed in the Description of Action.

- Qual è la dimensione prevista dei dati?

The affiliates to the spokes are stakeholders of world-level data intensive research initiatives, where the size of interesting data is already in the multi Petabyte to the Exabyte level (High Energy Physics, Space and Earth observation, Genomics, Climate, Material data repositories, ...). As such, the limit to what amount of data would be interesting resides more in what the CN will be able to store and process.

In the CN proposal, and in particular in the Spoke 0, a budget has been dedicated to the procurement and the operations of distributed storage systems; in particular, INFN, with its distributed centres, and CINECA, at Tecnopolo and in Naples, have provisioned for roughly 200 PB available to the spoke affiliates. On top of this, affiliates will have access to the existing domain research infrastructures, national and supranational, for a total largely exceeding the Exascale. While it is obviously unfeasible to ingest directly the total amount

in the CN, the middleware capabilities can still allow (when required) to deploy caching mechanisms in order to give final users an equivalent experience.

- A chi potrebbero essere utili?

The data generated in the project will be very beneficial to a variety of stakeholders including: policy makers, public funders, researchers, public administration, industry and citizens.

- *Scientific data from large instruments are essential for the work of researchers, but are also valuable for education (via for example Masterclasses) and for citizen science.*
- *Digital society data are useful to research institutions, local and state level administrations.*
- *Medical data are useful in research and in the*
- *Climate data are useful for research and for the planning of local and global response to changes.*

Information gathered will help identify challenges in these areas and serve as an input for policy design at national and European level in these areas. Finally, the data will be used for training and dissemination activities, towards the research environment and the wider public .

2. DATI FAIR

2.1. Rendere i dati reperibili, comprese le disposizioni relative ai metadati

- I dati prodotti e/o utilizzati nel progetto sono individuabili con metadati, identificabili e localizzabili tramite un meccanismo di identificazione standard (ad esempio, identificatori persistenti e unici come i *Digital Object Identifier*)?

The spokes are expected to operate within existing collaborations and follow common practices in these domains, including the utilization of services and standards for DOI and metadata. Any data generated by the CN will adhere to these standards.

On the other hand, the CN wants to foster cross-fertilization of domain data, by allowing interdisciplinary research. As such the realization of a mechanism for connecting heterogeneous data is envisaged and will be detailed in the DMP delivered at the end of the first year. Technically, the CN will use to the greatest possible extent a common middleware, which automatically ensures the interoperability of common low level services like data access and movement. For metadata, an existing task force created by ICDI (Italian Computing and Data Infrastructure, including Universities and Research Institutions) is trying to design an interoperable system between existing italian research infrastructures (allowing, for example, for metadata searches on the whole infrasctructure); this could be the basis of a CN common data infrastructure.

- Quali convenzioni di denominazione seguite?

It varies spoke by spoke, as explained in the previous answer, following domain level best practices. In general, all the specificities will be explained in the DMP document initially delivered at the end of the first year, and then continuously updated.

- Verranno fornite parole chiave di ricerca che ottimizzino le possibilità di riutilizzo?

It varies spoke by spoke, as explained in the previous answer, following domain level best practices. In general, all the specificities will be explained in the DMP document initially delivered at the end of the first year, and then continuously updated.

As detailed previously in this document, at metadata level there is the plan to make available CN-wide a metadata search system, to be accessed on the CN central portal.

- Fornite numeri di versione chiari?

It varies spoke by spoke, as explained in the previous answer, following domain level best practices. In general, all the specificities will be explained in the DMP document initially delivered at the end of the first year, and then continuously updated.

- Quali metadati verranno creati? Nel caso in cui non esistano standard di metadati nella vostra area, indicate che tipo di metadati verranno creati e come.

It varies spoke by spoke, as explained in the previous answer, following domain level best practices. In general, all the specificities will be explained in the DMP document initially delivered at the end of the first year, and then continuously updated.

2.2. Rendere i dati apertamente accessibili

- Quali dati prodotti e/o utilizzati nel progetto saranno resi apertamente disponibili di *default*? Se alcuni set di dati non possono essere condivisi (o devono essere condivisi con restrizioni), spiegarne il motivo, separando chiaramente le ragioni legali e contrattuali dalle restrizioni volontarie.

Each domain and use case has specific needs and policies for what concerns open access to data. In general, the CN will follow the FAIR data principles – Findable, Accessible, Interoperable and Reusable. It has to be noted, though, that FAIR does not necessarily imply that data will be openly available in all cases, as impossible for (e.g.) the sensitive data in Spoke 8. The paradigm to be followed is “As Open as Possible, as Closed as Necessary”, which implies a clear definition of the access patterns for each data object. The implementation of this paradigm requires an underlying technical infrastructure able to guarantee data safety and correct access granularities. These will be guaranteed by specific infrastructures as deployed and managed by the CN (e.g., ISO 27001 certified repositories and centres), and a state-of-the-art Authentication & Authorization Infrastructure (AAI), based for example on the AARC Blueprint architecture.

The specific status of each dataset will be initially defined inside the spoke, among the affiliated participating to the activity, and then in the Ethics and Data Management operational unit in the CN. This will be:

- *Coordinating the data management routine procedures of the CN, in which also specific data management figures operate within each spoke: it is responsible in the CN for the implementation of policies, supported by a team of stewards (at least 10) representing data domains, to ensure the quality and integrity of the data lifecycle*
- *Implementing executive decisions and ensuring adherence to procedures*
- *Enforcing the data governance policies of the Ethics and Data Governance Board*
- *Ensuring the routine ex ante and ex post ethical review of the projects of the CN;*

in doing so, the Unit avails itself of the advice of the Ethics and Data Governance Board, defined in the full project.

- Come saranno resi accessibili i dati (ad esempio, tramite deposito in un archivio)?

The data will reside on CN related computing centres; they will be interconnected connected via “datalake” middleware layer explained in the full project. Access and search protocols will be clearly defined via the

datalake, and the infrastructure will allow for an efficient data ingestion from all the connected processing facilities.

In some domains, data can be also made available via existing repositories, as (for example) for the Materials & Molecular Science spoke, which will establish formal partnerships with at least one of the two official repositories for materials data recommended by the EU Commission in Open Research Europe, Materials Cloud (MC) and NOMAD, with whom preliminary discussions are already taking place.

Data handles and / or generated in the context of large scientific collaborations (in High Energy Physics, for example) will be additionally made available to the collaboration members, and possibly egressed to existing repositories outside the centre.

- Quali metodi o strumenti software sono necessari per accedere ai dati?

The datalake will be the technical means for data access in the CN. It will provide multiple access and search protocols, in the form of a thin layer on top of existing infrastructures.

- È inclusa la documentazione sul software necessario per accedere ai dati?

The technical documentation on the datalake will be made available by the spoke 0; the specific spoke by spoke additional documentation, for example detailing the data formats used, will be provided along with the data management plan and published on the project website.

- È possibile includere il software pertinente (ad esempio, in codice open source)?

It is domain dependent. Some scientific domains (physics, for example) have moved to a paradigm of fully open software, made available via public repositories (github, for example) of specifically tailored repositories; other domains (related to some health-related activities, for instance) may need to resort to closed-source or proprietary software.

The CN fosters the development of open software, unless specific sensitivity or IP issues arise.

Technically, open source catalogs like those from Indigo-DataCloud and EOSC-hub are easily deployable on the datalake middleware, and will be made available whenever deemed necessary.

- Dove saranno depositati i dati e i metadati, la documentazione e il codice associati? Si dovrebbe dare la preferenza a depositi certificati che supportano l'accesso aperto, ove possibile.

The datalake supports data and metadata repositories in the form of services deployed on its infrastructure (i.e., its computing centres), including the support for ISO 27001 certified repositories for sensitive data.

- Avete esaminato gli accordi appropriati con il deposito individuato?

Not relevant (see previous answer)

- Se ci sono restrizioni all'uso, come sarà garantito l'accesso?

At the core of the datalake infrastructure, various services will be made available, including a federated Authentication and Authorization Infrastructure (AAI), compatible with the EOSC AAI. This will allow to control access to data according to the paradigm “As open as possible, as closed as necessary”. While the specific implementation design for AAI is not yet available, it will follow the AARC Blueprint architecture¹ and as such will offer federated access management solutions for international (research) collaborations, with the possibility to reuse existing identity providers.

- È necessario un comitato per l'accesso ai dati?

As detailed in the full proposal, the CN has defined an Ethics and Data Management operational unit and an Ethics and Data Management operational Governance Board.

The operational unit will be:

- *Coordinating the data management routine procedures of the CN, in which also specific data management figures operate within each spoke: it is responsible in the CN for the implementation of policies, supported by a team of stewards (at least 10) representing data domains, to ensure the quality and integrity of the data lifecycle*
- *Implementing executive decisions and ensuring adherence to procedures*
- *Enforcing the data governance policies of the Ethics and Data Governance Board*
- *Ensuring the routine ex ante and ex post ethical review of the projects of the CN;*

in doing so, the Unit avails itself of the advice of the Ethics and Data Governance Board, as defined in the full project.

- Esistono condizioni ben descritte per l'accesso (ad esempio, una licenza a lettura ottica)?

Not yet. This will be defined in the full DMP document delivered at the end of the first year.

- Come verrà accertata l'identità della persona che accede ai dati?

See above the discussion on the federated AAI. In most of the cases, the authentication will be delegated to existing identity providers from affiliated and trusted institutions.

2.3. Rendere i dati interoperabili

- I dati prodotti nel progetto sono interoperabili, ossia consentono lo scambio e il riutilizzo dei dati tra ricercatori, istituzioni, organizzazioni, Paesi, ecc. (ossia aderiscono agli standard per i formati, sono il più possibile conformi alle applicazioni software (aperte) disponibili e, in particolare, facilitano la ricombinazione con set di dati di origine diversa)?

¹ <https://aarc-project.eu/architecture/>

As explained above, the data formats will be varying use case by use case, even within the same spoke. They generally follow existing practices of the domains, including more or less open data formats and standards. This is not considered a problem, if such formats are stable and well documented. The interoperability guidelines and frameworks developed at international level, such as the ones made available within the European Open Science Cloud (EOSC), will be considered.

- Quali vocabolari di dati e metadati, standard o metodologie seguirete per rendere i vostri dati interoperabili?

Highly spoke dependent; the most appropriate vocabularies and formats will be chosen within each spoke according to the specific needs and experiences

- Utilizzerete vocabolari standard per tutti i tipi di dati presenti nel vostro insieme di dati, per consentire l'interoperabilità interdisciplinare?

As explained in a previous answer, each domain will define specific services. But, there is an attempt to prepare a cross-domain infrastructure for interdisciplinary searches on multiple data sources.

- Nel caso in cui sia inevitabile utilizzare ontologie o vocabolari non comuni o generare progetti specifici, fornirete mappature a ontologie più comunemente utilizzate?

Not decided currently

2.4. Aumentare il riutilizzo dei dati

- Come verranno concesse le licenze per consentire il più ampio riutilizzo possibile dei dati?

Data remain properties of the institutes generating / owning them. The same institutes will define access policies and restrictions.

- Quando i dati saranno resi disponibili per il riutilizzo? Se si chiede un embargo per dare tempo alla pubblicazione o alla ricerca di brevetti, specificare perché e per quanto tempo sarà applicato, tenendo presente che i dati della ricerca dovrebbero essere resi disponibili il prima possibile.

This is highly dependent on the spoke / use case. In general, hard sciences like astroparticle and collider physics open the access to data fully after an embargo period (years or less). Other domains involving sensitive or IP protected data simply cannot make the data available (see the "As Open as Possible, as Closed as Necessary" paradigm) unless via anonymization / obfuscation procedures. In all cases, the data owner will propose a policy for its data, which will be analyzed in the Ethics and Data Management Unit and Governance Board.

- I dati prodotti e/o utilizzati nel progetto sono utilizzabili da terzi, in particolare dopo la fine del progetto? Se il riutilizzo di alcuni dati è limitato, spiegarne il motivo.

See previous answer

- Per quanto tempo si prevede che i dati rimangano riutilizzabili?

Even if the CN project is 3 years, it is expected and modeled to be sustainable after that date, with the capability to continue serving the data (if not declared obsolete and removed)

- Sono descritti i processi di garanzia della qualità dei dati?

Not yet, this will be part of the DMP at the end of the first project year.

Oltre ai principi FAIR, i DMP devono anche affrontare i seguenti aspetti:

3. Allocazione delle risorse

- Quali sono i costi per rendere i dati FAIR nel vostro progetto?

The activities related to making the data/outputs FAIR are anticipated to be covered within the allocated budget for each spoke. The estimation of such cost is not possible at the moment because it depends on various elements, among which: 1) the possibility to implement a "FAIR by design approach", 2) the tools/frameworks chosen to make data FAIR if they are not so yet 3) the tools to assess the FAIR level, etc.

- Come saranno coperti?

Via the spokes' allocated budget.

- Chi sarà responsabile della gestione dei dati nel vostro progetto?

The CN has in its governance many bodies and professionals dedicated to data management.

In accordance with article 3713 of the GDPR, a Data Protection Officer will be appointed by the project; he/she will work in strict contact with the DPO of the single institutions' members of the CN, with data controllers and processors designated per each data source, and will be responsible for:

- *coordinating with the DPO of the institution's members of the CN;*
- *informing and providing advice to the data controller or the data processor as well as to the employees who carry out the processing on the obligations deriving from the GDPR, as well as from other national or European Union laws related with data protection;*
- *monitoring the compliance of the data controller or the data processor to the GDPR requirements, including the assignment of responsibilities, awareness-raising, and training of personnel involved in the processing and related control activities;*
- *providing an opinion on the impact assessment on data protection and monitor its performance according to art. 35 of the GDPR (if specifically requested);*
- *cooperating with the privacy authority (Garante per la Protezione dei Dati Personali) for any issue related with data processing and data protection, including early consultations (art. 36);*
- *supporting the data controller in keeping the processing register, follow the data controller's instructions.*

A single DPO will be designated for the whole CN; he/she will be assisted by experts from each spoke to cover the specificities in each domain.

As detailed in the full proposal, the CN has defined an Ethics and Data Management operational unit and an Ethics and Data Management operational Governance Board.

The operational unit will be:

- *Coordinating the data management routine procedures of the CN, in which also specific data management figures operate within each spoke: it is responsible in the CN for the implementation of policies, supported by a team of stewards (at least 10) representing data domains, to ensure the quality and integrity of the data lifecycle*
- *Implementing executive decisions and ensuring adherence to procedures*
- *Enforcing the data governance policies of the Ethics and Data Governance Board*
- *Ensuring the routine ex ante and ex post ethical review of the projects of the CN;*

in doing so, the Unit avails itself of the advice of the Ethics and Data Governance Board, as defined in the full project.

- *Sono state discusse le risorse per la conservazione a lungo termine (costi e valore potenziale, chi decide e come saranno conservati i dati e per quanto tempo)?*

Not yet. They will be included in the DMP at the end of the first year.

4. Sicurezza dei dati

- *Quali sono le disposizioni in vigore per la sicurezza dei dati (compreso il recupero dei dati e l'archiviazione e il trasferimento sicuro dei dati sensibili)?*

The CN, via its infrastructure, provides the capability to handle data which cannot be considered simply open, due to their sensitivity or IP value.

Specifically, the centre will provide:

- *ISO 2700x infrastructures for the secure handling of sensitive data, including wiping techniques for temporary storage (warm storage) of raw data during the data collection; The final (cold) storage will be endowed with encryption at rest procedure (AES256). The system will also be equipped with a disaster recovery storage located more than 50 km from the cold storage.*
- *Industry standard AAI, for the handling of permissions at the level of groups and single individuals*

Spoke #8 (In-silico Medicine & Omics data) hosts the most sensitivity-critical activities; it has detailed a list of measures to ensure compliance with the GDPR which is considered valid for all use cases with sensitivity issues:

- *adopt a procedure for the pseudonymization of personal data;*
- *adopt a strategy for the management of data breaches;*
- *perform a Privacy Impact Assessment (PIA) and report on the adopted methodological approach;*
- *update and validate all relevant privacy-related documents (e.g., reports, informed consents, etc.*
- *adopt and update a processing register;*

- *identify a person responsible for the protection of personal data or data protection officer, with the assignment of specific tasks for the effective implementation and effective application of compliance according to the GDPR (for example, providing day-by-day advice on the issues concerning data protection, performing risk assessments);*
- *define Standard Operating Procedures (SOP);*
- *carry out information and training activities on the requirements and obligations resulting from the GDPR;*
- *schedule periodic audit activities to monitor correct SOP application;*

Since the approach is risk-based and requires an impact assessment, it is important to adopt a solid risk management framework and perform a Data Protection Impact Assessment (DPIA).

- *identify the architecture, the parties involved, the roles and responsibilities for risk management;*
- *assess the confidentiality of personal data processed;*
- *implement the risk management through risk identification, assessment, prioritization, and mitigation strategies;*
- *review the risks (e.g. data loss, data integrity, unauthorized access, unauthorized or non-compliance of the processing, continuity of service, etc.) and related risk avoidance and risk mitigation strategies;*
- *assess of the residual risk and review of the risk avoidance and mitigation strategies;*
- *monitor and review of the risk management model for a continuous improvement of the data protection and risk management system (monitoring for continuous improvement).*

Following the European and national rules and regulations regarding personal data protection, it is always required to define minimum data retention time. Any processing of personal data carried out by employees, collaborators, and all those who, although not employed by but operate in various roles in the activities involving data processing, must respect privacy by design and by default principles. In full compliance with regulations in force and following the principles of necessity, correctness, pertinence, and minimisation, data processing takes place solely for determined, explicit, and legitimate purposes.

The information and documents will be managed and preserved on registered and approved systems for their technical and contractual elements, respecting the current applicable rules and regulations and the guidelines referring to the protection of personal data. The security measures adopted must be in line with the level of risk that has been identified.

- *I dati sono conservati in modo sicuro in archivi certificati per la conservazione e la cura a lungo termine?*

Certified repositories (for example, ISO 2700x) are used only when sensitive data is involved.

5. Aspetti etici

- *Esistono questioni etiche o legali che possono avere un impatto sulla condivisione dei dati? Questi aspetti possono essere discussi anche nel contesto della revisione etica.*

This is spoke dependent, but certainly true in some cases (spoke #8, for example). The CN will establish an Ethics and Data Governance Board.

The Board is composed of 7 outstanding experts in the field and is appointed by the Board of Directors. It is chaired by a Chairperson selected among its members. The Ethics and Data Governance Board provides independent advice, guidance, and feedback to the CN on ethical and data governance issues. From a policy and strategic perspective, the CN aims to maximize the positive societal impact of its activities. In particular, it adheres to Open Science standards (e.g., following the guideline "as open as possible, as closed as necessary") and to the principles of the upcoming EU Data Act in maximizing "the value of data in the economy by ensuring that a wider range of stakeholders gains control over their data and that more data is available for innovative use, while preserving incentives to invest in data generation".

- Il consenso informato per la condivisione dei dati e la conservazione a lungo termine è incluso nei questionari che trattano i dati personali?

Not yet defined, but it is a standard practice and a request induced by the GDPR

6. Altre questioni

- Vi avvalete di altre procedure nazionali/fondatori/settoriali/dipartimentali per la gestione dei dati?
Se sì, quali?

The CN will base its middleware on existing standards and interoperability frameworks and guidelines already well known and in operations at the affiliates (like solutions from RDA, Eudat, Zenodo, OpenAIRE, WLCG, EOSC projects, ...), minimizing the need for the development of new tools. The middleware is expected to be in the form of a thin connecting layer on top of existing infrastructures, reusing them without jeopardizing or interrupting existing functionalities. Using existing standards the CN data repositories will also be by design interoperable with extra national and domain specific infrastructures.

7. Ulteriore supporto per lo sviluppo del vostro DMP

La Research Data Alliance mette a disposizione un Repertorio degli standard di metadati che può essere consultato per individuare gli standard specifici della disciplina e gli strumenti associati.

Lo strumento EUDAT B2SHARE include una procedura guidata per le licenze che facilita la selezione di una licenza adeguata per i dati di ricerca.

Tra gli elenchi utili di depositi vi sono:

Registro degli archivi di dati di ricerca

Alcuni archivi, come Zenodo, una collaborazione di OpenAIRE e del CERN, consentono ai ricercatori di depositare sia le pubblicazioni che i dati, fornendo strumenti per collegarli.

