# **Template for Collecting ERF Community Requirements on Data**

*Version 1.1*

# **ERF Member Community Information**

|  |  |
| --- | --- |
| **ERF Member Name \*** | *ERF member community full name and short name if any*  |
| **ERF Member Description \*** | **Community Description**: *ERF member community description* **Website**: *website* **Community Objectives**: *Community objectives* **Scientific Challenges**: *Scientific challenges (problems and motivations for the collaborations)* *Expectation (for the new technology to be provided)**Impacts and benefits*  |
| **Information Provider \*** | *Contact of responsible person for the information provision* *Institution of the contact person*  |

# **Current System**

*Please provide information for current system.*

|  |  |
| --- | --- |
| **Data Policy** |  |
| Aspects covered by your community data policy |  |
| Reference or links | *Link to community data policy*  |
| Known issues | *Problems or consideration for new aspects to be added in Community Data Policy*  |
| **General Data Protection Regulation (GDPR)** |
| Practices in GDPR | *Does your community apply EU GDPR (*[*http://www.eugdpr.org/*](http://www.eugdpr.org/) *)? What are your experiences?*  |
| Reference or links | *Link to your community GDPR*  |
| Known issues | *Details of any problems and issues in practices*  |
| **Data Management Plan** |  |
| Aspects covered by your community’s Data Management Plan (DMP) | *What aspects are defined in your community’ DMP?* *Mention, in particular, if there are DMPs for individual researchers. What are the rules or regulations for them to create own DMPs?* |
| Formal DMP | *Do you have a formal DMP?**At what level is it defined /approved?* |
| Reference or links | *Link to your community’s DMP*  |
| Known issues | *Problems or consideration preventing new aspects to be added to your community’s DMP*  |
| **Data Lifecycle** |  |
| Data lifecycle stages | *Description of the different stages in handling scientific data in your infrastructure, e.g. data acquisition, curation, access, processing, deletion etc. Please use diagrams if possible.*  |
| Data management activities at each stage | *Description of any additional data management activities e.g. quality control, data transfer, data integration, provenance etc.*  |
| Tools/services used for handling data lifecycle |  |
| Limitations of current tools/services for handling data lifecycle | *Description of any performance limitations you face at present (e.g., difficulty for handling large-scale data). Include any requirements for new functionalities or improved performance.*  |
| **Open Data** |  |
| Community policy on open data | *Link if a policy regarding open data is available. This can include conditions, exceptions (e.g. embargoes)* |
| PIDs | *Detail if/how you use Persistent Identifiers to data (e.g. DOIs) including services (e.g. DataCite) or tools for this.* |
| Limitations of current tools/services for handling open data | *Any desirable tools/tools/services that are missing fromyour current system, or any performance limitation (e.g., difficulty for handling large-scale data). Requirements for new functionalities or improved performance.*  |
| **Data Objects** |  |
| Types of data objects | *Types of data objects handled by the infrastructure*  |
| Data size  | *The sizes of individual data objects and the whole collection* |
| Data velocity | *What is the increasing speed for data generation*  |
| Data format | *The format of datasets* |
| Data Identifiers | *Any identifier information* |
| Standards in use | *International or community standard in use*  |
| Data distribution | *Logical and physical location of data objects, indicating future changes if any*  |
| Data packaging | *How data is organised in unit and packaged to be handled by containers*  |
| **Metadata Objects** |  |
| Types of Metadata objects | *Types of metadata objects handled by the infrastructure*  |
| Metadata Identifiers | *Any identifier information, describing current and future status*  |
| Metadata size | *The sizes of Individual metadata objects and the whole collection, describing current and future status*  |
| Metadata Schema | *Name of schemas you use* |
| Link to Metadata Schema | *Link to any schemas you use* |
| Metadata format | *The format of metadata, describing current and future status*  |
| Standards in use | *The format of datasets, describing current and future status*  |
| Metadata Distribution | *Logical and physical location of metadata objects, indicating future changes if any*  |
| Limitations of current infrastructure in handling metadata | *New tools/tools/services are missing from current system, or any performance limitation (e.g., difficulty for handling large-scale data). Requirements for new functionalities or improved performance.*  |
| **Data Storage** |
| Storage types | *Is your data stored on disks or tape?. Single or multiple copies?*  |
| Storage period | *How long you keep the data*  |
| Data Migration | *Do you migrate data (e.g. from one format to another)* |
| Data storage tools/software | *What tools/software do you use in current system for data storage (e.g. hierarchical storage management systems)* |
| Limitation of current tools/software for data storage | *Are there any tools/tools/services missing from current system that would be beneficial, or any performance limitations? Requirements for new functionalities or improved performance.*  |
| **System Architecture**  |
| System Architecture Description | *Please describe system architecture, use diagrams if possible*  |
| Known issues |  |
| **Authentication Authorisation and Accounting** |
| Authentication service | *How does current system handle user authentication? What protocols and types of services are in use?*  |
| Authorisation service | *How does current system handle user authorisation? What protocols and types of services are in use?*  |
| Accounting service | *How does current system track user activities? What protocols and types of services are in use?*  |
| Limitations of current system in handling AAA requirements | *New tools/tools/services are missing from current system, or any performance limitation. Requirements for new functionalities or improved performance.*  |
| **Computing Capacity** |
| Infrastructure capacity | *What are capacity for computing, storage and network?*  |
| Limitations of current capacity | *Problems and issues*  |
| **Human Resources** |
| IT staffs | *How many IT staffs*  |
| Data staffs |  |
| Cost/budgets |  |

# **Requirements for future systems**

*If there are limitations for current system in handling community data, what are your requirements for new services/tools. Please consider concrete usage scenarios and describe them as case studies.*

|  |  |
| --- | --- |
| **Case Study** | **Case Study** *n (the number of case study):* *A Case Study is an implementation of a research method involving an up-close, in-depth, and detailed examination of a subject of study (the case), as well as its related contextual conditions.* **User Story *n (the number of User Story):*** *The Case Study is based on a set of User Stories, which which are tokens representing requirements and describing the value of functionalities from a user’s perspective, i.e. how the researcher describes the steps necessary to solve each part of the problem addressed. In practice, the user community should be notified that the selection of the use stories is representative reflecting both of the research challenge and complexity, and of the possible solutions offered by the investigation project.* *Use Case n (the number of User Case):* *User Stories are the starting point of Use Cases, where they are transformed into a description using software engineering terms (like the actors, scenario, preconditions, etc.) Use Cases are useful in capturing the requirements that will be handled by the technology provider and can be tracked, e.g., by a Backlog system from an OpenProject tool. Specify, in particular, the following information:* **Community Roles**: List of stakeholders who will use the system and their responsibilities**Workflow***A sequence of actions to fulfil each case studies -- explicitly indicate who(actor) wants to do what need what services/functions and handle what information objects (data, metadata, signals etc.). Using figure if possible.*  |
| **Functional Requirements** | *﻿Describe the functionalities needed*  |
| **Non Functional Requirements** | *Describe performance requirements, indicating requirement levels (High/Middle/Low)**Availability* *Accessibility* *Throughput or Response time* *Security* *Utility* *Reliability* *Scalability* *Efficiency* *Others* |
| **Requirements for using e-Infrastructure** | *Please indicate if the current solution is already using an e-Infrastructure (like GEANT, EGI, PRACE, EUDAT, a Cloud provider, etc.) and if so what middleware is used.* *Any issues of using e-Infrastructure resources and services***Computational requirements: ﻿****Storage requirements:** **Network requirements:** **Software requirements:**  |