

Appendix A Common Requirements of Environmental Research Infrastructures

The following tables describe the common requirements environmental research infrastructures. The requirements are divided in five sets that correspond to the five stages of the datalifecycle. The requirements highlighted on each table are the minimal model.

Data Acquisition (A)

| | Functions | Definitions |
|----------|---|---|
| A .1 | Instrument Integration | Functionality that creates, edits and deletes a sensor. |
| A .2 | Instrument Configuration | Functionality that sets-up a sensor or a sensor network. |
| A .3 | Instrument Calibration | Functionality that controls and records the process of aligning or testing a sensor against dependable standards or specified verification processes. |
| A .4 | Instrument Access | Functionality that reads and/or updates the state of a sensor. |
| A .5 | Configuration Logging | Functionality that collects configuration information or (run-time) messages from a sensor (or a sensor network) and outputs into log files or specified media which can be used by routine troubleshooting and in incident handling. |
| A .6 | Instrument Monitoring | Functionality that checks the state of a sensor or a sensor network which can be done periodically or when triggered by events. |
| A .7 | Parameter/Data Visualisation | Functionality that outputs the values of parameters and measured variables a display device. |
| A .8 | <i>Real-Time Parameter/Data Visualisation</i> | <i>Specialisation of (Parameter) Visualisation which is subject to a real-time constraint.</i> |
| A .9 | Process Control | Functionality that receives input status, applies a set of logic statements or control algorithms, and generates a set of analogue / digital outputs to change the logic states of devices. |
| A .10 | Data Collection | Functionality that obtains digital values from a sensor instrument, associating consistent timestamps and necessary metadata. |
| A .11 | <i>Real-Time Data Collection</i> | <i>Specialisation of Data Collection which is subject to a real-time constraint.</i> |
| A .12 | Data Sampling | Functionality that selects a subset of individuals from within a statistical population to estimate characteristics of the whole population. |
| A .13 | Noise Reduction | Functionality that removes noise from scientific data. |
| A .14 | Data Transmission | Functionality that transfers data over communication channel using specified network protocols. |
| A .15 | <i>Real-Time Data Transmission</i> | <i>Specialisation of Data Transmission which handles data streams using specified real-time transport protocols.</i> |
| A .16 | Data Transmission Monitoring | Functionality that checks and reports the status of data transferring process against specified performance criteria. |

Data Curation (B)

| | Functions | Definitions |
|--|-----------|-------------|
|--|-----------|-------------|

| | | |
|----------|-----------------------------|---|
| B .1 | Data Quality Checking | Functionality that detects and corrects (or removes) corrupt, inconsistent or inaccurate records from data sets. |
| B .2 | Data Quality Verification | Functionality that supports manual quality checking. |
| B .3 | Data Identification | Functionality that assigns (global) permanent unique identifiers to data products. |
| B .4 | Data Cataloguing | Functionality that associates a data object with one or more metadata objects which contain data descriptions. |
| B .5 | Data Product Generation | Functionality that processes data against requirement specifications and standardised formats and descriptions. (optional /may be null) |
| B .6 | Data Versioning | Functionality that assigns a new version to each state change of data, allows to add and update some metadata descriptions for each version, and allows to select, access or delete a version of data. |
| B .7 | Workflow Enactment | Functionality that interprets predefined process descriptions and control the instantiation of processes and sequencing of activities, adding work items to the work lists and invoking application tools as necessary. |
| B .8 | Data Storage & Preservation | Functionality that deposits (over long-term) the data and metadata or other supplementary data and methods according to specified policies, and makes them accessible on request. |
| B .9 | Data Replication | Functionality that creates, deletes and maintains the consistency of copies of a data set on multiple storage devices. |
| B .10 | Replica Synchronisation | Functionality that exports a packet of data from on replica, transports it to one or more other replicas and imports and applies the changes in the packet to an existing replica. |

Data Publishing (C)

| | Functions | Definitions |
|----------|--------------------------------|--|
| C .1 | Access Control | Functionality that approves or disapproves of access requests based on specified access policies. |
| C .2 | Resources Annotation | Functionality that creates, changes or deletes a note that reading any form of text, and associates them with a computational object. |
| C .3 | <i>Data Annotation</i> | <i>Specialisation of Resource Annotation which allows to associate an annotation to a data object.</i> |
| C .4 | Metadata Harvesting | Functionality that (regularly) collects metadata (in agreed formats) from different sources. |
| C .5 | Resource Registration | Functionality that creates an entry in a resource registry and inserts resource object or a reference to a resource object in specified representations and semantics. |
| C .6 | <i>Metadata Registration</i> | <i>Specialisation of Resource Registration, which registers a metadata object in a metadata registry.</i> |
| C .7 | <i>Identifier Registration</i> | <i>Specialisation of Resource Registration, which registers an identifier object in an identifier registry.</i> |
| C .8 | <i>Sensor Registration</i> | <i>Specialisation of Resource Registration which registers a sensor object to a sensor registry.</i> |
| C .9 | Data Conversion | Functionality that converts data from one format to another format. |
| C .10 | Data Compression | Functionality that encodes information using reduced bits by identifying and eliminating statistical redundancy. |

| | | |
|----------|---------------------------|---|
| C .11 | Data Publication | Functionality that provides clean, well-annotated, anonymity-preserving datasets in a suitable format, and by following specified data-publication and sharing policies to make the datasets publicly accessible or to those who agree to certain conditions of use, and to individuals who meet certain professional criteria. |
| C .12 | Data Citation | Functionality that assigns an accurate, consistent and standardised reference to a data object, which can be cited in scientific publications. |
| C .13 | Semantic Harmonisation | Functionality that unifies similar data (knowledge) models based on the consensus of collaborative domain experts to achieve better data (knowledge) reuse and semantic interoperability. |
| C .14 | Data Discovery and Access | Functionality that retrieves requested data from a data resource by using suitable search technology. |
| C .15 | Data Visualisation | Functionality that displays visual representations of data. |

Data Processing (D)

| | Functions | Definitions |
|----------|--------------------------------------|--|
| D .1 | Data Assimilation | Functionality that combines observational data with outputs from a numerical model to produce an optimal estimate of the evolving state of the system. |
| D .2 | Data Analysis | Functionality that inspects, cleans, and transforms data, providing data models which highlight useful information, suggest conclusions, and support decision making. |
| D .3 | Data Mining | Functionality that supports the discovery of patterns in large data sets. |
| D .4 | Data Extraction | Functionality that retrieves data out of (unstructured) data sources, including web pages ,emails, documents, PDFs, scanned text, mainframe reports, and spool files. |
| D .5 | Scientific Modelling and Simulation | Functionality that supports the generation of abstract, conceptual, graphical or mathematical models, and to run an instances of those models. |
| D .6 | <i>Scientific Workflow Enactment</i> | <i>Functionality provided as a specialisation of Workflow Enactment supporting the composition and execution of computational or data manipulation steps in a scientific application. Important processing results should be recorded for provenance purposes.</i> |
| D .7 | Scientific Visualisation | Functionality that graphically illustrates scientific data to enable scientists to understand, illustrate and gain insight from their data. (optional or may be null) |
| D .8 | Service Naming | Functionality that encapsulates the implemented name policy for service instances in a service network. |
| D .9 | Data Processing Control | Functionality that initiates calculations and manages the outputs to be returned to the client. |
| D .10 | Data Processing Monitoring | Functionality that checks the states of a running service instance. |

Data Use (E)

| | Functions | Definitions |
|---------|--------------------------|--|
| E. 1 | Authentication | Functionality that verifies a credential of a user. |
| E. 2 | Authorisation | Functionality that specifies access rights to resources. |
| E. 3 | Accounting | Functionality that measures the resources a user consumes during access for the purpose of capacity and trend analysis, and cost allocation. |
| E. 4 | <i>User Registration</i> | <i>Specialisation of Resource Registration which registers a user to a user registry.</i> |

| | | |
|---------|------------------------------|---|
| E. 5 | Instant Messaging | Functionality for quick transmission of text-based messages from sender to receiver. |
| E. 6 | Interactive Visualisation | Functionality that enables users to control of some aspects of the visual representations of information. |
| E. 7 | Event Notification | Functionality that delivers message triggered by predefined events. |