**ENVRIplus Theme 2:**

Requirements information gathering exercise

***Your RI name here***

*RI representative(s):*

* *Firstname Lastname,*

*Institute*

Version date

**1. Identification and citation**

*While some of the questions should be quite easy to answer, others could require a bit of thinking and effort, and then there are also a few that are probably not applicable to your Research Infrastructure.*

*In any case, it is preferable that you give a rough estimate or a very schematic response, than leaving a question unanswered. If you get stuck, either get in touch with me (Maggie Hellström at margareta.hellstrom@nateko.lu.se), or simply write "can't answer", "don't understand" or "not applicable"...*

*Please send the completed questionnaire to Maggie by* ***18:00 on Tuesday, December 15 2015*** *at the very latest!*

**IDENTIFICATION**

1. What granularity do your RI’s data products have:
   1. Content-wise (all parameters together, or separated e.g. by measurement category)?

<your answer here>

* 1. Temporally (yearly, monthly, daily, or other)?

<your answer here>

* 1. Spatially (by measurement station, region, country or all together)?

<your answer here>

1. How are the data products of your RI stored - as separate “static” files, in a database system, or a combination?

<your answer here>

1. How does your RI treat the “versioning” of data - are older datasets simply replaced by updates, or are several versions kept accessible in parallel? How do you identify different version of the same dataset?

<your answer here>

1. Is it important to your data users that:
   1. Every digital data object is tagged with a unique & persistent digital identifier (PID)?

<your answer here>

* 1. The metadata for data files contains checksum information for the objects?

<your answer here>

* 1. Metadata (including any documentation about the data object contents) is given its own persistent identifier?

<your answer here>

* 1. Metadata and data objects can be linked persistently by means of PIDs?

<your answer here>

1. Is your RI currently using, or planning to use, a standardized system based on persistent digital identifiers (PIDs) for:
   1. “Raw” sensor data?

<your answer here>

* 1. Physical samples?

<your answer here>

* 1. Data undergoing processing (QA/QC etc.)?

<your answer here>

* 1. Finalized “publishable” data?

<your answer here>

1. Please indicate the kind of identifier system that are you using - e.g. Handle-based (EPIC or DOI), UUIDs or your own RI-specific system?

<your answer here>

1. If you are using Handle-based PIDs, are these handles pointing to “landing pages”? If so, are these pages maintained by your RI or an external organization (like the data centre used for archiving)?

<your answer here>

1. Are costs associated with PID allocation and maintenance (of landing pages etc.) specified in your RI’s operational cost budget?

<your answer here>

**CITATION**

1. How does your “designated scientific community” (typical data users) primarily use your data products? As input for modelling, or for comparisons?

<your answer here>

1. Do your primary user community traditionally refer to datasets they use in publications:
   1. By providing information about producer, year, report number if available, title or short description in the running text (e.g. under Materials and Methods)?

<your answer here>

* 1. By adding information about producer, year, report number if available, title or short description in the References section?

<your answer here>

* 1. By DOIs, if available, in the References section?

<your answer here>

* 1. By using other information?

<your answer here>

* 1. By providing the data as supplementary information, either complete or via a link

<your answer here>

1. Is it important to your data users to be able to refer to specific subsets of the data sets in their citation? Examples:
   1. Date and time intervals

<your answer here>

* 1. Geographic selection

<your answer here>

* 1. Specific parameters or observables

<your answer here>

* 1. Other

<your answer here>

1. Is it important to be able to refer to many separate datasets in a collective way, e.g. having a collection of “all data” from your RI represented by one single DOI?

<your answer here>

1. What strategy does your RI have for collecting information about the usage of your data products?

<your answer here>

* 1. Downloads/access requests

<your answer here>

* 1. Visualization at your own data portal

<your answer here>

* 1. Visualization at other data portals

<your answer here>

* 1. References in scientific literature

<your answer here>

* 1. References in non-scientific literature

<your answer here>

* 1. Scientific “impact”

<your answer here>

1. Who receives credit when a dataset from your RI is cited?
   1. The RI itself

<your answer here>

* 1. The RI’s institutional partners (all or in part, depending on the dataset contents)

<your answer here>

* 1. Experts in the RI’s organization (named individuals)

<your answer here>

* 1. “Principal investigators” in charge of measurements or data processing (named individuals)

<your answer here>

* 1. Staff (scientists, research engineers etc.) performing the measurements or data processing (named individuals)

<your answer here>

1. What steps in tooling, automation and presentation do you consider necessary to improve take up of identification and citation facilities and to reduce the effort required for supporting those activities?

<your answer here>