

Provenance in IS-ENES2

Context of provenance in IS-ENES2

Summary of IS-ENES2 requirements for provenance

Detailed requirements

1. **Do you already have data provenance recording in your RI?** Yes, depending on the data analysis activity

If so:

2. **Where/when do you need it, e.g., in the data processing workflows, data collection/curation procedures, versioning control in the repositories etc.?**

Mostly in data collection procedures as well as data processing workflows

3. **What systems are you using?** Community tools e.g. to manage what has been collected from where, and what is the overall transfer status or e.g. to generate provenance log files in workflows.

4. **What standards are you using?**

i. **Advantages/disadvantages** No standard by now, first experiments toward the use of PROV-O in a specific analysis project.

ii. **Have you ever heard about the PROV-O standard?** Yes.

5. **Do you need provenance tracking?**

i. **If so, which information should be contained?** Input data characteristics (names, characterizing facets, checksum, unique ids), tools used (git svn tags), output files, timing information, platform/environment information.

6. **What information do you need to record regarding the following:**

i. **Scientific question and working hypothesis?**

The data has been produced following a very details experimental protocol. We need to collect all the information needed to assess how exactly the protocol has been followed (facets, control vocabulary, documentation : es-doc.org).

ii. **Investigation design?** Authors information.

iii. **Observation and/or measurement methods?**

iv. **Observation and/or measurement devices?**

v. **Observation/measurement context (who, what, when, where, etc.)?**

vi. **Processing methods, queries?**

vii. **Quality assurance?**

Performed quality assurance procedures, results of QA software.

7. **Do you know/use controlled vocabularies, e.g. ontologies, taxonomies and other formally specified terms, for the description of the steps for data provenance?** Not yet.

8. **What support, e.g. software, tools, and operational procedures (workflows), do you think is needed for provenance tracking?**

Agreements on what information to record and simple APIs to be able to be integrated in analysis tools and frameworks.

9. **How does your community use/plan to use the provenance information?**

-For catalogues as additional metadata for data products.

-For end users to understand the derivation history of data products.

-For tools to automatically "replay" specific analysis parts.

i. **Do you have any tools or services in place/planned for this purpose?** No generic ones – specific loggers, etc.

Formalities (who & when)

Go-between	Yin Chen
RI representative	Sylvie Joussaume < sylvie.joussaume@lsce.ipsl.fr > Francesca Guglielmo < francesca.guglielmo@lsce.ipsl.fr >

Period of requirements collection	Oct -Nov 2015
Status	Completed