

Provenance in LTER

Context of provenance in LTER

Summary of LTER requirements for provenance

Detailed requirements

1. Provenance

a. Which information should be retained if your RI needs provenance tracking?

Provenance should include the generation as well as the modification of the data. Further the use (e.g. download) as well as storage (where are the data) should be tracked. This links also to the use of community services, e.g. EUDAT B2SHARE, for data storage and exchange.

The line between metadata describing the 'dataset' and provenance needs to be defined. It can be seen as part of the MD.

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

i. Scientific question and working hypothesis?

ii. Investigation design?

yes

iii. Observation and/or measurement methods?

Yes – current part of the data set MD – method; use of a method registry would be appreciated in order to harmonise and share method information.

iv. Observation and/or measurement devices?

Yes – currently part of the method description. For SOS (sensor observation service) part of the sensor specification.

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

Yes – currently part of the dataset MD

c) Do you know/use controlled vocabularies, e.g. ontologies, taxonomies and other formally specified terms, for the description of the steps for data provenance?

Yes – EnvThes (<http://vocabs.lter-europe.net/EnvThesDev.html>). Is currently transferred to TopBraid software. The thesaurus is used for keyword tagging, but will also be used as a registry for reference lists used in the data context. In addition it should contain concepts (e.g. substances) used in the data

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

Still open – hope ENVRI could work on that

Formalities (who & when)

Go-between	@Barbara Magagna
RI representative	@Johannes Peterseil
Period of requirements collection	@20150901
Status	gathered