Documenting Data Provenance

If you have used other data sources in the creation of your data package, it is important to document data **provenance**, i.e., the origin or history of data used in the creation of the current data package. Provenance metadata contributes to making your research transparent and reproducible, and it lets you give credit to the creators of data you have used.

To specify provenance for your data package, EDI recommends that you create a **Method Step** and add to it one or more **Data Sources** that refer to the data package(s) that were used to create the current data package. Such a provenance method step is a sort of "dummy" method step, a placeholder to which data sources can be attached. It is customary to place such provenance method steps at the end of the method steps list.

To create a provenance method step, click **Methods** in the Contents pane, then click **Add Method Step**. In the method step's **Description** text box, enter something like the following: "This method step describes provenance-based metadata as specified in the LTER EML Best Practices."

Method Step

Enter the method step description and (optionally) instrumentation and data sources below:

Description *

> This method step describes provenance-based metadata as specified in the LTER EML Best Practices.

ⓘ

Instrumentation (Optional)

ⓘ

Data Sources (Optional)

Data Source Title

[Add Data Source] [Fetch Data Source Info from EDI]

ⓘ

[Save and Continue] [Cancel]

Then, specify the data sources you have used. If a data source is a data package that has been published in EDI's data repository, you can use **Fetch Data Source Info from EDI** to retrieve the provenance metadata for that data source and fill it in automatically. First, select the *scope* to which the data source package belongs by clicking one of the links in the list (only part of the list is shown below).

Fetch Data Source Info from EDI ⓘ

Select a data package *scope* value to view a list of associated identifiers:

ecotrends
edi
knb-lter-and
knb-lter-arc
knb-lter-bes
knb-lter-ble
knb-lter-bnz
knb-lter-cap
knb-lter-cce
knb-lter-cdr

Let's say we want to fetch provenance metadata for data package edi.100.4. We click the **edi** link. This brings up a list of *identifier* values for packages in the edi scope (only part of the list is shown below). Click the link for edi.100.

Fetch Data Source Info from EDI ⓘ

Select the data package *scope.identifier* for the data source:

edi.1
edi.2
edi.3
edi.4
edi.5
edi.6
edi.7
edi.8
edi.9
edi.12
edi.13
edi.14

If only one *revision* exists for the data package, its provenance metadata will be retrieved immediately. Otherwise, you will see a list of available revisions. Click the one you want. For edi.100, there are two revisions at the time of this writing.

Fetch Data Source Info from EDI ⓘ

Select a *revision* of data package **edi.100**:

1
4

Cancel

We click the link for revision 4, and the provenance metadata is fetched and added as a data source:

Method Step

Enter the method step description and (optionally) instrumentation and data sources below:

Description *

> This method step describes provenance-based metadata as specified in the LTER EML Best Practices.

Instrumentation (Optional)

Data Sources (Optional)

Data Source Title		
LAGOS-NE-LOCUS v1.01: A module for LAGOS-NE, a multi-scaled geospat...	Edit	Remove

Add Data Source   Fetch Data Source Info from EDI

Save and Continue   Cancel

As usual, to inspect or edit the data source, click the **Edit** button for it. This brings up a page that looks like the following:

Data Source ⑦

Title *

LAGOS-NE-LOCUS v1.01: A module for LAGOS-NE, a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. Lakes: 1925-2013

Online Description (Recommended)

This online link references an EML document that describes data used in the creation of this derivative data

URL (Recommended)

https://pasta.lternet.edu/package/metadata/eml/edi/100/4

Creators ⑦

Patricia A Soranno, Michigan State University   ▼   Edit   Remove

Kendra S Cheruvelil, Michigan State University   ▲   Edit   Remove

Add Creator

Contacts ⑦

Patricia A Soranno, Michigan State University   ▼   Edit   Remove

Information Manager   ▲   Edit   Remove

Add Contact

Save and Continue   Cancel

Everything has been filled in automatically. The **URL** points to the data source's metadata in the EDI data repository. The **Online Description** describes the link, and a suggested description has been provided, which you can modify as you see fit.

If your data source is not one that has been published in EDI's data repository, use **Add Data Source** to bring up an empty Data Source page, which you can then fill in. If the data source is available online, in place of a URL you can, if you prefer, use a DOI or other globally unique identifier. If the data source is not available online, leave the Online Description and URL fields empty.

Perform these steps for each of the data sources used in the creation of your data package. All of the data sources can be attached to a single provenance method step.