Customizing a DDI compliant repository: a detailed course with Dataverse

Geneviève Michaud, Baptiste Rouxel

SciencesPo
CENTRE DE DONNÉES SOCIO-POLITIQUES

Customizing a DDI compliant repository: a detailed course with Dataverse

Baptiste Rouxel, Geneviève Michaud,
Center for Socio-Political Data

EDDI 2020, 12th Annual European DDI Users Conference
DOI: 10.5281/zenodo.4298929
December, 2, 2020

Customizing a DDI compliant repository: a detailed course with Dataverse

Tackling issues, sharing tools and processes

Outline

1. Metadata model customization
2. Setting up multilingualism
3. User interface, branding

Project objectives

Set up a CESSDA Metadata Model (CMM) compliant Dataverse repository

- Customize Dataverse metadata model

Migrate CDSP study level metadata from Nesstar

- Upgrade DDI files and harmonize metadata
- Keep study + variable level metadata as DDI files for each dataset in the repository

Metadata model customization

Dataverse metadata model customization: .tsv files

Customizing a Dataverse original metadata block involves modifying a tab separated value file. A ".tsv" file targets a specific "metadata block". Operation include:

1. uploading the modified TSV file using the native API
2. triggering a SolR index update

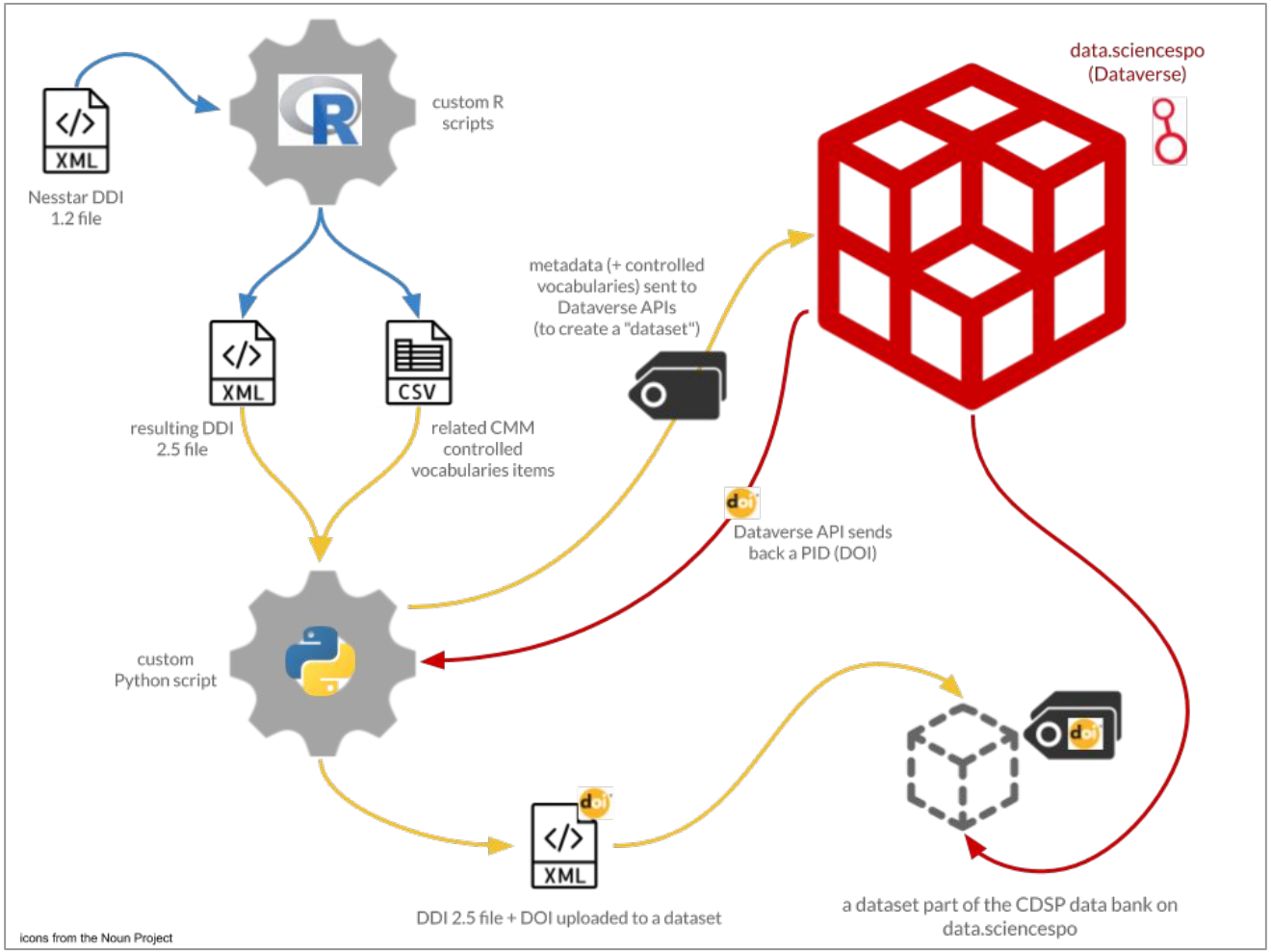
Our custom TSV is available on GitHub here:

CDSP-SCPO/dataverse-controlledvocabulary

CDSP databank migration from Nesstar to Dataverse

Custom R script
CDSP-SCPO/Nesstar2Dataverse

Custom Python scripts
CDSP-SCPO/dataverse-controlledvocabulary
CDSP-SCPO/dataverse-metadataupdate
CDSP-SCPO/dataverse-jsonimport
CDSP-SCPO/dataverse-ddiimport



Dataverse metadata model customization: issues

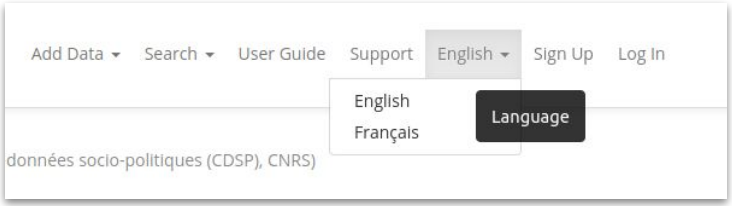
Adding metadata fields to the metadata model is quite easy (Create feature).
But this must be done **early** in the process, since revisions (Update and Delete features) require specific requests on underlying database with admin privileges. Further revisions should definitely be avoided (complexity + risk).

The DDI import API might not work properly if the metadata model is customized.

Multilingualism creates an additional layer of complexity.

Setting up multilingualism

Multilingualism: user interface



Adding languages is possible from Dataverse v4 onwards.

Steps (link to Dataverse official documentation):

- Declare new languages that will appear in the dropdown menu
- Upload a language bundle file (see the documentation to get your languages and create the bundle)

Language translations are maintained by the community and you can ask to contribute.

GlobalDataverseCommunityConsortium/dataverse-language-packs

Managing multilingualism: controlled vocabularies

a language bundle = multiple language property files for each language

In the language property files, you can translate a controlled vocabulary by adding a property for each language available.

> **Example:**
> In the metadatablock TSV file, if `Self-administered questionnaire: Web-based (CAWI)` is a controlled vocabulary item value for `ModeOfCollection`, add the following line in your language property file as described in the Dataverse documentation:

```
controlledvocabulary.modeofcollection.self-administered_questionnaire:_web-based_(cawi)=TRANSLATED_VALUE
```

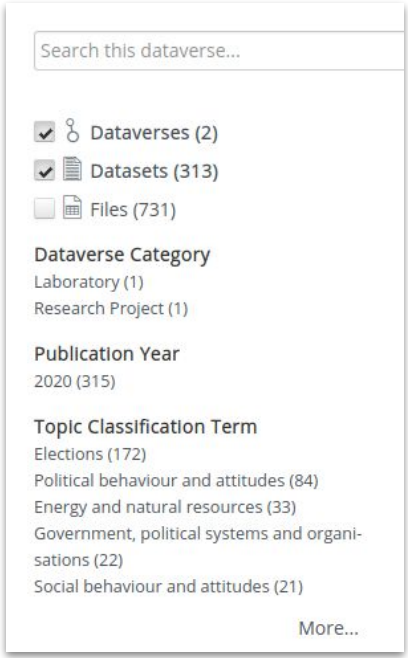
Managing multilingualism: controlled vocabularies

This process can be bit tedious and has a high risk of error if it is done manually, so we wrote a script that:

- Creates language property files for each language and different metadata blocks
- Transforms TSV controlled vocabulary values to property names
- Adds translated value to each property

CDSP-SCPO/dataverse-controlledvocabulary

Managing multilingualism: controlled vocabularies



Managing multilingualism: remaining issues

There is room for improvement in the way Dataverse follows DDI schema:

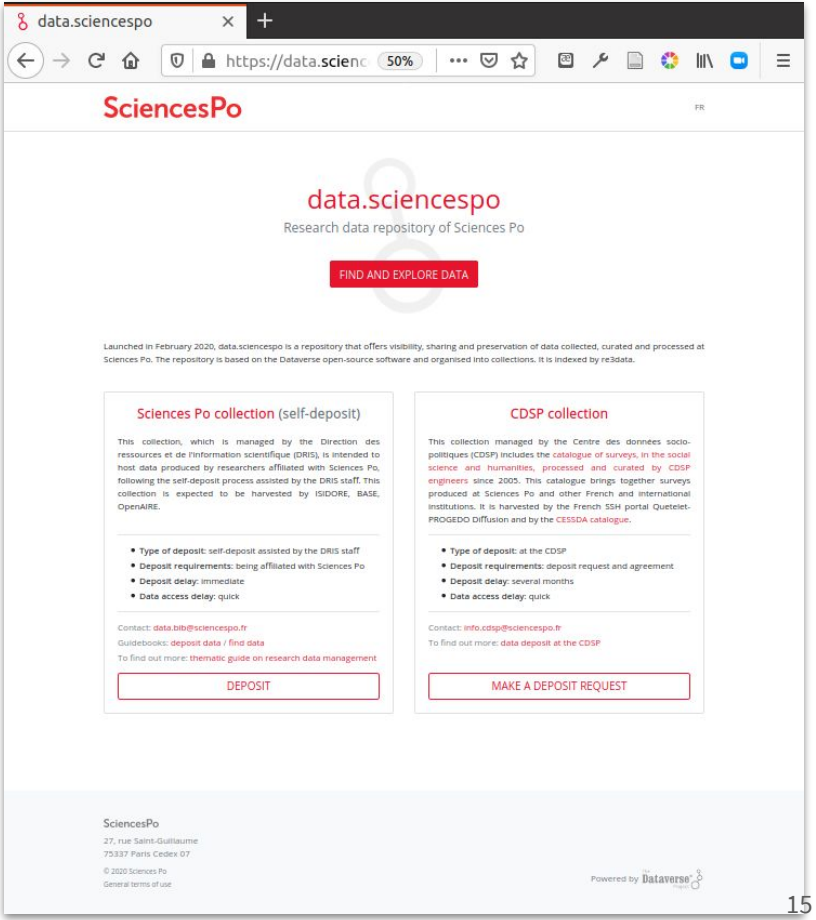
- As an example, it's not possible to save and display a **title** in more than one language.
- A specific group in Dataverse community is actively working on **flexible metadata.**

User interface, branding

Repository landing page

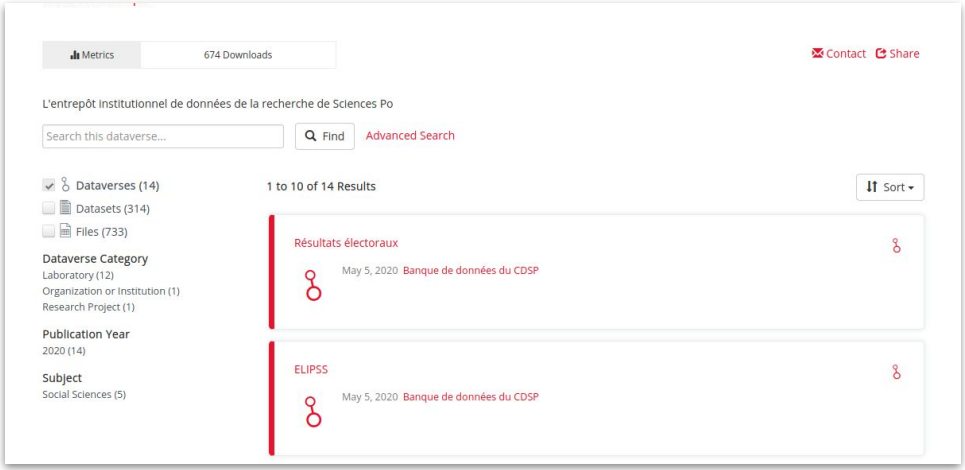
- Dataverse landing page, headers and footers are customizable
- A clear visual setup to help users navigate between different services
  - Search for data
  - Deposit data (self deposit)
  - Contact data and metadata curators (CDSP)
- Available in french and english

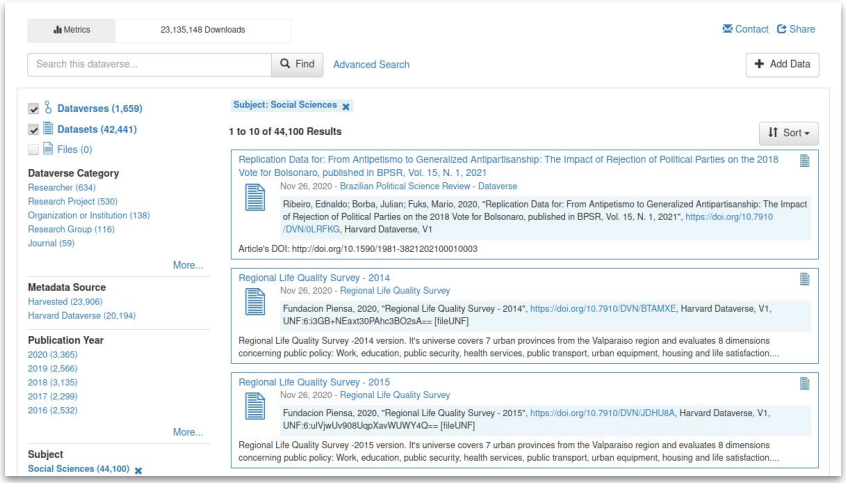
FIND AND EXPLORE DATA



CDSP-SCPO/dataverse-branding

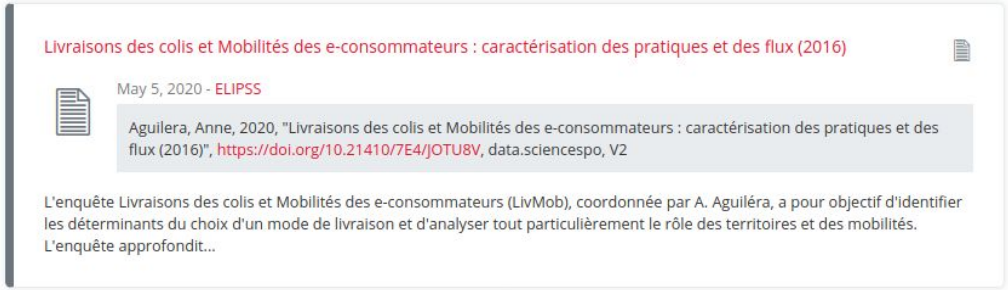
User interface: consistency with our design guidelines

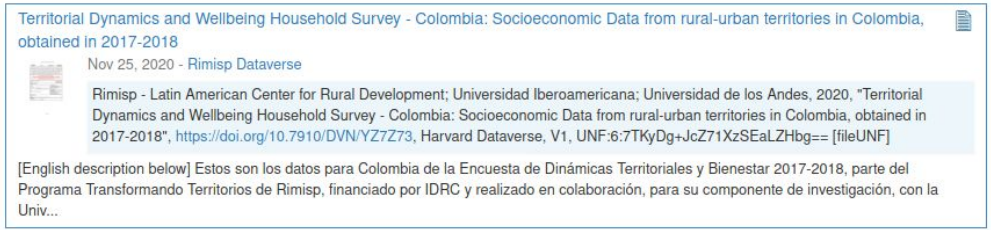




CDSP-SCPO/dataverse-branding

User interface: consistency with our design guidelines





CDSP-SCPO/dataverse-branding

Next steps: improving our repository

We are currently collaborating with Dataverse on:

- Multilingualism (issue #6607)
- DDI compliance (DDI import/export, OAI-PMH/OAI-DDI) (issues #7388 #7387 #6751)

Merci !

If you have any question, contact us!

itcdsp-scpolt@sciencespo.fr