



HAL
open science

Dataverse as a Single Source of Truth in a DDI-Compliant Environment

Quentin Agren, Simon Dellac, Geneviève Michaud, Romain Mougin, Baptiste Rouxel

► **To cite this version:**

Quentin Agren, Simon Dellac, Geneviève Michaud, Romain Mougin, Baptiste Rouxel. Dataverse as a Single Source of Truth in a DDI-Compliant Environment. Annual European DDI User Conference, Dec 2018, Berlin, Germany. hal-02874088

HAL Id: hal-02874088

<https://sciencespo.hal.science/hal-02874088>

Submitted on 19 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Dataverse as a Single Source of Truth in a DDI-Compliant Environment

Quentin AGREN

Simon DELLAC

Geneviève MICHAUD

▶ Romain MOUGIN

▶ Baptiste ROUXEL (all CDSP, Sciences Po)

10th Annual European DDI User Conference

December 4 - 5, 2018

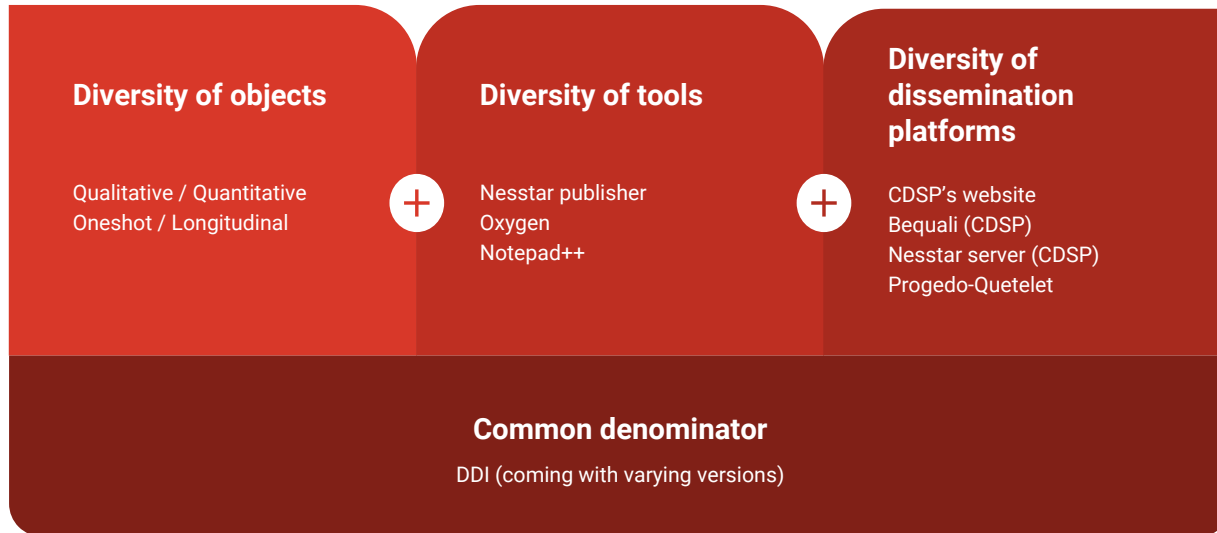
Berlin, Germany

Presentation outline

1. A diverse documentation landscape
2. Towards a SSOT: implementing DDI import into Dataverse
3. Enforcing consistency between published content and documentation

1. A diverse documentation landscape

Current state of our documentation landscape



2. Towards a SSOT: implementing DDI import into Dataverse

Why Dataverse?

- Open source development model with an active community
- OAI-PMH harvesting
- PID for each dataset / file
- Faceted search
- DDI export
- Multiple citation formats

No satisfactory DDI import functionality available

An import procedure of datasets exists, but designed for internal use:

- feature undocumented and only available to super users
- poor error reporting
- violations of DDI semantics (multiplicity, content, tag types, version inference)

➔ Let's contribute a user-facing API to import DDI files!

Lessons learned

- Interest and abundant feedback from the dataverse community on the topic of DDI import
- Steep learning curve to get started (environment, codebase...)
- Responsiveness and support from core developers
- Cost-effectiveness compared to home-brewed solution

3. Enforcing consistency between published content and documentation

Enquête sur les valeurs, l'environnement et l'énergie - vague 2 (2015)

Enquête quantitative / Opinions politiques, Environnement, ELIPSS

Auteurs : Béatrice HAMMER, Mathieu BRUGIDOU, Monique WACH

Producteurs : CDSP, INED

Présentation Fiche technique

Présentation

Résumé

L'Enquête sur les valeurs, l'environnement et l'énergie (EVALENE), coordonnée par Béatrice Hammer du Groupe de Recherche Energie Technologie Société (GRETS) a été menée dans la phase test du dispositif ELIPSS.

Elle porte sur la perception de l'environnement, les préoccupations environnementales, le changement climatique, le rôle et l'avenir des différentes énergies, la maîtrise de la consommation d'énergie ainsi que les Intentions d'actions pro-environnementales.

Deux approches contrastées des valeurs sont mises à l'épreuve dans EVALENE : le modèle décontextualisé de Schwartz et l'approche contextualisée par les discours. Le parti-pris du projet est de tester dans une même enquête ces deux approches - généralement dissociées dans les protocoles d'enquête - pour apprécier leur complémentarité.

Après une première vague en octobre 2013, l'enquête EVALENE a fait l'objet d'une nouvelle administration en novembre 2015. Cette seconde vague était l'occasion de tester une autre formulation des valeurs du modèle de Schwarz avec un mode de questionnement plus proche de celui de Rokeach.

Fiche technique

Méthodologie

L'étude longitudinale par Internet pour les sciences sociales (ELIPSS) est un panel Internet représentatif de la population résidant en France métropolitaine. Les panélistes sont sélectionnés aléatoirement. Ils répondent tous les mois à des enquêtes élaborées par des chercheurs, à l'aide d'une tablette et d'un abonnement Internet mobile mis à leur disposition. Les informations démographiques et socio-économiques proviennent de l'enquête annuelle.

Le questionnaire de l'enquête EVALENE (vague 2) a été auto-administré sur Internet, via l'application ELIPSS, installée sur la tablette 7 pouces mise à disposition des répondants.

Explorer les données

Commander les données

Attitudes towards environment and Schwartz values - 2nd wave (2015)

Quantitative survey / Political opinions, Environment, ELIPSS

Authors : Béatrice HAMMER, Mathieu BRUGIDOU, Monique WACH

Producers : CDSP, INED

Presentation Technical report

Presentation

Abstract

The survey "Attitudes towards environment and Schwartz values" (EVALENE), coordinated by Béatrice Hammer from the Energy Technology Society Research Group (GRETS), was conducted in the test phase of the ELIPSS panel. It focuses on the perception of the environment, environmental preoccupations, climate change, the role and future of different energy sources, the control of energy consumption, as well as intentions for pro-environmental actions. Two contrasting approaches to values are tested in EVALENE: Schwartz's decontextualised model and the model of contextualisation through discourse. The objective of the project is to test these two approaches - which are generally separated in survey protocols - within a single survey, in order to assess their complementarity. After a first wave in October 2013, the EVALENE survey was conducted again in November 2015. This second wave was an opportunity to test another formulation of the values in the Schwartz model with a method of questioning more like the Rokeach model.

Explore data

Download data

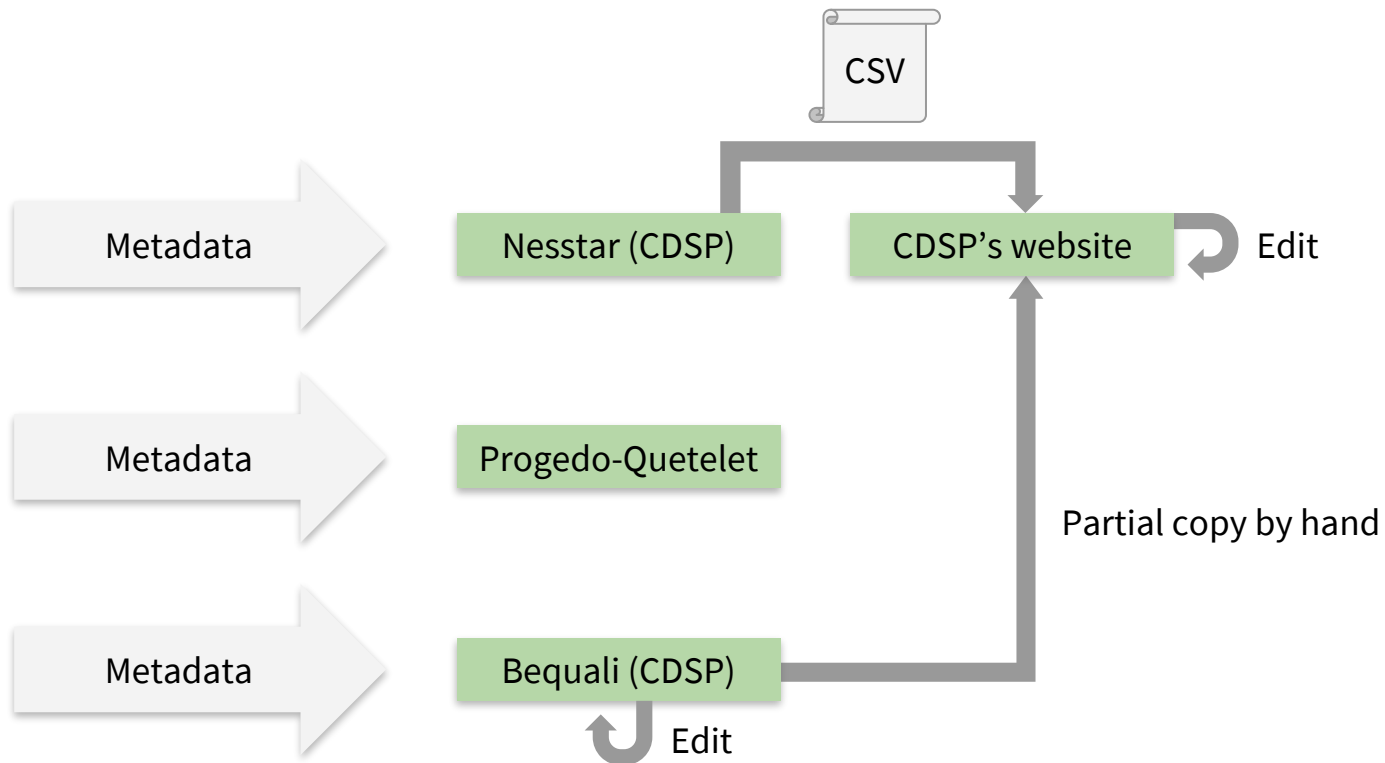
Technical report

Methodology

Self-administered online questionnaire. Access via the ELIPSS survey app, installed on the 7-inch tablet supplied to panellists.

Sampling

The sampling method used is a multi-stage sampling carried out by INSEE's Survey Division. From the 2011 census, INSEE extracted a sample of 4500 dwellings using stratified cluster sampling on the following criteria: Strata: Region (excluding Corsica) and type of municipality (rural/urban, large/small). Clusters: ZAE (survey action zones). This sample was divided into three subsamples: a main section (3500 dwellings) and two reserve sections in order to provide 1000 additional addresses for use if necessary. In the end, all 4500 dwellings extracted by INSEE were used. After the agreement of the household had been obtained (1349 households), the final step was to select one individual from the eligible individuals in the household

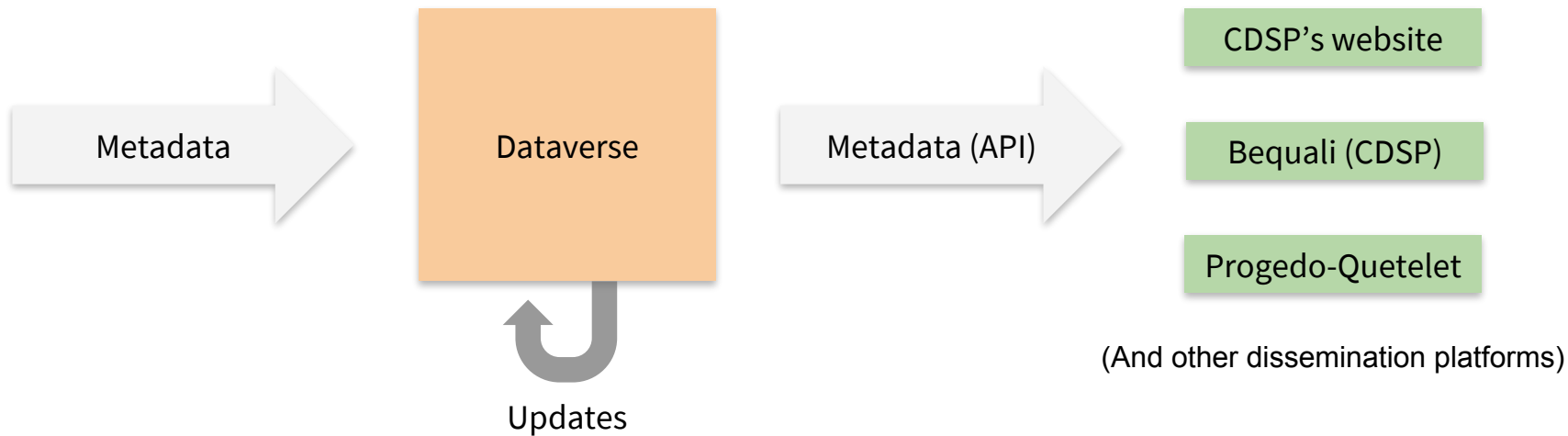


Brittle solution: edit metadata on the publishing platforms

Robust solution: Dataverse as SSOT accessed through web APIs

Dissemination platforms query metadata through web APIs

- Changes reflected everywhere automatically
- Simplified editing and publishing workflow



Robust solution: Dataverse as SSOT accessed through web API

Next Step

Development prospect: Dataverse webhooks

- Make the SSOT push changes to users, rather than having users periodically poll for changes
- Idea: allow specifying URLs at the Dataverse collection level, which should be POSTed to each time metadata changes are made

Do you have any questions?

romain.mougin@sciencespo.fr
baptiste.rouxel@sciencespo.fr