Providing DDI metadata for OAI-PMH harvesting a Dataverse repository

Geneviève Michaud^{*1}, Baptiste Rouxel¹, and Tom Villette¹

¹Centre de données socio-politiques de Sciences Po (Sciences Po, CNRS) – Sciences Po, Centre National de la Recherche Scientifique : UAR828 / UMS828 – France

Abstract

The approach that leads to the adoption of DDI as a standard to document resources from a research data bank, is, in a great share, the promise of users being able to use and compare different data sources. To facilitate discovery and searching, aggregator portals gather different sources of metadata in one place. The OAI-PMH protocol (Open Archives Initiative's Protocol for Metadata Harvesting) is widely used in this context for compiling metadata from various sources, including but not limited to data repositories.

The DDI documentation standard has been a toolbox of choice at the CDSP of Sciences Po from the start of its documentation activities in 2006. Since 2016, the CDSP has been carrying out experiments with the open-source repository software Dataverse of the Dataverse Project[®] that is being developed at Harvard Institute for Quantitative Social Science (IQSS), along with collaborators and contributors worldwide and supported by the Global Dataverse Community Consortium. First dedicated to the Human and Social Sciences, Dataverse is now widely used beyond this first domain, and is the underlying infrastructure of several national repositories as in the Netherlands, Canada, France or more recently Denmark. From 2020 onwards, the CDSP is in charge of Science Po's institutional research data repository, data.sciencespo.

Among numerous assets, Dataverse features an OAI-PMH server that provides embedded metadata either Dublin Core or DDI. The CDSP relies on this facility to act as a metadata provider for several portals of various scope and characteristics (domain, languages of sources and research output types). In the course of our dialogue with the OAI-PMH service providers, we have identified several points of attention that we will share here. In addition, we will also highlight how metadata describing data repositories available on portals such as re3data or FAIRsharing.org, can contribute to support data discovery by end users.

^{*}Speaker