
Towards Machine-assisted Disclosure Assessment with DDI-CDI, DPV and sdcMicro

Deidre Lungley^{*1}, Darren Bell¹, Thomas Gilders¹, and Artiom Rumiancev¹

¹UK Data Archive – United Kingdom

Abstract

At the UK Data Archive (UKDA) we have a vision for interoperability of sensitive and disclosive data in the next decade. We will demonstrate a "Data Product Builder": a tool which guides researchers through selecting and building a subsetting data product from one or more linked datasets and dynamically filters the responses and variables as appropriate for their research. With this tool, powered by DDI-CDI metadata, we aim to break open the current binary utility/risk trade-off of Secure Access or Open Access when it comes to sensitive/disclosive datasets, particularly in Social Science. Study and variable metadata, enhanced with Data Privacy Vocabulary (DPV) concepts, feed into risk functions provided by sdcMicro (a well-used command-line tool we have rewritten as Python web services) which drive machine-assisted disclosure assessment, which in turn feeds a decision tree of real-time access outcomes, providing researchers with richer and more flexible choices based on real-time mitigations of key variable sensitivity e.g. (global recoding, top/bottom recoding, etc.). As well as a live demo of the tool, we outline the wider context of interoperability for sensitive data including other ontologies such as Data Use ontology and Open Digital Rights Language.

^{*}Speaker