



Contribution ID: 46

Type: **Poster**

The Path to Open Access for Restricted Data with the Data Product Builder

Wednesday, 29 November 2023 13:05 (20 minutes)

With the Data Product Builder (DPB) the UK Data Service (UKDS) aims to allow researchers access to on-demand linked subsets of data, dynamically assessed for emergent disclosure risk in real time. Such a system depends both on sophisticated upstream metadata and powerful downstream computation. We detail the pipeline components required to take original curated data in traditional dissemination formats, e.g. SPSS, and make it available as a secure RDF linked data resource. These components encompass:

- Transforming the binary files into a highly granular structural representation in DDI-CDI
- Using machine learning models to aid automation of metadata enhancement, e.g. determining 'key' variables for Disclosure Risk Analysis
- Aligning study variable representation with aggregate census variable representation to allow us to benchmark 'risk' using population statistics
- Transforming the user determined data product into RDF triples, as well as permitting further deserialization into multiple binary formats - SPSS, STATA, Excel etc., - if required for download and/or desktop analysis.

This pipeline results in a new type of digital resource, where data products are dynamically built by the researcher based on their individual research need, augmented by real-time disclosure risk mitigations, unblocking their access to data which has traditionally been difficult and time-consuming to procure.

Primary authors: LUNGLEY, Deirdre; Mr GILDERS, Thomas (University of Essex)

Presenter: LUNGLEY, Deirdre

Session Classification: Posters

Track Classification: Proposed Topics of the Conference: Software and Tools