

Sensitive Data Interest Group kickoff discussion

Attendees:

Johannes Darms - Germany national library of medicine. Challenges to centralize RDM due to sensitivity of data.

Katie Mika - Harvard Library & IQSS

Falco (Denmark) - 3 national services: dataverse, storage for open data and storage for sensitive data. How to integrate

Dylan Dearborn (Borealis) - Pilot tech implementation (U of T dataverse with institutions having some control over data in their collections) for secure storage for sensitive data Institutional shared custodianship model (onus on institution).

Rita - University of Aveiro (Portugal) - Interested in learning more!

Nada Chaya (Lebanon) - Social science funding agency. Dataverse installation. Only accept deidentified datasets. Help researchers deidentify data if necessary. Interested in learning more to avoid reid risk

Lina Harper (Borealis) - Sensitive data governance strategy to craft a national framework for Borealis. Curation events and trainings.

Discussion

Are regulations compatible globally? More regulations are happening, likely to get more diverse, not more coordinated.

Levels of sensitivity do not map well between institutions and countries

Want to start with an interest group (monthly?). Slack channel to be created in community space

Potential invited speaker (Proposal from Johannes - follow up!)

Workflows for technological implementation.

Data Use Ontology?

Read Jim's white paper on Data Tags and adding flexibility: link?

Access methods?

Data Use agreements - can we add support for DUAs directly in dataverse?

Ownership of data - who has what rights to research data and how does this differ by institution by country?

Do institutions have close relationships with their research administration and compliance offices?

General problems with researcher enabled access to data, for sensitive data, but also non sensitive data.

References! Can we add better structure to the metadata fields that describe “cited datasets” in the same way as the related pub field? But what about different types of reuse? For sensitive data there’s use cases of adding sensitive variables/values/etc to derivative datasets.

Use cases of sensitive data reuse?

GBIF! DOIs for collections

Registries and remote storage

- How do we direct users to the correct storage locations?
- Can we have datasets that have files in different storage locations? Should they be different DOIs? Is it another digital artifact if it lives in a different place?
- Would a linked data type graph make things like this clearer and more machine readable? How on earth would we get a user to enter this information???
- The general LOD in dataverse convo is relevant for sensitive data