Research Data Alliance Activities on FAIR Sensitive Data

Open Science with Sensitive Data 15.5.2019
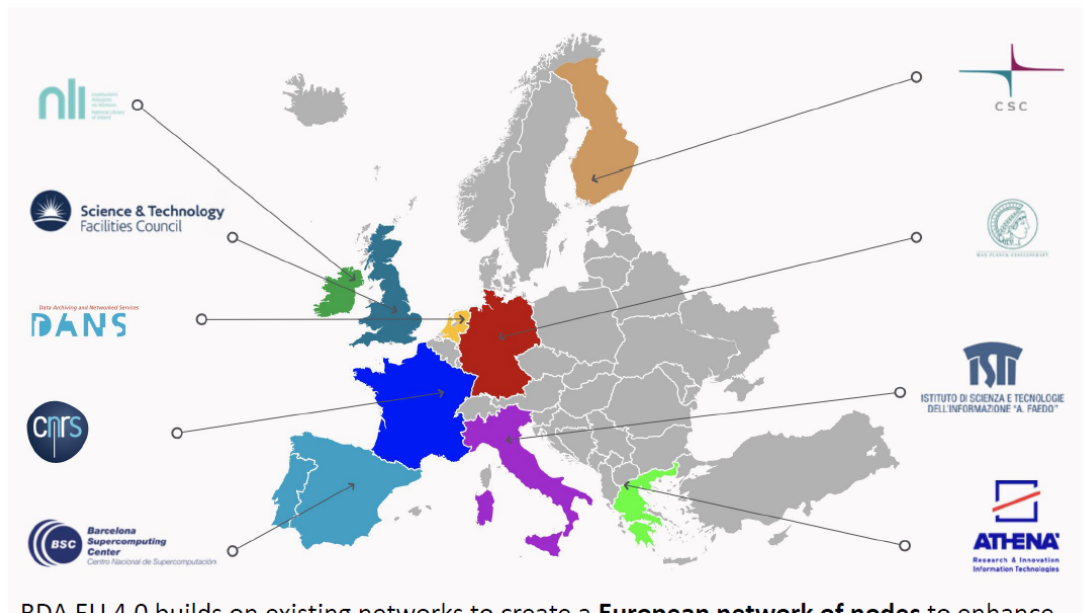
Heidi Laine, CSC

CSC
ICT Solutions for Brilliant Minds

CSC Coordinates RDA Finland

RDA EU 4.0 and the National Nodes

RDA EU 4.0 builds on existing networks to create a **European network of nodes** to enhance Member State cooperation and support of RDA in many European countries & vice versa the representation of EU countries in RDA Global
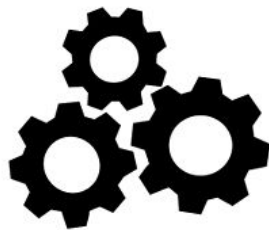
New gold standard for Open Science compliant data

Findable **A**ccessible **I**nteroperable **R**eusable

Spoiler!

RDA hasn't produced much on sensitive data

Possible factors:

- RDA is more of a community of data managers and service providers then researchers, not a lot of topic specific outputs overall

- Sensitive data management is an emerging concern

- Thinking about sensitive data is hard and hurts your head

- We have all the recommendations, standards, protocols, and best practices that we can possibly need for sensitive data!

Types of RDA Outputs

RDA Endorsed Recommendations (10)

  ○ RDA Recommendations are the official, endorsed results of RDA and considered our "flagship" Outputs.

Recommendations with RDA Endorsement in Process (8)

RDA Supporting Outputs (12)

  ○ Supporting Outputs are useful solutions from our RDA Working and Interest Groups, but may not be as clearly adoptable by organisations as our RDA Recommendations.

All found here: https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

RDA outputs across the Workflow
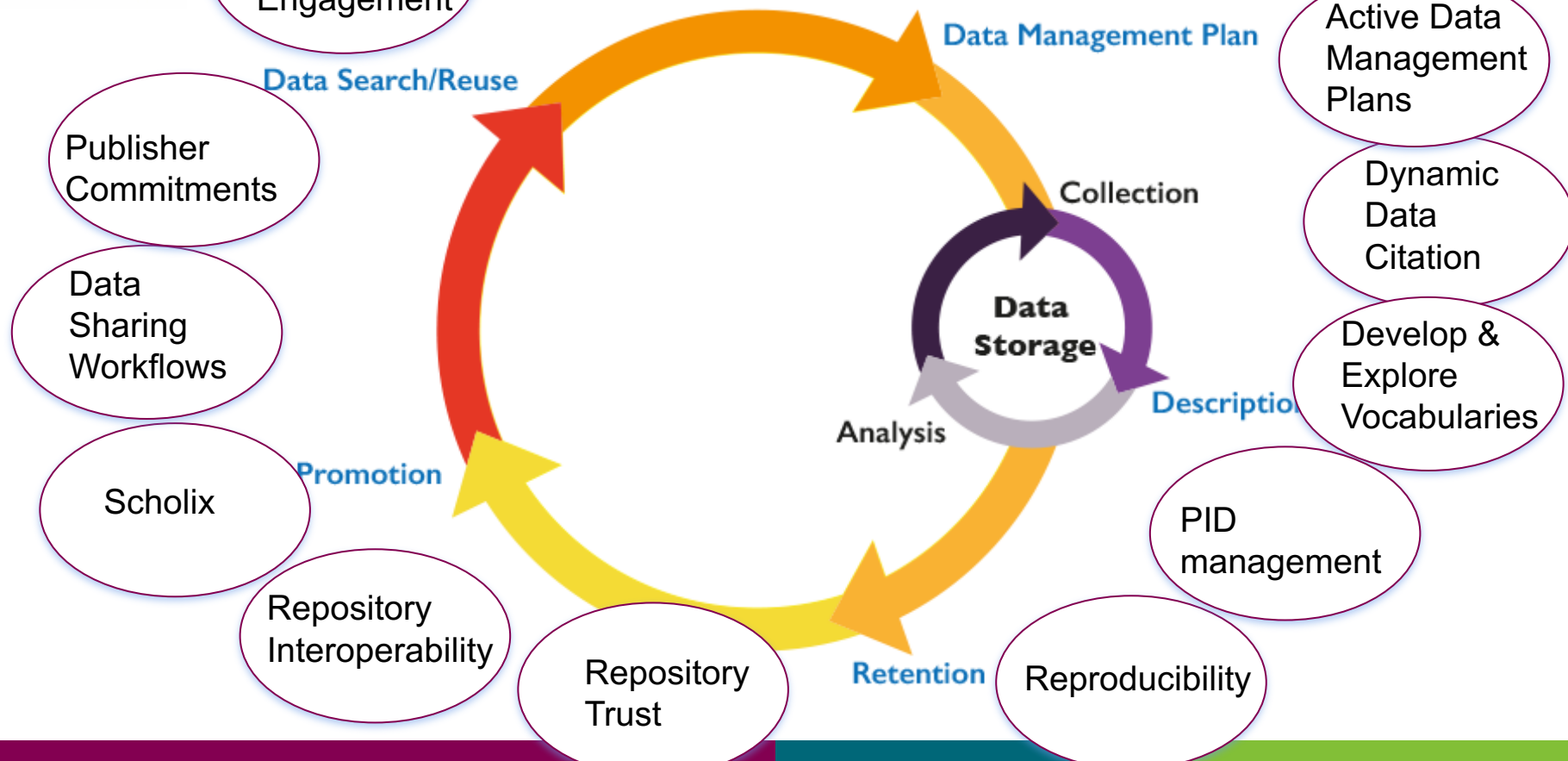
RDA Endorsed Recommendations

- Scalable Dynamic-data Citation Methodology

- Data Description Registry Interoperability Model

- Basic Vocabulary of Foundational Terminology Query Tool

- Data Type Model and Registry

- Persistent Identifier Type Registry

- Machine Actionable Policy Templates

- Repository Audit and Certification Catalogues

- Recommendation on Research Data Collections

- Workflows for Research Data Publishing: Models and Key Components

- Wheat Data Interoperability Guidelines, Ontologies and User Cases

Recommendations with endorsement in process

- The FAIRsharing Registry and Recommendations: Interlinking Standards, Databases and Data Policies

- Sustainable Business Models for Brokering Middleware to support Research Interoperability

- Metadata Standards Directory

- Recommendation on PID Kernel Information

- Framework for Summer Schools in Data Science and Cloud Computing

- RDA/TDWG Attribution Metadata Working Group: Final Recommendations

- An open, universal literature-data cross-linking service

- Research Data Repository Interoperability WG Final Recommendations

RDA Supporting Outputs

- 23 Things: Libraries For Research Data

- A survey of current practices in data search services

- Addressing the Gaps: Recommendations for Supporting the Long Tail of Research Data

- Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories

- Eleven Quick Tips for Finding Research Data

- Federated Identity Management for Research Collaborations

- Income Streams for Data Repositories

- Legal Interoperability of Research Data: Principles and Implementation Guidelines

- Matrix of use cases and functional requirements for research data repository platforms

- Persistent identifiers: Consolidated assertions

- Research Data Repository Interoperability Primer

- Summary of Virtual Layer Recommendations

Federated Identity Management for Research Collaboration

- White paper published July 2018
  - doi: 10.5281/zenodo.1296031

- Second version in the works

- Recommendations
  - Governance and Coordination
    - Increase research representation in FIM governance
    - Sustain operation of critical FIM services
    - Provide avenues for ongoingcoordination
  - Baseline of Research User Experience
    - Release Research & Scholarship attributes
    - Provide usability essentials
    - Remove interoperability barriers in eduGAIN metadata processes
    - Admit research organisations to federation
    - Enable researcher mobility
  - Security Incident Response Readiness
  - Harmonisation of Research Community Proxy Operations and Practices
    - Follow the proxy model and related AARC guidelines
    - Re-use shared AAI and related services
  - Sensitive Research User Experience

CSC

RDA Work Related to Sensitive Data

- Reproducible Health Data Services WG

- Blockchain Applications in Health WG

- Health Data Interest Group

- ELIXIR Bridging Force IG

- Digital Practices in History and Ethnography IG

- Data for Development IG

- International Indigenous Data Sovereignty IG

- Social Sciences & Humanities Research Data IG

- RDA/NISO Privacy Implications of Research Data Sets IG

- Ethics and Social Aspects of Data IG

CSC

Blockchain Applications in Health WG

- Endorsed and active

- Aim is
  - to analyse and compare usages of the blockchain in healthcare, implementations of blockchain architectures, associated legal and socio-economic impacts and perspectives
  - to assess the potential of blockchain-based self-enacting smart contracts in handling consent and data permission systems minimizing transaction costs
  - to assess if and how the blockchain can ensure compliance with advanced data protection requirements (such as GDPR), yet making it happen seamlessly and efficiently, at scale

- Final deliverable: set of guidelines for establishing a scalable blockchain-based data sharing system in healthcare
  - Guidelines will include a state-of-the-art report and a report on regulatory and legal issues, focussing on blockchain applications in health.

- Chairs: Edwin Morley-Fletcher, President of Lynkeus, David Manset, Almerys, Gnubila, & Cerebro, and Aggelos Kiayias, Uni. Edinburgh

Reproducible Health Data Services WG

- Not yet endorsed

- Goal: to improve the reuse of health data by providing recommendations for reproducible data curation and brokerage workflow services.

- Will generate
  - (i) recommendation statements to identify, capture and store curation metadata, and
  - (ii) adoption and training guide to improve the uptakes of our outputs.
  - "Documentation of workflow best practices will be shared as a data dictionary of materials [which] will be developed into templates within the CEDAR metadata registry tool, which will provide an interface for data entry, storage, and export, as well as a display of the existing metadata standards and ontologies mapped to each element within the Health Data Service Workflow. [Templates] will be shared with all CEDAR users, as well as exported as JSON and RDF schema. […] templates will be hosted and shared on a project Github, Open Science Framework, and shared Google drive. An adoption guide will be created to assist adopters in the use of the metadata collection templates, as well as best practices […]. This adoption guide will also be made available within a project Github, Open Science Framework, and shared Google drive, and potentially disseminated in the form of a publication."

- Chairs: Oya Bean, Aachen Uni., Anthony Juehne, RDA Secretariat / Washington Uni., and Ludovika Durst, Lynkeus

What social and technical bridges, that could be solved by community standards and solutions, are in the way of responsible sharing of sensitive data?

("It can't be shared", is the wrong answer.)

HEIDI LAINE

**Coordinator
Data Management and
Interoperability
Research Data Alliance Node Finland**

**CSC - IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P. O. Box 405,
FI-02101 Espoo, Finland
Tel. +358 40 513 95 93
heidi.laine@csc.fi
www.csc.fi**

facebook.com/CSCfi

twitter.com/CSCfi

youtube.com/CSCfi

linkedin.com/company/csc---it-center-for-science

github.com/CSCfi