

Advancing Rights Management Metadata Best Practices across Open and Closed Data Sharing Communities



Sam Grabus

RDA Data Share Fellow; Research Assistant, Metadata Research Center;
College of Computing and Informatics, Drexel University



Statement of Need:

A common data sharing obstacle across many disciplines is that agencies and organizations don't share their data because of the risks associated with the mishandling of sensitive or private information. Robust rights management metadata can help to ensure that researcher data sharing needs are communicated throughout the entire lifecycle of the dataset.

Goal:

Facilitate data sharing through the recommendation of rights management metadata best practices for obtaining Institutional Review Board (IRB) approval.

Literature Review on IRBs and Data Sharing

Review of the existing literature regarding data sharing and IRBs.

- Results of a survey from 60 institutions about the handling of biomedical research data determined that ambiguous or inconsistent IRB policies and practices regarding data protection makes data sharing challenging.
- Recommendations: calls for interoperable policies regarding the secondary use of data.** ¹

- Surveys of 208 IRB professionals about sharing genetic data found inconsistent beliefs about the risks and repercussions involved with the re-identification of research participants. Discusses the need for understanding of researcher data sharing needs.
- Recommendations: calls for regulatory measures or best practices to maintain consistency and avoid duplication of effort. Calls for regional/national IRB consortia to establish best practices.** ²

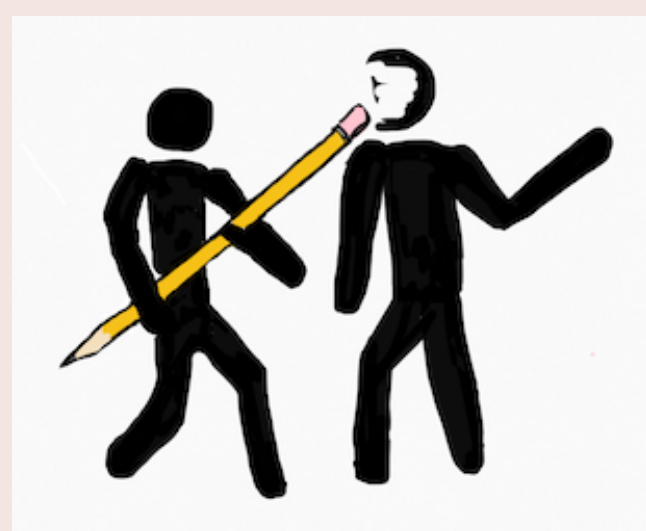
De Wolf, et al., discuss the many risks involved with sharing even de-identified or anonymized data.

Recommendations: suggests a three-prong approach for IRBs to ensure that they are taking appropriate precautions: calls for IRB understanding of data sharing risks, consulting experts (e.g., statisticians), and IRB implementation of a data protection checklist, to ensure that researchers are taking appropriate precautions with their data to avoid re-identification. ³

Related progress through the Northeast Big Data Innovation Hub's Data Sharing Spoke Project

Steps taken to Identify researcher data sharing needs:

- Collected a sample of 26 organizational data sharing agreements from industry, academia, and government
- Performed content analysis and language clustering to determined common attributes discovered in the agreements
- Identified 90+ specific attributes within the sample of 26 agreements, specifying researcher data sharing needs (e.g., Individual identifiers removed prior to transfer, encryption, establishment of data ownership).



- Sought insights from the data analysis to determine most common data sharing needs.
- Over half (16 out of the 26) of all agreements specify which rights and obligations will remain in effect after agreement termination—pointing to the need to convey comprehensive rights management metadata throughout the entire lifecycle of the dataset to ensure communication of metadata such as copyright information, provenance, ownership, and modification date.** ⁴

Identifying existing metadata standards that support rights management of data access, use, and reuse

- Currently examining existing descriptive metadata standards that incorporate rights metadata elements, such as *Latest Date*, *Copyright statement*, *Provenance description*, *Legal status*, and *Owner/Agent*, and have initially identified CDWA, VRA, METS, MARC/AACR, DACS, and EAD.
- Becoming familiar with the host of metadata standards that support rights management and licensing. Initial work relating to the NE Big Data Hub project has identified several standards, schemas, and models to guide work in this area, as demonstrated in **figure 1** below.
- Examining other licensing initiatives, such as Creative Commons and rightsstatements.org

REQUIREMENTS	EXAMPLE METADATA STANDARDS
DATA PUBLICATION; DOMAIN DISCOVERY	
Persistent Identifiers	Product (Schema.org), DOI (Digital Object Identifiers), Handle system, OAIS (Open Archival Information System)
Domain specific schemes	Schema.org, RDA metadata directory or other resources
IDENTIFICATION/DESCRIPTION	
Personal Identifiable Information	Person (Schema.org) vCard (Virtual Business Card), VIAF (Virtual International Authority File), ORCID (Open Researcher and Contributor ID)
Organization profile	Organization (Schema.org), ORCID, NAF (Name Authority File), EAC (Encoded Archival Context) for Organizational Bodies
Attribution	Same as PII
LICENSING AND USE	
Access	MODS, The Recommended Practice Access and License Indicators (NISO RP-22-2015)
Restriction on Use	Embargos and Leases (Project HYDRA), PCDM (Portland Common Data Model: Rights Extension), METS, PREMIS (Preservation Metadata Data Dictionary)
Training/user requirements	Technical metadata, operational (see 'Technical Format' and 'Restriction on Use')
Technical format	Accessibility (Schema.org), W3C MS Global Access for All (AIA) Information Model Data Element Specification, PREMIS
Privacy	EHR (Electronic Health Records)
LIFE-CYCLE MANAGEMENT	
Workflow	Protocols found via scientific research, such as Taverna and Kepler will aid this work.
Provenance	PROV-Model (Provenance Model, W3C), PREMIS
Accountability/Authenticity	PREMIS

Figure 1. From The NSF spoke proposal, "A Licensing Model and Ecosystem for Data Sharing."

Next Steps:

- IRB Approval
- Semi-structured interviews with IRB officers
- Perform crosswalk analysis of existing rights management metadata standards
- Propose rights metadata best practices for the IRB process

Resources:

- This research is supported by the RDA Data Share Fellowship and funded, in part, by National Science Foundation Award #1636788.
- Acknowledgements to the RDA/CODATA Legal Interoperability Interest Group, the RDA/NISO Privacy Implications of Research Data Sets Interest Group, Jane Greenberg, Paul Ulhir, Bob Chen, Simon Hodson, Todd Carpenter, and others, for their support.

- Goldenberg, A. J., Maschke, K. J., Joffe, S., Botkin, J. R., Rothwell, E., Murray, T. H., . . . Rivera, S. M. (2015). IRB practices and policies regarding the secondary research use of biospecimens. *BMC Medical Ethics*, 16(1), 32. doi:10.1186/s12910-015-0020-1
- Lemke, A. A., Smith, M. E., Wolf, W. A., Trinidad, S. B., & GRRIP Consortium. (2011). Broad data sharing in genetic research: Views of institutional review board professionals. *IRB: Ethics & Human Research*, 33(3), 1-5.
- Virginia A. de Wolf, Sieber, J. E., Steel, P. M., & Zarate, A. O. (2006). Part III: Meeting the challenge when data sharing is required. *IRB: Ethics & Human Research*, 28(2), 10-15.
- Grabus, S., Greenberg, J. (2017, forthcoming). Toward a metadata framework for sharing sensitive and closed data: An analysis of data sharing agreement attributes. *In MTSR-2017: Proceedings of the 11th Metadata and Semantics Research Conference*. Tallinn, Estonia, November 28-December 1st, 2017.