**Joint NISO-RDA Interest Group on Privacy Implications of Research Data Sets**

**Introduction**

Joint NISO-RDA Interest Group on Privacy Implications of Research Data Sets will explore issues related to scientific research data sets that contain human subject information, as well as related datasets that have the potential to be combined in a way that can expose private information. The benefits of this interest group include:

1) Improving the understanding of the privacy issues that relate specifically to research data from distinct stakeholder perspectives.

2) Support of a worldwide dialogue about the privacy issues that surround the sharing, combination, and reuse of research data.

3) Reduction of the risk of an unintentional release of personally identifiable information through the sharing or reuse of research data.

4) The creation and adoption of a framework to reduce the potential risk to scientific discovery writ large that might be caused by the unintentional but significant exposure of personal data.

**User Scenarios/Use Cases**

The information that could come out of the discussions, task groups, and projects will enhance the privacy of people worldwide whose personal data become the subject of research, as well as offer guidelines for those involved with the collection, preservation, sharing, use, and re-use of that data. The latter group of potential users who could benefit from this interest group is very broad, as the number of fields that are using—or could potentially benefit from applying—human subject research data is tremendous. Medicine and psychology are obvious examples, but data science is also being integrated into other fields such as the humanities and social sciences. Some disciplines have developed ethics for researchers in these situations and human-subjects review protocols ensure proper treatment of research subjects during studies, but no generalized guidelines yet exist for these same privacy issues in the deposit, preservation, and re-use of such datasets. Institutional Review Boards (IRBs) vet a variety of research processes, including data management and reuse polices primarily within the United States, but even there a 2014 report by the National Academies[[1]](#footnote-1) recommended a number of changes to the IRB system—changes that could be informed by the recommendations produced by this group. Other parties that might benefit from this effort are research funding bodies, governments, and academic data repositories.

**Objectives**

Through conversations, research, and work in task groups, the Interest Group will compile a bibliography of resources and build awareness of the privacy implications of research-data sharing. We also hope to develop a framework for how researchers and repositories should appropriately manage human-subject datasets and to develop a metadata set to describe the privacy-related aspects of research datasets. While privacy is related to the ethical, legal, and data-publishing issues surrounding data management of which privacy is a part, this interest group is focused specifically on privacy-related concerns and will support, where appropriate, the related work of other RDA groups.

**Participation**
Joint NISO-RDA Interest Group on Privacy Implications of Research Data Sets is open to all RDA members to participate. The following participants are especially relevant:

* Legal, Information Ethics, and Privacy scholars
* Library professionals involved in research data management and curation
* Research funders
* Information policy professionals
* Managers involved in any combination of the activities mentioned above.
* Repository managers

NISO and RDA are both involved in related work. Related efforts are also being undertaken by outside organizations, and this interest group will in some cases include individuals involved in those endeavors. In other cases, the results of the outside work being undertaken will be studied and, where applicable, applied to this project.

The following are some related projects.

NISO has completed a project funded by the Mellon Foundation related to privacy of patron data in library, publisher, and software-provider systems[[2]](#footnote-2). This effort created a high-level set of principles that will provide the scholarly communications community with a benchmark to relate to these issues. The principles were distributed and discussed in late 2015. While that project was explicitly focused on the U.S. market and not focused on data, but rather on publisher and library end-user services, it is related to, and will inform, this work.

The Research Data Alliance has a number of groups that are exploring related issues as well. An interest group within the RDA is focused on Legal Interoperability for data sets[[3]](#footnote-3). This group has been developing a core set of principles and guidelines that include best practices through which legal interoperability can be achieved. For human subject data, a core component of legal interoperability deals directly with privacy issues.

A new RDA working group has formed that will explore security and trust as it relates to research data[[4]](#footnote-4). The group will be focused primarily on the technological aspects of security and trust building necessary for security of potentially injurious data, if released. Certainly, security is a component of protection of privacy and there are many examples of efforts to securely share information, although a significant portion of privacy-related issues are policy focused, not necessarily technology focused.

Yet another intersecting group within RDA is centered on the topic of Ethics and Social Aspects of Data (ESAD)[[5]](#footnote-5). This group is studying a broad set of issues surrounding data sharing and the culture of scientists. It is creating an annotated bibliography and plans to pursue two additional deliverables, producing educational materials and case studies of ethical dilemmas faced by researchers working with data. Privacy is among many concerns that the group is focused on, although many ethical issues extend well beyond privacy. Conversations between ESAD and this group have already begun and the efforts should be complimentary.

Several additional groups have similar connections to privacy. As this project develops, liaisons and points of contact with other groups will be explored and fostered. One such effort is the work by the Data & Society Institute and its Council for Big Data, Ethics, and Society (BDES). Data & Society is “a research institute in New York City that is focused on social, cultural, and ethical issues arising from data-centric technological development.” In 2015, the project produced a survey report entitled Human-Subjects Protections and Big Data: Open Questions and Changing Landscapes[[6]](#footnote-6), which outlines some of the challenges related to scholarly data resources and privacy. In September of that year, BDES announced a new network to “facilitate information sharing, discussion, and community building among academics, practitioners, researchers, and others who seek to raise important questions, share opportunities, and ask for help navigating complex data ethics issues.”

There is a significant project led by the Harvard School of Engineering and Applied Sciences entitled Privacy Tools for Sharing Research Data[[7]](#footnote-7). This effort is part of a larger National Science Foundation Secure and Trustworthy Cyberspace Project[[8]](#footnote-8) that has received additional support from the Sloan Foundation and Google, Inc. That project’s goals are “to help enable the collection, analysis, and sharing of sensitive data while providing privacy for individual subjects.” A good deal of its work has been around tools to support differential privacy risk assessments as a framework for decision making about the risks and controls necessary to support privacy. The group has developed open course materials, hosted seminars, and produced a variety of papers and presentations. It has also organized a public symposium hosted by the Harvard Institute for Applied Computational Science Privacy in a Networked World, held Friday, January 23, 2015. The symposium included speakers Edward Snowden, Bruce Schneier, John DeLong, John Wilbanks, Lee Rainie, and Cynthia Dwork. This initiative is primarily, though not exclusively, focused on the technological and computational elements of privacy and data protection. NISO has worked closely with several members of this team and plans to include them in this interest group.

**Outcomes**

The interest group will work to achieve the following specific outcomes:

1. Lead discussions at RDA Plenary meetings on research data privacy topics.
2. Gather and share a bibliography of data-and-privacy-related materials for public use.

The interest group will host task groups to work on the following outcomes and that may become working groups:

1. Development of a framework that explains, at a high level, the precautions that data creators, repositories, aggregators and scientists should use in creating, using, preserving, and providing access to research data. That task group’s [draft aims and outputs](https://docs.google.com/document/d/1KVMa-vj9O0wzX2rQAh4L0-SENEu5wJEAwZK4mHc36hw/edit#heading=h.m4hwwzzhrcq6) are being drafted online
2. An exploration of the policy implications at an institutional level and provide guidance that will support sharing of personal and/or sensitive data. This [task group’s aims and objectives](https://docs.google.com/document/d/1rW7siwKpSBBCfxCvXAZRI9NP28Cd5iXdxmLSY0v6gKY/edit#heading=h.t2hgo9phlylu) are being drafted online.
3. Identification of key areas of variance in privacy laws or regulations at national and international levels that are significant when sharing data worldwide. The aims and scope of this task force are still being drafted.
4. An outline of use cases and situations where the privacy principles would be applied.
5. Definition of a set of technical metadata that can be used to describe privacy-related information contained within a data set, parameters for use, and description of where it should be applied. This output was determined to be a later-stage activity and would be undertaken as a joint effort with the Metadata IG.
6. Advancement of adoption of the principles through an outreach and communications campaign.

**Mechanisms and Coordination**

The group will meet face-to-face during the RDA Plenaries to build interest and awareness, lead discussions, and to share its work with other working and interest groups regarding privacy and data sharing.

The task groups will meet virtually as needed during the interim periods between plenaries.

Meetings and group communications will be coordinated by NISO staff and the other co-chairs.

NISO hosted a public symposium on research data and privacy, in coordination with the RDA P8 in Denver, CO in the fall of 2016. Recordings of that meeting are available on the [Symposium website](http://www.niso.org/news/events/2016/september_joint/).

All documents related to this project will be publicly available on both the RDA website and mirrored on the NISO website.

**Timeline & Work Plan**

The group will explore world-wide legal frameworks and the impacts these frameworks have on data sharing, especially with human-subject data. The group has created task groups which may lead to working groups. They will consider crafting a set of principles that will provide guidance to the researcher and repository communities on how to manage these data when they are received. crafting a set of use cases on how the principles will be applied. After these elements are completed, an effort to advance the principles through promotion and community outreach will be developed and executed.

The proposed interested group met at P7 in Tokyo to discuss the value of the group and needs of the RDA community in this topic area. Since that meeting, the members of the group have met and created task groups to focus in on specific areas. The members have also started a bibliography via Zotero. The members meet again face to face at P8 in Denver and began discussing this task groups and their outputs as well as ways to work with other RDA groups.

The interest group also co-sponsoring a joint meeting on Responsible Sharing on Confidential Research Data which will include experts speaking on emerging challenges of new sources and forms of confidential data and legal and ethical responsibilities of data repositories. It’s a joint meeting of IG Domain Repositories, IG RDA/NISO Privacy Implications of Research Data Sets and IG Ethics and Social Aspects of Data.

Following the Denver plenary meeting, the task groups will make progress and determine if, at some point, they want to turn into working groups, although the scope of their ouputs as currently defined are more suited for IG outputs. We plan to meet face to face at RDA Plenaries and continue the work regularly between plenaries through our task groups.

**Potential Group Members**

We have a list of over 55 members on our [case statement page](https://rd-alliance.org/group/rdaniso-privacy-implications-research-data-sets-wg/case-statement/rdaniso-privacy-implications). We have representatives from the US, UK, Poland, Denmark, Greece, Australia, Sweden, Germany, Austria, Italy, and Canada. We will work to increase membership from more regions of the world.

**Proposed Chairs of the Group**

Todd Carpenter, Executive Director, National Information Standards Organization (NISO)

Bonnie Tijerina, Research Fellow, Data & Society

Libby Bishop, Manager, Producer Relations – Research Ethics, UK Data Service, University of Essex (UK)

Freyja van den Boom, Researcher FutureTDM, Open Knowledge International

1. National Research Council. (2014). *Proposed Revisions to the Common Rule for the Protection of Human Subjects in the Behavioral and Social Sciences.* Committee on Revisions to the Common Rule for the Protection of Human Subjects in Research in the Behavioral and Social Sciences. Board on Behavioral, Cognitive, and Sensory Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. [↑](#footnote-ref-1)
2. ##  NISO. (2015). Consensus Framework to Support Patron Privacy in Digital Library and Information Systems. Retrieved at http://www.niso.org/topics/tl/patron\_privacy/.

 [↑](#footnote-ref-2)
3. #  RDA. (2016). RDA/CODATA Legal Interoperability IG. Retrieved at

 https://rd-alliance.org/groups/rdacodata-legal-interoperability-ig.html [↑](#footnote-ref-3)
4. https://rd-alliance.org/groups/working-group-data-security-and-trust.html [↑](#footnote-ref-4)
5. https://rd-alliance.org/group/ethics-and-social-aspects-data/case-statement/rda-%E2%80%93-ethics-and-social-aspects-data-esad [↑](#footnote-ref-5)
6. #  Metcalf, Jacob. (2015, April 22). Human-Subjects Protections and Big Data: Open Questions and Changing Landscapes. Retrieved at http://bdes.datasociety.net/council-output/human-subjects-protections-and-big-data-open-questions-and-changing-landscapes/

 [↑](#footnote-ref-6)
7. Harvard School of Engineering and Applied Sciences. (2014). Privacy Tools for Sharing Research Data, A National Science Foundation Secure and Trustworthy Cyberspace Project. Retrieved at http://privacytools.seas.harvard.edu/ [↑](#footnote-ref-7)
8. https://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504709 [↑](#footnote-ref-8)