



Blane Dahl
National Science Board Office
Committee on Strategy and Budget
Task Force on Data Policies
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

The National Science Board's Task Force on Data Policies comes at a watershed moment in the development of an infrastructure for data-intensive science based on sharing and interoperability. The NISO community applauds this effort and the focused attention on the key issues related to a robust and interoperable data environment.

Thank you for the opportunity to comment on the NSB's *Digital Research Data Sharing and Management* report. The National Information Standards Organization (NISO) respectfully submits the following comments.

NISO is an ANSI-accredited non-profit industry trade organization that supports the development of technical standards for the creation, distribution, discovery and preservation of information. Since 1939, NISO has worked within the publishing, library, media and system-provider communities to develop consensus standards and best practices to improve interoperability of information distribution, particularly in research, education and libraries. As such, we play an important role in the scholarly landscape and we have been working on a variety of data-related issues over the past several years.

NISO has standardized several key data-sharing related standards. Among the projects most tied to research data are the Digital Object Identifier (DOI) syntax (ANSI/NISO Z39.84) and the Dublin Core metadata (ANSI/NISO Z39.85) standards. Identifiers and metadata for information distribution have long been staples in the NISO standards portfolio. NISO's historical engagement in these areas goes back decades, and includes some of the most widely known and used identifiers in the trade of information. NISO was responsible for key components in the formulation of the precursor to the ISSN (International Standard Serial Number) identification system, the ISBN (International Standard Book Number) identification system and the MARC (MACHine-Readable Cataloging) descriptive cataloging and bibliographic information exchange standards.

NISO has particular interest in *Key Challenge #4: The reproducibility of scientific findings requires that digital research data be searchable and accessible through documented protocols or method*. Beyond its historical involvement in these issues, NISO is actively engaged in forward-looking projects related to data sharing and data citation. NISO, in partnership with the National Federation of Advanced Information Services (NFAIS), is nearing completion of a best practice for how publishers should manage supplemental materials that are associated with the journal articles they publish. With a funding award from the Alfred P. Sloan Foundation and in partnership with the Open Archives Initiative, NISO began work on ResourceSync, a web protocol to ensure large-scale data repositories can be replicated and

maintained in real-time. We've also had conversations with the DataCite group for formal standardization of their IsCitedBy specification. I serve as a member of the ICSTI/CODATA task force working on best practices for data citation and NISO is looking forward to promoting and formalizing any recommendations and best practices that derive from that work. NISO is also presently in conversations about a potential project, hopefully with the support of the Sloan Foundation and NSF, on the question of data equivalence, i.e., how one might ensure researchers have the same dataset after it goes through transformations related to packaging (say from raw data into SASS, SPSS, Excel or other data management programs.) These projects have the potential to add significantly to the data exchange environment and we hope additional work related to data exchange will be forthcoming in the coming months. We also hope that the NSB and NSF will encourage the use of these recommendations and standards.

We strongly urge that any further development of data-related best practices and standards take place in neutral forums that engage all relevant stakeholder communities, such as the one that NISO provides for consensus development. As noted in Appendix F of the report, *Summary Notes on Expert Panel Discussion on Data Policies*, standards for descriptive and structural metadata and persistent identifiers for all people and entities in the data exchange process are critical components of an interoperable data environment. We cannot agree more with this statement from the report of the meeting: “Funding agencies should work with stakeholders and research communities to support the establishment of standards that enable sharing and interoperability internationally.”

There is great potential for NSF to expand its leadership role in fostering well-managed use of data. This would include not only support of the repository community, but also in the promulgation of community standards. In partnership with NISO and using the consensus development process, NSF could support the creation of new standards and best practices. More importantly, NSF could, through its funding role, provide advocacy for—even require—how researchers should use these broad community standards and best practices in the dissemination of their research. We note that there are more than a dozen references to standards in *Digital Research Data Sharing and Management* report, so we are sure that this point is not falling on unreceptive ears.

The engagement of all relevant stakeholders in the establishment of data sharing and management practices as described in *Recommendation #1* is critical in today's environment—at both the national and international levels. While the promotion of individual communities of practice is a laudable one, it does present problems and issues when it comes to systems interoperability. A robust system of data exchange by default must be one grounded on a core set of interoperable data. More often than not, computational systems will need to act with a minimum of human intervention to be truly successful. This approach will not require a single schema or metadata system for all data, which is of course impossible and unworkable. However, a focus on and inclusion of core data elements and common base-level data standards is critical. For example, geo-location, bibliographic information, identifiers and discoverability data are all things that could be easily standardized and concentrated on to foster interoperability. Domain-specific information can be layered over this base of common and consistent data in a way that maintains domain specificity without sacrificing interoperability.

One of the key problems that the NSB and the NSF should work to avoid is the proliferation of standards for the exchange of information. This is often the butt of standards jokes, but in reality it does create significant problems. It is commonplace for communities of interest to review the landscape of existing standards and determine that existing standards do not

meet their exact needs. That community then proceeds to duplicate seventy to eighty percent of existing work to create a specification that is custom-tailored to their specific needs, but which is not necessarily compatible with existing standards. In this way, standards proliferate and complicate interoperability. The NSB is uniquely positioned to help avoid this unnecessary and complicating tendency. Through its funding role, the NSB should promote the application, use and, if necessary, extension of existing standards. It should aggressively work to avoid the creation of new standards, when relevant standards already exist.

The sharing of data on a massive scale is a relatively new activity and we should be cautious in declaring fixed standards at this state. It is conceivable that standards may not exist to address some of the issues in data sharing or that it may be too early in the lifecycle for standards to be promulgated in the community. In that case, lower-level consensus forms, such as consensus-developed best practices or white papers could advance the state of the art without inhibiting the advancement of new services, activities or trends. The NSB should promote these forms of activity as well, when standards development is not yet an appropriate path.

Thank you once again for this opportunity to respond to the Request for Information: *Digital Research Data Sharing and Management*. The National Information Standards Organization looks forward to the formal issuance of these recommendations and its attendant policies. We also look forward to the opportunity to assist in any way in promoting the increased use and long-term sustainability of a robust data exchange environment for research and learning. If there are any inquiries related to this response, please feel free to contact me at your convenience.

Respectfully submitted,



Todd A. Carpenter
Managing Director

National Information Standards Organization (NISO)
One North Charles Street
Suite 1905
Baltimore, MD 21212
(301) 654-2512
tcarpenter@niso.org