RESEARCH PAPER

# The Landscape of Rights and Licensing Initiatives for Data Sharing

Sam Grabus and Jane Greenberg

Metadata Research Center (MRC), College of Computing and Informatics (CCI), Drexel University, Philadelphia, PA, US
Corresponding author: Sam Grabus (sam.grabus@drexel.edu)

Over the last twenty years, a wide variety of resources have been developed to address the rights and licensing problems inherent with contemporary data sharing practices. The landscape of developments is this area is increasingly confusing and difficult to navigate, due to the complexity of intellectual property and ethics issues associated with sharing sensitive data. This paper seeks to address this challenge, examining the landscape and presenting a Version 1.0 directory of resources. A multi-method study was pursued, with an environmental scan examining 20 resources, resulting in three high-level categories: *standards*, *tools*, and *community initiatives*; and a content analysis revealing the subcategories of *rights*, *licensing*, *metadata & ontologies*. A timeline confirms a shift in licensing standardization priorities from open data to more nuanced and technologically robust solutions, over time, to accommodate for more sensitive data types. This paper reports on the research undertaking, and comments on the potential for using license-specific metadata supplements and developing data-centric rights and licensing ontologies.

## 1. Introduction

Today's data sharing movement continues to be encumbered by the need to protect sensitive and proprietary information, which can make the data sharing process prohibitively difficult. For some researchers, the advantages of data sharing can be outweighed by the risks associated with sharing personally-identifiable information (PII), intellectual property, and other sensitive data types (Fecher, Friesike, & Hebing, 2015). Fortunately, a number of resources have been pursued over the last twenty years, addressing rights and licensing challenges.

As the data sharing movement grows across all sectors, navigating the landscape of rights and licensing resources has become increasingly complicated given the diversity of the resources addressing these challenges. Where is the best place for a researcher or an organization to learn about facilitating the complex process of rights management? Which standardized licenses would be most appropriate for sharing a particular type of data, and which metadata standards and ontologies can help address these needs? The landscape can be complicated for researchers to navigate, due to the varying scopes and impact of the initiatives, as well as the international nature of data sharing and its challenges. This current environment points to a need for frameworks that can help researchers identify the resources best suited for their data sharing needs.

The research presented in this paper addresses this need. The paper reports results from an environmental scan of resources supporting data sharing through their focus on rights and licensing. The emphasis is on resources that are potentially applicable to research data. The work presented was conducted over a six-month period, from August 2017-January 2018. The work was motivated, in part, by current work on the NSF Spoke Initiative, A Licensing Model and Ecosystem for Data Sharing (Metadata Research Center, 2018) (Greenberg et al., 2017), and by research conducted as a Research Data Alliance (RDA US) data share fellow

(Grabus & Greenberg, 2018). The following section of this paper presents the background, covering information ethics and legal challenges of data sharing, followed by the research objectives and review the method supporting the environmental scan. Next, the results are presented in two sections: *first*, the standards, tools, and community initiatives covering rights and licensing are described, and *second*, a set of visualized results and initiative descriptions are presented as a framework for understanding how these rights and licensing developments have progressed and interrelate. The results are followed by a directory (Version 1.0) of basic initiative information, a contextual discussion of the environmental scan, and the conclusion that highlights key findings and identifies future initiative direction.

## 2. Background
### 2.1 Ethics in Data Sharing
Sharing research data, while crucial to the development of solutions and innovations, is encumbered with many ethical issues. Data sharing and information ethics are unavoidably interconnected in the contemporary global information society, spanning privacy, accuracy, property, and accessibility of information (PAPA), also known as focal points for developing a social contract to protect "threats to their intellectual capital" (Parrish, 2010, p. 187). Privacy, in particular, has gained much attention in the public eye over the last several years, particularly with high profile incidents, such as the Cambridge Analytica Facebook data breach (Granville, 2018). In essence, information privacy relates to our ability to control the flow of information about ourselves (Bélanger & Crossler, 2011). These privacy restrictions may complicate researcher and corporate endeavors to maintain a competitive edge and promote innovation through information insights.

Concerns about information privacy frequently prohibit the sharing of data between researchers. Researchers are concerned with losing control or even knowledge over who has access to the data, as well as how the data is accessed and ultimately used (Fecher, Friesike, & Hebing, 2015). The major factors that contribute to this apprehension are protecting personally-identifiable information categories (PII), such as the 18 Health Insurance Portability and Accountability Act (HIPAA) identifiers, intellectual property, and other sensitive data categories. These other sensitive data categories may include indigenous data (Harding et al., 2012), endangered or invasive species data (Jarnevich, Graham, Newman, Crall, & Stohlgren, 2007), same-disease data (Liu et al., 2016), and quasi-identifiers, such as gender, date of birth, and zip code, which, when combined, can uniquely identify between 63 and 87% of the US population (Liu et al., 2016).

### 2.2 Legal Issues in Data Sharing
There are many legal liability data sharing barriers that operate in conjunction with the challenges of complying with privacy concerns. Complex data sharing agreements are frequently required in order to ensure that appropriate measures are taken to protect the privacy of PII, intellectual property, and other sensitive data types. Particularly with biomedical data, institutional policies require data sharing agreements that prohibitively complicate the data sharing process (Tenopir, 2011). Contractual agreements between organizations typically specify permissions and restraints for how the data can be handled. These specifications can include clauses regarding data updates, access controls, quality guarantees, how the data can be copied and displayed, whether it can be disseminated, how the original source will be credited, and who is responsible for remedying data breaches (Swarup, Seligman, & Rosenthal, 2006). These data sharing agreements may also specify limitations for research subject re-identification, data transferability, requirements for IRB review, and use of the data solely for research purposes. Legal aspects in data sharing can become even more complicated when a singular project integrates multiple datasets held in systems with differing data security requirements (Rockhold, Nisen, & Freeman, 2016).

Considering the collective momentum towards open access, open data, and open science, it is essential to remember that protecting individual privacy, intellectual property, and national security must be balanced against this impetus (National Research Council, 1997; National Science and Technology Council Committee on Science, & National Science and Technology Council Interagency Working Group on Digital Data, 2009). Careful measures regarding rights management and data licensing can help to ensure that researchers are able to maintain the relationship of trust with research subjects that is necessary to ensure that the research will be able to continue safely well into the future. Informatics solutions must address the concerns and repercussions regarding information privacy and legal requirements, which frequently requires extensive rights management and licensing measures.

### 2.3 The Landscape of Technical and Informatics Solutions

Open data has become an international movement, particularly among STEM disciplines, although not all STEM data can be open or free. The progress has nevertheless helped to highlight ideas sharing closed data, which can be supported through reduced complexity and providing guidance for the usage of sensitive data types (Janssen, Charalabidis, & Zuiderwijk, 2012). In other words, "[data] sharing should not be an all-or-nothing choice" (Sweeney, Crosas, Bar-Sinai, 2015, p. 2), considering the many risks and challenges associated with sharing sensitive data. Moving forward, we need to develop technological and informatics solutions for sharing sensitive data to both diminish the risks and make it a less burdensome process for organizations to undergo.

This proliferation of the open data and open science movements has been an impetus for the development of an increasing variety of technological and informatics solutions for licensing and sharing data. Despite this, researcher confusion about the complex nuances of legal protection, licensing options, republishing, and data sharing prevails. (Else, 2016; Oxenham, 2016). The landscape of initiatives related to enabling to these data sharing facets is extensive, with each catering to a specific piece of the data sharing puzzle.

Some initiatives, such as the Research Data Alliance (2017c), serve to bring disciplines together to discuss and advance data sharing practices and possibilities, whereas other initiatives exist solely to develop a standard. Standards most often refer to regulatory outputs that have been formally endorsed by standard governing bodies, such as the International Organization for Standardization (ISO, 2018), World Wide Web Consortium (W3C, 2018), the European Committee for Standardization (CEN, 2018), and most recently, the Research Data Alliance has also gained traction as a global standards-creating organization in the data sharing space.

As these developments continue to grow, it is increasingly challenging for a newcomer to grasp the scope of issues such as licensing, rights management, and standards related to the data sharing process. Even those who have been engaged in addressing data sharing challenges have trouble keeping up. Currently, there is no single vetted resource for learning about the full extent of these developments and how they may address associated data sharing challenges. To this end, it seems there is a growing need for frameworks to better understand this evolving landscape. Furthermore, a directory or open list where individuals and communities can help to identify and share use information about such developments could be of tremendous value to any community or individual pursuing data sharing. The work reported on in this paper considers the complex landscape of technical and informatics data sharing solutions, and takes initial steps present as a framework and initial directory to help any community or individual seeking to navigate and learn more about sharing research data across both open and closed environments.

## 3. Objectives

The overriding goal of this work is to provide clarity by offering a framework for understanding the landscape of data sharing initiatives at the intersection of rights and licensing. A secondary goal was to present a basis for a directory of initiatives in this area, which will evolve into an online, community-driven resource. These objectives were shaped by engagement in the North East Big Data Innovation Hub (NEBDIH), as well as work taking place within the Research Data Alliance and related communities. The next section of this paper reports on our methods and the steps taken to address these objectives.

## 4. Method

The above objectives were pursued by conducting a multi-method approach combining an environmental scan and content analysis. Environmental scan methods are often pursued in marketing to understand the landscape and identify opportunities and threats, and to detect trends (Cooper & Schindler, 2012). Content analysis is a common method guiding the examination of an artifacts, such as a documents, images or collection of resources, and looking for patterns. The method used in the information and data area, draws from Krippendorff (2012). The combined approach, integrating an environmental scan and a content analysis was pursued to allow more thorough investigation of this topic.

The protocol for performing this research involved the following steps:

Environmental scan steps
1. *Data collection.* Journal publications, reports, slides, outputs of working groups or communities, and other artifacts associated with data sharing, rights, privacy, sensitive data, restricted data, licensing, and the intersection of these areas were collected. Steps were taken to be as comprehen-

sive as possible, but we also considered practical research constraints. Data collection was limited to: 1) English language, 2) materials that showed sufficient community impact through either duration of some time (e.g., a few years), or active participation through publications and other outputs. Endorsement or activity within major organizations addressing data licensing and rights management, such as the Research Data Alliance (RDA), CODATA, ESIP (Earth Science Information Partners), DPLA, and Europeana, were also considered.

2. *The first phase analysis.* This step drew upon the formal environmental scan methodology to identify trends. This step involved reading initiative documentation and establishing high-level categories to differentiate between the various types of initiatives identified. Our first-pass at high level categories were 1. Data licensing standardization, and 3. Metadata initiatives

3. *Category refinement.* After iterative review, feedback, and additional data collection, it became clear to the researchers that further refinement of these high-level categories were required. The environmental scan for the work presented in this paper yielded key types of initiatives: 1. Standards, 2. Tools, and 3. Community initiatives. Conceptualization of these high level categories were as follows:

> *Standard*: a uniform technical procedure or practice as developed through expertise-driven consensus.
> *Tool*: a technical application to help automate or otherwise streamline a procedure.
> *Community initiative*: an initiative developed by a group of people who share a concern or a passion for a rights or licensing topic within the open data community, and learn how to do it better as they interact regularly. This definition reflects the fundamental social nature of human learning.

Content analysis steps

1. *Template development.* A second-phase examination was pursued, building on the above steps, and a template was designed to methodically capture the content about the 1. Standards, 2. Tools, and 3. Community Initiatives.

2. *Categorization.* The second phase analysis also helped in identifying a set of sub-categories that was refined through an iterative process with members of the research team, and through feedback from individuals engaged in the Research Data Alliance.

## 5. Results

The results of the environmental scan and content analysis are presented below. The initial environmental scan identified 20 initiatives falling into three broad categories: standards, tools, and community initiatives. As reported in **Table 1**, we identified 11 standards, three tools, and six community initiatives. **Table 1** presents the high-level framework, showing how these 20 initiatives fall into the three broad categories.

For the content analysis, each of the initiatives were further classified by the subcategories of rights, licensing, metadata & ontologies, and informational resources. Each initiative was assigned to an average of two sub-categories. Three of the initiatives were classified with one sub-category, 13 had two categories, and three fit into three sub-categories. **Table 2** presents the results of dividing the initiatives into subcategories.

**Table 1:** Initiative Categories.

| Standards | Tools | Community Initiatives |
|---|---|---|
| · Creative Commons<br>· Open Data Commons<br>· The Open Government License<br>· RightsStatements.org<br>· Linked Content Coalition<br>· The Data Use Ontology<br>· The Neurona Data Protection Ontology<br>· W3C Permissions & Obligations Expression<br>· ONIX-PL<br>· RightsDeclarationMD Extension Schema<br>· Open Digital Rights Language | · ShareDB: A Licensing Model and Ecosystem for Data Sharing<br>· DataTags<br>· Legal Assessment Tool (LAT) | · Research Data Alliance<br>· Datasets Licensing Project<br>· DCC's How to License Research Data<br>· The (Re)usable Data Project<br>· FAIRsharing.org<br>· The Federal Demonstration Partnership: Data Transfer and Use Agreement Pilot |

**Table 2:** Initiative Subcategories.

| | Rights | Licensing | Metadata and Ontologies | Informational Resources |
|---|---|---|---|---|
| Creative Commons | ✓ | ✓ | | |
| Datasets Licensing Project | | ✓ | ✓ | |
| DataTags | ✓ | | | |
| The Data Use Ontology | ✓ | | ✓ | |
| DCC's How to License Research Data | | ✓ | ✓ | ✓ |
| FAIRsharing.org | | | ✓ | ✓ |
| The FDP Data Transfer and Use Agreement Pilot | | ✓ | | ✓ |
| Legal Assessment Tool (LAT) | | ✓ | | ✓ |
| Linked Content Coalition | ✓ | | ✓ | |
| The Neurona Data Protection Ontology | ✓ | | ✓ | |
| ONIX-PL | | ✓ | ✓ | |
| Open Data Commons | ✓ | | | |
| Open Digital Rights Language (ODRL) | ✓ | | ✓ | |
| The Open Government License | ✓ | | | |
| The (Re)Usable Data Project | | ✓ | | ✓ |
| Research Data Alliance | ✓ | ✓ | ✓ | |
| RightsDeclarationMD Extension Schema | ✓ | | ✓ | |
| RightsStatements.org | ✓ | | | |
| ShareDB: A Licensing Model & Ecosystem for Data Sharing | | ✓ | | |
| W3C Permissions & Obligations Expression | ✓ | ✓ | ✓ | |

Another output of the content analysis is a timeline of when these initiatives started (**Figure 1**), in order to identify any insights regarding the progression of initiative scope and emphasis over time.

The timeline begins with the development of Creative Commons, ODRL, and RightsDeclarationMD in 2001, and the last initiatives reported in this research are the (Re)usable Data Project, Datasets Licensing Project, The Data Use Ontology, and the FDP Data Transfer and Use Pilot, all of which started in 2017. Beginning with 2008, 16 out of the 20 initiatives (80%) started in the second half of this 16-year time span, with the remaining 4 initiatives (20%) starting between 2001 and 2007. This timeline also shows that the "open" licensing standardization efforts (Creative Commons, Open Data Commons, and The Open Government License) were developed between 2001 and 2009, while the other two licensing initiatives (ShareDB and the Datasets Licensing Project), started far more recently (in 2016 and 2017 respectively), and with substantial technological components. This may suggest a shift in prioritization due to the need for more nuanced solutions.

## 6. Directory (Version 1.0)

The initiatives explored in this environmental scan are described more extensively in this directory (**Table 3**), reporting the following details: name, sub-categories, date initiated, founded by, current URL, followed by the goals and status in the *Appendix*. The goal of providing these more significant descriptions is to provide readers with a concise glimpse of the scope and purpose for each initiative, as well as what types of data are appropriate for the various standardization efforts and technological infrastructures.

## 7. Discussion

The above data analysis presented broad categories, subcategories, a timeline, and directory (Version 1.0) of initiative efforts. The classification of these initiatives demonstrates the complexity of these various initiatives, since most initiatives address more than one need, and vary in purpose and scope. The results show that we can look at these initiatives both at a top level, in terms of being a standard, tool, or community
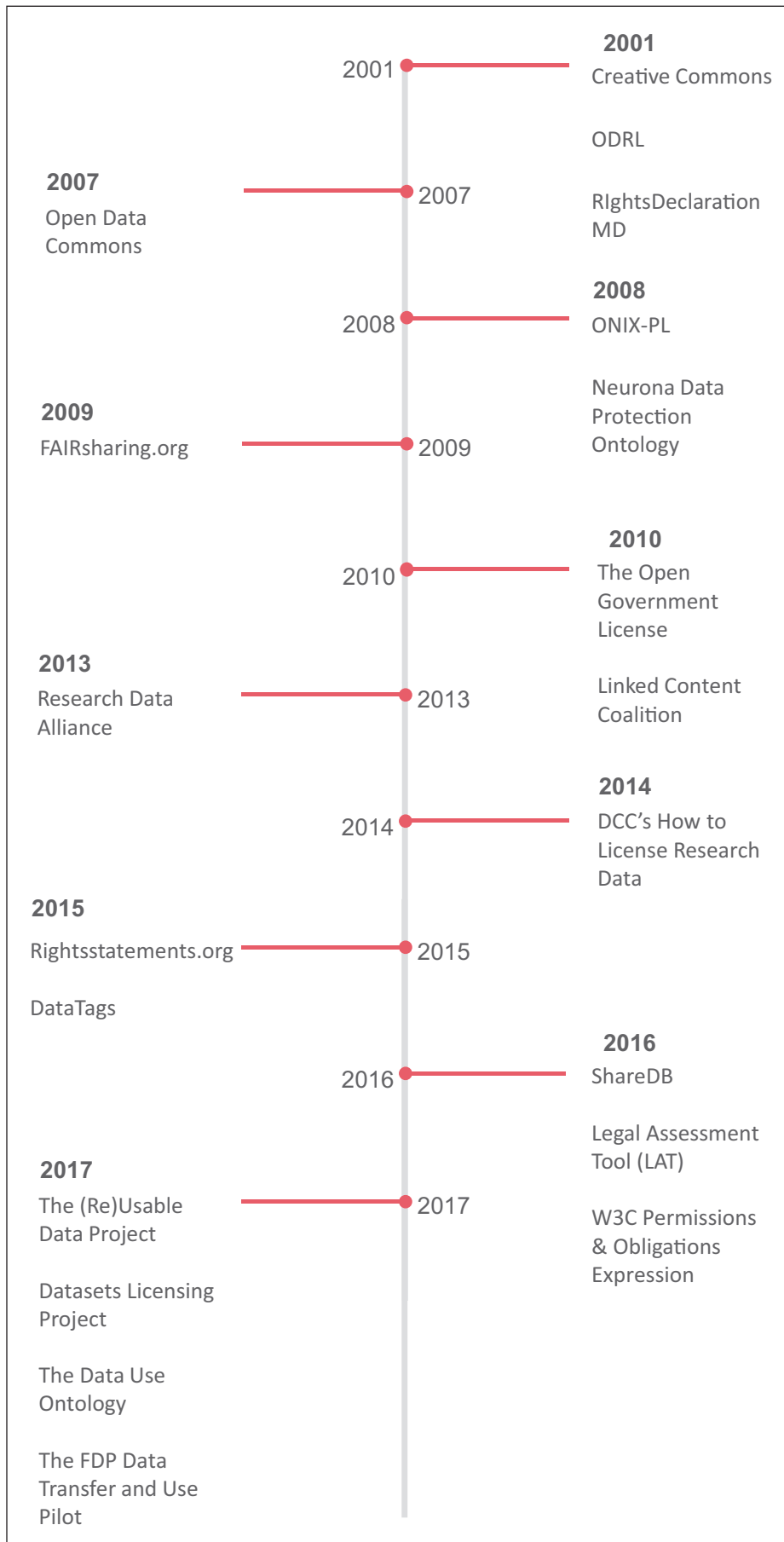
**Figure 1:** Timeline of Initiatives.

initiative, and at a more specific level, regarding the multiple ways that many of these initiatives approach the challenges of sharing data. Our top-level classification showed a heavy emphasis on the development of standards and community initiatives, with far fewer tools to facilitate the process. The classification of initiatives into subcategories provided further insights. The vast majority of these initiatives fell into two or more subcategories, demonstrating that the majority of standards, tools, and communities at the intersection of

**Table 3:** Directory (Version 1.0).

| | Initiatives | Sub-Categories | Date Initiated | Founded By | Current URL |
|---|---|---|---|---|---|
| **Standards** | **Creative Commons** | Licensing, Rights | 2001 | Lawrence Lessig | https://creativecommons.org/ |
| | **Open Data Commons** | Licensing, Rights | 2007 | Open Knowledge Foundation | https://opendatacommons.org/ |
| | **The Open Government License** | Licensing, Rights | 2010 | UK National Archives | http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ |
| | **RightsStatements.org** | Rights | 2015 | DPLA and Europeana | http://rightsstatements.org/en/ |
| | **Linked Content Coalition** | Rights, Metadata & Ontolgoies | 2010 | European Publisher's Council | http://www.linkedcontent-coalition.org/ |
| | **The Data Use Ontology** | Rights, Metadata & Ontolgoies | 2017 | Global Alliance for Genomics and Health | https://github.com/EBIS-POT/DUO |
| | **The Neuron a Ontology** | Rights, Metadata & Ontolgoies | 2008 | S21 sec security company and the Institute of Law and Technology at the Universitat Autonoma de Barcelona | N/A |
| | **W3C Permissions & Obligations Expression** | Rights, Licensing, Metadata & Ontolgoeis | 2016 | W3C | https://www.w3.org/2016/poe/wiki/Main Page |
| | **ONIX-PL** | Licensing, Metadata & Ontologies | 2008 | Digital Library Federation's Electronic Resource Management Initiative (ERMI) and EDItEUR/NISO | http://www.editeur.org/21/ONIX-PL/ |
| | **RightsDeclaration MD Extension Schema** | Rights, Metadata & Ontolgoies | 2001 | Digital Library Foundation for digital library objects | https://www.loc.gov/standards/rights/METSRights.xsd |
| | **Open Digital Rights Language (ODRL)** | Rights, Metadata & Ontologies | 2001 | W3C Permissions & Obligations Expression Working Group | https://www.w3.org/community/odrl/ |
| **Tools** | **ShareDB: A Licensing Model and Ecosystem for Data Sharing** | Licensing | 2016 | Drexel University's Metadata Research Center, MIT, Brown University | https://cci.drexel.edu/mrc/rescarch/a-licensing-model-and-ecosystem-for-data-sharing/ |
| | **DataTags** | Rights | 2015 | Harvard's Dataverse | https://datatags.org/ |
| | **Legal Assessment Tool (LAT)** | Informational Resource, Licensing | 2016 | BioMedBridges | N/A |

(Contd.)

| | Initiatives | Sub-Categories | Date Initiated | Founded By | Current URL |
|---|---|---|---|---|---|
| **Community Initiaves** | **Research Data Alliance** | Rights, Licensing, Metadata & Ontolgoeis | 2013 | European Commission, the US National Science Foundation (NSF), and the Australian Government's Department of Innovation | https://www.rd-a1liance.org/ |
| | **Datasets Licensing Project** | Licensing, Metadata & Ontologies | 2017 | Jisc, The University of Glasgow, and CREATe | httns://datasetlicencing.wordpress.com/ |
| | **DCC's How to License Research Data** | Informational Resource, Licensing, Metadata & Ontologies | 2014 | Digital Curation Centre | http://www.dcc.ac.uk/resources/how-guides/license-research-data |
| | **The (Re)Usable Data Project** | Informational Resource, Licensing | 2017 | National Center for Advancing Translational Sciences (NCATS) Biomedical Data Translator and the Monarch Initiative | http://reusabledata.org/ |
| | **FAIRsharing.org** | Informational Resource, Metadata & Ontologies | 2009 | University of Oxford e- Research Centre | https://fairsharing.org/ |
| | **The Federal Demonstration Partnership: Data Transfer and Use Pilot** | Informational Resource, Licensing | 2017 | The Federal Demonstration Partnership | http://thefdp.org/default/committees/research-compliance/data-stewardship/ |

rights and licensing are multi-faceted. As discussed above, the timeline of initiatives demonstrated a shift in licensing standardization priorities, which may suggest that while the open license standardization efforts have been successful in meeting the needs of a particular segment of the data sharing community, there are still too many barriers that prevent researchers from sharing their data, and these data sharing challenges need to be met with more nuanced, robust, and interoperable licensing initiatives that can ensure the protection of more sensitive data types.

This research also produced additional key observations that could inform future research, but will require further analysis. An interesting metadata observation from the environmental scan results is that none of the rights or licensing-related standards and schemas were developed specifically for use with research data. Despite the proliferation of rights-related and licensing metadata schemas, one of the challenges is implementing commerce or library-centric metadata schemas for data-centric data sharing needs. Perhaps the use of multiple metadata formats could be encouraged in order to allow researchers to append their discipline-specific metadata standards with interoperable rights or licensing standards to communicate essential privacy and intellectual property requirements and limitations. The idea is to employ rights or licensing-specific metadata supplements as boundary objects that reach across communities (Star & Griesemer, 1989), facilitating interoperability between disparate data sharing communities within industry, academia, and government.

The two ontologies discovered, however, *are* specific to research data. The Data Use Ontology was developed specifically the facilitate the sharing of genomics data, which would most likely not be appropriate when sharing other types of research data. The Neurona Data Protection Ontology, while pertinent to data protection and security, is only relevant within the Spanish legal system and European Union data protection guidelines, and thus may not be appropriate for more widespread application. One potential avenue

forward to address this gap in research data-specific rights and licensing metadata standards is to develop a generic or cross-discipline ontology or standard for expressing rights and licensing metadata for the purposes of data sharing. By identifying cross-disciplinary rights management and licensing requirements for sharing private and sensitive data types, an information model could be developed to enable the sharing of disparate research data types across multiple domains.

The current landscape of initiatives seeking to address the rights management and licensing complications of data sharing is encouraging, but there are challenges regarding the implementation of these various efforts. For example, there are different applicable standards and policies for data sharing, not just between different disciplines and communities, but also between US-centric and international efforts. Data sharing initiatives in Europe may not be appropriate to meet data sharing needs in the United States, due to the disparate community-specific, local, and national regulations for protecting privacy.

The directory of data sharing initiatives examined in this paper is not exhaustive, and there are undoubtedly many other ongoing efforts to address the rights management and licensing challenges of sharing private and sensitive data types. Identifying all of the initiatives may not be possible, due to the varying progress, publicity, and impact level of initiatives, from local domain-specific repositories, to national or global efforts. Another limitation of this research is that the categories and subcategories used for this environmental scan are subjective in nature, established iteratively by the researchers, and could be categorized in different ways. However, the categories and sub-categories created by the researchers are intended to provide users with a quick glance at the scope and purpose of these rights and licensing efforts. Similarly, an additional challenge is that people from different backgrounds and perspectives within data communities may have varying notions of what qualifies as a standard, tool, or community initiative. For this study, effort was made to follow what seemed to be most consistent for our purposes and within the context of how these topics are generally understood within the RDA community.

## 8. Conclusion and Next Steps

The objective of this research was to provide clarity by offering a framework of the landscape of data sharing initiatives at the intersection of rights and licensing, based on the categories and subcategories used. This was accomplished through an environmental scan, which was performed through the collection, categorization, and presentation of results, including the development of a resource directory (Version 1.0). The results demonstrated how these 20 initiatives interrelated and differed, as well as how the trend of rights and licensing efforts have progressed over the last 16 years. Over time, efforts shifted from the development of open licensing standardization initiatives to more nuanced and technologically-focused efforts, which can accommodate for more sensitive and private data types. The directory was developed as a contribution for researchers, as a one-stop resource for understanding what organizations and people developed the initiative, when it was developed, what are the goals and current status, as well as where to find more information. Gathering information for the directory also identified insights and opportunities in the metadata and ontology community, including the need for universal rights and licensing metadata standards and ontologies specifically for use with research data.

As the landscape of data sharing initiatives continues to grow, clear next steps include connecting this resource to the Northeast Big Data Innovation Hub's data sharing spoke initiative, Drexel's Metadata Research Center, and the Research Data Alliance. We will provide a template to these organizations, for wider and further vetting and contribution to this directory. Additional next steps include further engagement with developing data sharing standards and best practices with the Research Data Alliance global community, as well as promoting the continued development of standards, tools, and communities that specifically promote the sharing of sensitive and private data types. Through the development of these initiatives and solutions, the prohibitively difficult process of sharing data will become easier, which is essential to support scientific research and innovation.

## Additional File

The additional file for this article can be found as follows:

· **Appendix.** Expanded directory: Standards, Tools, and Community Initiatives. DOI: https://doi.org/10.5334/dsj-2019-029.s1

## Acknowledgements

## Competing Interests

The authors have no competing interests to declare.

## References

**Bélanger, F** and **Crossler, RE.** 2011. Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly* 35(4): 1017–1041. DOI: https://doi.org/10.2307/41409971

**CEN.** 2018. European Committee for Standardization. Retrieved April 17, 2018 from https://www.cen.eu/Pages/default.aspx.

**Cooper, D** and **Schindler, P.** 2013. *Business research methods*, 12th Edition, McGraw Hill.

**Else, H.** 2016. Half of academics confused about open data. *Times Higher Education.* Retrieved April 24, 2018 from https://www.timeshighereducation.com/news/half-academics-confused-about-open-data#survey-answer.

**Fecher, B, Friesike, S** and **Hebing, M.** 2015. What Drives Academic Data Sharing? *PLoS ONE* 10(2): e0118053. DOI: https://doi.org/10.1371/journal.pone.0118053

**Grabus, S** and **Greenberg, J.** 2018. Resources for understanding the data sharing landscape: Rights, licensing, and related initatives. *Poster presented at the Research Data Alliance 11th Plenary Meeting.* Berlin, Germany.

**Granville, K.** 2018. Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens. *The New York Times.* Retrieved June 12, 2018 from https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html.

**Harding, A, Harper, B, Stone, D, O'Neill, C, Berger, P, Harris, S** and **Donatuto, J.** 2012. Conducting research with tribal communities: Sovereignty, ethics, and data-sharing issues. *Environmental Health Perspectives* 120(1): 6–10. DOI: https://doi.org/10.1289/ehp.1103904

**ISO.** 2018. International Organization for Standardization: When the world agrees. Retrieved April 20, 2018 from https://www.iso.org/home.html.

**Janssen, M, Charalabidis, Y** and **Zuiderwijk, A.** 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29(4): 258–268. DOI: https://doi.org/10.1080/10580530.2012.716740

**Jarnevich, CS, Graham, JJ, Newman, GJ, Crall, AW** and **Stohlgren, TJ.** 2007. Balancing data sharing requirements for analyses with data sensitivity. *Biological Invasions*, 9(5): 597–599. DOI: https://doi.org/10.1007/s10530-006-9042-4

**Krippendorff, K.** 2012. Content analysis: An introduction to its methodology (3rd ed.). Thousand Oaks, CA: Sage Publications.

**Liu, X, Li, X-B, Motiwalla, L, Li, W, Zheng, H** and **Franklin, PD.** 2016. Preserving Patient Privacy When Sharing Same-Disease Data. *Journal of Data and Information Quality* 7(4): 1–14. DOI: https://doi.org/10.1145/2956554

**National Research Council.** 1997. Bits of power: Issues in global access to scientific data. Washington, DC: The National Academies Press. DOI: https://doi.org/10.17226/5504

**National Science and Technology Council (U.S.), Committee on Science, & National Science and Technology Council (U.S.)** and **Interagency Working Group on Digital Data.** 2009. *Harnessing the power of digital data for science and society: Report of the interagency working group on digital data to the committee on science of the national science and technology council.* Washington, D.C.: Interagency Working Group on Digital Data.

**Oxenham, S.** 2016. Legal confusion threatens to slow data science. *Nature.* Retrieved April 24, 2018 from https://www.nature.com/news/legal-confusion-threatens-to-slow-data-science-1.20359. DOI: https://doi.org/10.1038/536016a

**Parrish, JL, Jr.** 2010. PAPA knows best: Principles for the ethical sharing of information on social networking sites. *Ethics and Information Technology* 12(2): 187–193. DOI: https://doi.org/10.1007/s10676-010-9219-5

**Rockhold, F, Nisen, P** and **Freeman, A.** 2016. Data sharing at a crossroads. *The New England Journal of Medicine* 375(12): 1115–1117. DOI: https://doi.org/10.1056/NEJMp1608086

**Star, SL** and **Griesemer, JR.** 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science* 19(3): 387–420. DOI: https://doi.org/10.1177/030631289019003001

**Swarup, V, Seligman, L** and **Rosenthal, A.** 2006. Specifying data sharing agreements. *Proceedings – Seventh IEEE International Workshop on Policies for Distributed Systems and Networks, Policy 2006*, 157–160. DOI: https://doi.org/10.1109/POLICY.2006.34

**Tenopir, C, Allard, S, Douglass, K, Aydinoglu, AU, Wu, L, Read, E, Frame, M,** et al. 2011. Data sharing by scientists: Practices and perceptions. *PLoS ONE* 6(6): 1–22. DOI: https://doi.org/10.1371/journal.pone.0021101