Expanded directory: Standards, Tools, and Community Initiatives (Appendix for: The Landscape of Rights and Licensing Initiatives for Data Sharing)

Cite this Appendix as:

Grabus, S. and Greenberg, J. 2019. Expanded directory: Standards: Tools, and Community Initiatives. Data Science Journal, 18: 29. DOI: <https://doi.org/10.5334/dsj-2019-029.s1>

Appendix is associated with this article:

Grabus, S. and Greenberg, J. 2019. The Landscape of Rights and Licensing Initiatives for Data Sharing. Data Science Journal, 18: 29, pp. 1–11. DOI: <https://doi.org/10.5334/dsj-2019-029>

Table of Contents

6.1 Standards	1
6.2 Tools	6
6.3 Community Initiatives	7
6.4 References	9

6.1 Standards

Name: Creative Commons

Sub-categories: Licensing, Rights

Date Initiated: 2001

Founded by: Lawrence Lessig

Current URL: <https://creativecommons.org/>

Goals: Develop a set of free copyright licenses that reduce barriers to creativity, yet also specify conditions for acceptable use (Creative Commons, 2017a).

Status: At the time of writing, there are seven types of Creative Commons license or dedication options, with specifications for sharing, attribution, derivative works (remixing), and commercial use permissions. The six licensing options are: Attribution (CC BY), Attribution Share-Alike (CC BY-SA), Attribution-NoDerivs (CC BY-ND), Attribution-NonCommercial (CC BY-NC), Attribution-NonCommercial-ShareAlike (CC BY-NC-SA), and Attribution-NonCommercial-NoDerivs (CC BY-NC-ND). The seventh and most recently-developed addition to Creative Commons is CC Ø (CC Zero), released in 2009, which is a “no rights reserved” public domain waiver option, as a means to unambiguously provide unfettered re-use possibilities as fully as the licensor can (Creative Commons, 2017b). In addition, the Creative Commons Rights Expression Language (*ccREL*) was proposed by Creative Commons in 2008, but has not yet been endorsed by the W3 (W3C, 2008). Applied through the use of Resource Description Framework (RDF) triples, ccREL is a machine-readable way to express licensing metadata (W3C, 2008).

Name: Open Data Commons

Sub-categories: Licensing, Rights

Started: 2007

Founded by: Open Knowledge Foundation

Current URL: <https://opendatacommons.org/>

Goals: Develop a set of free standardized licenses for data and databases that are interoperable with the Creative Commons licenses.

Status: The Open Data Commons developed three standardized licenses for data and databases: The Public Domain Dedication and License (PDDL), the Open Data Commons Attribution License (ODC-BY), and the Open Database License (ODbL) (Open Data Commons, 2017). The licenses differ from Creative Commons in that CC was designed for use with content, whereas the Open Data Commons licenses are designed for use with data and databases, which requires different considerations (Open Data Commons, 2017). The Public Domain Dedication and License (PDDL) is similar to and interoperable with the Creative Commons CC 0 (CC Zero) public domain specification, designed to standardize the communication of open data licensing information using simple and transparent language (Open Data Commons, 2017). The Open Data Commons Attribution License (ODC-BY) is the data/database equivalent to the Creative Commons attribution license, commonly known as CC-BY, allowing the user to share, create, and adapt the data or database, but requires attribution to the original source (Open Data Commons, 2017). The Open Database License (ODbL) is an attribution and share-alike license for data and databases (equivalent to the Creative Commons CC BY-SA license), which allows the user to freely share, create, and adapt the data/database, as long as the user attributes to the original source, as well as sharing the adapted work under the same ODbL License (Open Data Commons, 2017).

Name: The Open Government License

Sub-categories: Licensing, Rights

Started: 2010

Founded by: UK National Archives

Current URL: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Goals: Develop a free standardized license for public sector information within the UK, compatible with the Creative Commons and Open Data attribution licenses (The National Archives, n.d.).

Status: The Open Government License (OGL) allows users to copy, publish, distribute, adapt, and commercialize data, with the caveat that the original source must be attributed, and that the license is governed by the laws presiding over the place of business at which the licensor is employed (The National Archives, n.d.).

Name: RightsStatements.org

Sub-categories: Rights

Started: 2015

Founded by: DPLA (Digital Public Library of America) and Europeana

Current URL: <http://rightsstatements.org/en/>

Goals: Develop a set of standardized rights statements as a means to communicate the rights status for works in which the party displaying the information object may encounter legal or factual uncertainty about the specifics of the copyright information and usage (International Rights Statements Working Group, 2015).

Status: The 12 rights statements categorize entities in three ways: In copyright, no copyright, other, with a few specific varieties for each. For example, copyrighted materials can be described in the following ways: In Copyright, EU Orphan Work, Educational Use Permitted, Non-Commercial Use Permitted, and Unknown Rightsholder. Using the “Other” category, an organization or person can specify that the copyright status of the particular information object has either not been evaluated, is undetermined, or is not known to have a copyright, beyond a reasonable doubt.

Name: Linked Content Coalition

Sub-categories: Rights, Metadata & Ontologies

Started: 2010

Founded by: European Publisher’s Council

Current URL: <http://www.linkedcontentcoalition.org/>

Goals: To promote digital data reuse and more efficiently communicate copyright information for digital objects, focusing their efforts on the interoperability of identifiers and rights metadata (Linked Content Coalition, 2018). To develop tools and prototypes, which can be implemented into systems.

Status: In 2013, the Linked Content Coalition developed the LCC Framework, which consisted of an abstract data model called the LCC Rights Reference Model, for representing intellectual property rights information; principles of identification; and the rights data supply chain. Under the LCC Rights Reference Model, rights entities can be represented through the following types of context: Right, RightsAssignment, Assertion, and RightsConflict (Rust, 2013). From 2013-2015, the LCC implemented the Rights Data Integration (RDI) Project, which served as an instantiation of the LCC Framework for data from 20 different sources, to demonstrate the feasibility of integrating and using standardized data (Rights Data Integration, 2016).

Name: The Data Use Ontology

Sub-categories: Rights, Metadata & Ontologies

Started: 2017

Founded by: Global Alliance for Genomics and Health

Current URL: <https://github.com/EBISPOT/DUO>

Goals: Develop an ontology to facilitate the sharing of genomics research data (Courtot, 2017).

Status: The Data Use Ontology is based on data use “consent codes,” which were divided into primary and secondary categories based on data types and restrictions (Courtot, 2017). These codes were based on guidance from the National Institutes of Health (NIH); surveys of conditions for data use within large research institutions and a variety of genomics organizations; and guidance from the Global Alliance Data Working Group and Matchmaker Exchange Project (Dyke, et al., 2016).

Name: The Neurona Data Protection Ontology

Sub-categories: Rights, Metadata & Ontologies

Started: 2008

Founded by: S21sec security company and the Institute of Law and Technology at the Universitat Autònoma de Barcelona (Casellas, Nieto, Meroño, Roig, & Torralba, 2008).

Current URL: N/A

Goals: Develop a non-domain-specific ontology for data protection compliance within the Spanish legal system.

Status: The development of this ontology was informed by guidance from legal experts within the Spanish legal system, regarding personal data, consent, purpose of use, and data security, combined with an analysis of data protection regulations, corporate data protection concepts, and technological security. (Casellas, Nieto, Meroño, Roig, & Torralba, 2008). Some of the key concepts of the ontology are data, consent, purpose, security measures, person, treatment process, and security degree (Casellas, Nieto, Meroño, Roig, & Torralba, 2008).

Name: W3C Permissions & Obligations Expression

Sub-categories: Rights, Licensing, Metadata & Ontologies

Started: 2016

Founded by: W3C

Current URL: https://www.w3.org/2016/poe/wiki/Main_Page

Goals: Create a new semantic information model for expressing permissions, obligations, rights statements, and content licensing (W3C, 2016). As an initiative that allows for the expression of more complex and business-specific obligations, the initiative seeks to address different business needs than those addressed by Creative Commons.

Status: This information model was structured beginning with the ODRL (Open Digital Rights Language) specifications, but builds upon this by factoring in the “additional needs of the wider community including commercial and non-commercial publishers, consumers and re-users of human readable content and machine readable data” (W3C, 2016). The projected end date for the development of the information model is projected for April 2018 (W3C, 2016).

Name: ONIX-PL

Sub-categories: Licensing, Metadata & Ontologies

Started: 2008

Founded by: Digital Library Federation’s Electronic Resource Management Initiative (ERMI) and EDItEUR/NISO (Editeur, 2009a).

Current URL: <http://www.editeur.org/21/ONIX-PL/>

Goals: Create an XML structure to document publication licensing agreements, originally developed for publication commerce purposes (Editeur, 2009a).

Status: This structure expresses licensing terms in an XML format, allowing for machine-readable iterations of expression detail, license detail, definitions, license grant, usage terms, supply terms, continuing access terms, payment terms, general terms, and the license document text itself (Editeur, 2009a). The OPLE open source browser software enables users to create and edit expressions using ONIX-PL (Editeur, 2009b).

Name: RightsDeclarationMD Extension Schema

Sub-categories: Rights, Metadata & Ontologies

Started: 2001

Founded by: Digital Library Foundation for digital library objects

Goals: Create rights-based extension of METS (Metadata Encoding and Transmission Standard), called the RightsDeclarationMD Extension Schema

Current URL: <https://www.loc.gov/standards/rights/METSRights.xsd>

Status: RightsDeclarationMD (or METS Rights) specifies permissions and constraints for users regarding how that information object can be handled (The Library of Congress, 2016). For example, permissions to discover, display, copy, duplicate, modify, delete, or print, as well as constraints for quality, format, watermark, unit, payment, count, attribution, re-use, time, and transfer permissions (The Library of Congress, 2016).

Name: Open Digital Rights Language (ODRL)

Sub-categories: Rights, Metadata & Ontologies

Started: 2001

Founded by: W3C Permissions & Obligations Expression Working Group

Goals: The development of a Digital Rights Management (DRM) language for the purposes of standardizing the communication of rights information for digital resources (About ODRL, 2018).

Current URL: <https://www.w3.org/community/odrl/>

Status: The ODRL information model specifies permissions, constraints, requirements, and contextual elements. The schema includes permissions for usage, transfer, asset management, and re-use. Constraint elements include those for the user, devices, boundary constraints, temporal restraints, aspect constraints, and rights. Requirement elements include payment, interactions requirements, and usage requirements. Rights holder elements include royalty. Contextual elements include the following: UID, name, role, remark, date, physical location, digital location, external reference, service, version, transaction, and event.

Additional Standards with a Rights Entity

- Dublin Core
- Metadata Object Description Schema (MODS)
- CopyrightMD, Extension for use with the Metadata Encoding and Transmission Standard (METS)
- Preservation Metadata: Implementation Strategies (Premis)
- Categories for the Description of Works of Art (CDWA) Lite
- Learning Object Metadata (LOM)
- Visual Resources Association (VRA) Core

- Public Broadcasting Metadata Dictionary (PBCore)

6.2 Tools

Name: ShareDB: A Licensing Model and Ecosystem for Data Sharing

Sub-categories: Licensing

Started: 2016

Founded by: Drexel University's Metadata Research Center, MIT, Brown University

Current URL: <https://cci.drexel.edu/mrc/research/a-licensing-model-and-ecosystem-for-data-sharing/>

Goals: Facilitate the process of data sharing among industry, academia, and government (Metadata Research Center, 2018). The project has three interconnected components: a licensing framework/generator; a data sharing platform, called ShareDB, to enforce some of the privacy and compliance aspects of the licenses; and robust metadata for the datasets as well as the developed licenses, to communicate data handling and rights specifications in a machine-readable way.

Status: The first workshop, held in September 2016, gathered data sharing perspectives from industry, academia, and government representatives, to identify success stories, failures, challenges, and opportunities for technological and legal frameworks to facilitate the complex data sharing process (Greenberg et al., 2017). Researchers have constructed a preliminary data sharing agreement ontology based on a content analysis of data sharing agreements (Grabus & Greenberg, 2017), and are currently developing technological solutions for enforcing data protection requirements, such as differential privacy.

Name: DataTags

Sub-categories: Rights

Started: 2015

Founded by: Harvard's Dataverse

Current URL: <https://datatags.org/>

Goals: Develop a system for tagging datasets in repositories by their varying levels of sensitivity and data sharing restrictions (Sweeney, Crosas, & Bar-Sinai, 2015).

Status: This Datatags initiative takes the complex governing restrictions and regulations that surround sharing data and reduces it to a more simplified process, through a data sharing decision tree. This simplified process incorporates a graduated scale of severity for access requirements and restrictions, enforced with levels of security features such as password protection, registration, approval, click-through or signed data use agreements, and two-factor authentication. A datatags-compliant repository will integrate two or more datatags into the design of the repository, with each dataset designated under one datatag level. Ultimately, this initiative helps to facilitate data sharing by enabling a repository to store and disseminate datasets with disparate privacy and security restrictions.

Name: Legal Assessment Tool (LAT)

Sub-categories: Informational Resource, Licensing

Started: 2016

Founded by: BioMedBridges

Current URL: N/A

Goals: Develop a query workflow user interface that assesses the types of data being shared and presents results for what restrictions and requirements are applicable (Kuchinke et al., 2016).

Status: The project implements a strategy of requirements engineering through usage scenarios to create legal requirement clusters. While researchers first attempted to develop a data sharing ontology based on the privacy restrictions, they felt that ontologies "...are still not suitable to represent the complexity and intricacy associated with many of the rules and combination of rules [they] encountered" (Kuchinke et al., 2016).

6.3 Community Initiatives

Name: Research Data Alliance

Sub-categories: Rights, Licensing, Metadata & Ontologies

Started: 2013

Founded by: European Commission, the US National Science Foundation (NSF), and the Australian Government's Department of Innovation (The Research Data Alliance, 2017).

Current URL: <https://www.rd-alliance.org/>

Goals: Facilitate data re-use through the development of interest groups and working groups that collaborate to develop outputs such as policies, recommendations, standards, codes, and best practices (Research Data Alliance, 2017c).

Status: Specifically related to rights management and licensing, the RDA/CODATA Legal Interoperability Interest Group focuses on developing guidelines and principles toward the contractual compatibility of data, considering the difficulties of reconciling disparate privacy and rights-related policies and needs (Research Data Alliance, 2017a). The RDA/NISO Privacy Implications of Research Data Sets Interest Group also focuses their efforts on sharing sensitive data types. This interest group seeks to develop a framework for developing awareness about the privacy issues involved with sharing research data (Research Data Alliance, 2017b).

Name: Datasets Licensing Project

Sub-categories: Licensing, Metadata & Ontologies

Started: 2017

Founded by: Jisc, The University of Glasgow, and CREATE

Current URL: <https://datasetlicencing.wordpress.com/>

Goals: Provide licensing support and guidance for data creators and users (Jisc, CREATE, & The University of Glasgow, 2017). Use insights gained from the Datasets Licensing Project to inform future efforts for the Jisc Research Data Shared Service (Jisc, 2018).

Status: The first Datasets Licensing Project workshop was held in November 2017. Talking points from this workshop included the implementation of plain language terminology (through provision of a glossary), funder recommendations for license usage, automating the process of applying and embedding licensing metadata for multiple files, and machine-actionable licenses. (Jisc, CREATE, & The University of Glasgow, 2017).

Name: DCC's How to License Research Data

Sub-categories: Informational Resource, Licensing, Metadata & Ontologies

Started: 2014

Founded by: Digital Curation Centre

Current URL: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

Goals: Develop a guide for licensing research data, including information on why to license research data, licensing concepts, prepared licenses, bespoke licenses, standard licenses, multiple licensing, mechanisms for licensing data, and licensing-related information (Ball, 2014).

Status: The mechanisms for licensing data section of these guidelines provides an overview of various access and rights-related metadata standards and data formats to facilitate the data licensing process.

Name: The (Re)usable Data Project

Sub-categories: Informational Resource, Licensing

Started: 2017

Founded by: National Center for Advancing Translational Sciences (NCATS) Biomedical Data Translator and the Monarch Initiative (The Reusable Data Project, 2017).

Current URL: <http://reusabledata.org/>

Goals: To facilitate legal interoperability through conversation about accessibility, comprehension, clarify, and licensing decisions that affect data sharing within the scientific community. (The Reusable Data Project, 2017). Evaluate biological and biomedical data sources for their data reusability and redistribution potential.

Status: Biology and biomedical resources are evaluated on a scale of 0-5 Stars, determined by these data re-use metrics. For example, at the time of writing, the Dryad Digital Repository is evaluated at 4.5 stars, meaning that while there may be some terms of use, the repository unambiguously allows for the re-use and redistribution of data (The Reusable Data Project, 2017).

Name: FAIRsharing.org

Sub-categories: Informational Resource, Metadata & Ontologies

Started: 2009

Founded by: University of Oxford and the Oxford e-Research Centre

Current URL: <https://fairsharing.org/>

Goals: Provide a comprehensive information resource to guide researchers toward increasing data sharing and supporting reproducible research through providing recommendations, collections, and educational resources about data sharing policies, standards, and databases.

Status: Users can contribute to the website by adding databases, policies (e.g., organization data sharing policies), standards (e.g. discipline-specific schemas and ontologies), recommendations (e.g., repositories

recommended by specific journals, funders, or organizations) and collections (e.g., Repositories recommended by NIH), all of which makes the resource more discoverable to a wider audience, and aids the FAIRsharing community in maintaining their comprehensive hub of resources.

Name: The Federal Demonstration Partnership: Data Transfer and Use Agreement Pilot

Sub-categories: Informational Resource, licensing

Started: 2017

Founded by: The Federal Demonstration Partnership

Current URL: <http://thefdp.org/default/committees/research-compliance/data-stewardship>

Goals: The broad goal of the Federal Demonstration Partnership (FDP), a cooperative effort among 10 United States federal agencies and 154 institutional federal funding recipients, is to mitigate administrative loads related to research grants and contracts (The Federal Demonstration Partnership, 2018). As it relates to data stewardship, the FDP has conducted a Data Transfer and Use Agreement pilot towards the creation of data transfer and use agreement templates (The Federal Demonstration Partnership, 2018). These templates were developed to reduce the complex and time-consuming process of crafting individualized data transfer and use agreements for each research collaboration.

Status: Finalized and published template sections include project-specific details, identification of permitted collaborators, and attachments for sharing de-identified human subject data or limited data sets. FDP has also published a chart for guiding researchers on how to classify human subject data.

6.4 References

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4), 1017-1041.

Board on Research Data and Information, Policy and Global Affairs, Council, N. R., & National Academies (U.S.). Committee on the Future Career Opportunities and Educational Requirements for Digital Curation. (2015). *Preparing the workforce for digital curation*. Washington, D.C: National Academies Press.

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., & Hebing, M. (2015). What Drives Academic Data Sharing? *PLoS ONE* 10(2): e0118053. <https://doi.org/10.1371/journal.pone.0118053>

Grabus, S., & Greenberg, J. (2018). Resources for understanding the data sharing landscape: Rights, licensing, and related initiatives. 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>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Harding, A., Harper, B., Stone, D., O'Neill, C., Berger, P., Harris, S., & Donatuto, J. (2012). Conducting research with tribal communities: Sovereignty, ethics, and data-sharing issues. *Environmental Health Perspectives*, 120(1), 6–10. <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., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <https://doi.org/10.1080/10580530.2012.716740>
- Jarnevich, C. S., Graham, J. J., Newman, G. J., Crall, A. W., & Stohlgren, T. J. (2007). Balancing data sharing requirements for analyses with data sensitivity. *Biological Invasions*, 9(5), 597–599. <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., & Franklin, P. D. (2016). Preserving Patient Privacy When Sharing Same-Disease Data. *Journal of Data and Information Quality*, 7(4), 1–14. <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. <https://doi.org/10.17226/5504>
- National Science and Technology Council (U.S.). Committee on Science, & National Science and Technology Council (U.S.). 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>
- Parrish, J.L., 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:10.1007/s10676-010-9219-5
- Rust, G. (2013). *The LCC rights reference model*. Retrieved February 2, 2018 from <http://www.linkedcontentcoalition.org/index.php/rights-data-network/lcc-framework>
- Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, ‘Translations’ and Boundary Objects : Amateurs and Professionals in Berkeley ’ s Museum of Vertebrate Zoology , 1907-39 Author (s) : Susan Leigh Star and James R . Rockhold, F., Nisen, P., & Freeman, A. (2016). Data sharing at a crossroads. *The New England Journal of Medicine*, 375(12), 1115-1117. doi:10.1056/NEJMp1608086. Griesemer Source : Social Studies of Science , Vol . 19 , No . 3. *Social Studies of Science*, 19(3), 387–420.

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

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