Journal of eScience Librarianship

Volume 1 | Issue 3 Article 4

2013-03-13

Proprietary vs. General Data Formats: Opportunity for Librarians

James Schroeder
MIT Lincoln Laboratory

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/jeslib

Part of the Library and Information Science Commons

Repository Citation
Schroeder J. Proprietary vs. General Data Formats: Opportunity for Librarians. *Journal of eScience Librarianship* 2013;1(3): e1031. https://doi.org/10.7191/jeslib.2012.1031. Retrieved from https://escholarship.umassmed.edu/jeslib/vol1/iss3/4

Creative Commons License

This work is licensed under a Creative Commons Attribution 4.0 License.
This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.

UMass Chan MEDICAL SCHOOL

JESLIB 2012; 1(3): 118-119
doi:10.7191/jeslib.2012.1031

Journal of eScience Librarianship
putting the pieces together: theory and practice

Proprietary vs. General Data Formats: Opportunity for Librarians

James Schroeder

Massachusetts Institute of Technology Lincoln Laboratory Library, Lexington, MA, USA

The article *Lurking in the Lab: Analysis of Data from Molecular Biology Laboratory Instruments* in this issue highlights an important evolving issue for science information professionals. That is the inexorable increase in proprietary information formats and the accompanying need to seamlessly convert the information generated through these formats into generic formats for preservation, migration, and dissemination.

During my last position at Cold Spring Harbor laboratory, one of the world's most prominent genomics laboratories, there was much talk and effort around the ongoing conversion of research laboratories from "wet" labs (where researchers primarily work "on the bench" doing hands-on experiments) to "dry" labs (where experiments are highly automated and software driven, and where researchers are spending more of their time in front of screens interpreting the data). This is being driven by extraordinary technological advancements such as gene sequencing technologies. Sequencing has advanced from the technologies utilized in the 1990s Human Genome Project (which required roughly eight years and three billion dollars to create a top level sequence), to the next-generation nanopore sequencers currently in development that promise a whole genome sequence in a week for roughly $5,000! Indeed, some of Cold Spring Harbor's most prominent labs and researchers are now exclusively devoting their research to developing software for processing and visualizing the greater flood of data from the increasingly automated labs of their colleagues. One of the most challenging, well funded, "hot" areas of research is not generating experimental data but creating software to interpret it.

Obviously then proprietary software that drives these advancements is not only necessary but will continue to grow in complexity and scope. Also obvious is that the seamless, accurate conversion of the data and metadata generated by these proprietary software formats into data formats that can be seamlessly preserved, migrated, and disseminated will need to keep pace with the ongoing development of these proprietary formats. As members of the scientific community who are professionally committed to the accurate, comprehensive preservation of scientific information that is the lifeblood of their institutions, I believe that it is librarians who have the skills and the training to play a vital role in creating institutional standards, protocols, ontologies, and communities of practice around the curation of information from proprietary to generic formats. These issues present an exciting opportunity and increasingly important institutional imperative for science librarians.

Correspondence to James Schroeder: james.schroeder@ll.mit.edu
Keywords: research data, data formats, proprietary software

JESLIB 2012; 1(3): 118-119
doi:10.7191/jeslib.2012.1031

All content in Journal of eScience Librarian-ship, unless otherwise noted, is licensed under a Creative Commons Attribution 4.0 International License.

http://creativecommons.org/licenses/by/4.0

ISSN 2161-3974