University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England Area Librarian e-Science Symposium

2012 e-Science Symposium

Apr 4th, 12:00 AM - 2:00 PM

Lurking in the Lab: Analysis of Data on Molecular Biology Laboratory Instruments

Jen Ferguson
Northeastern University

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

 Part of the Library and Information Science Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

Repository Citation

Ferguson, J. (2012). Lurking in the Lab: Analysis of Data on Molecular Biology Laboratory Instruments. *University of Massachusetts and New England Area Librarian e-Science Symposium.* https://doi.org/10.13028/a3hp-sx15. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2012/posters/3

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.
This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Lurking in the lab: Analysis of data from molecular biology laboratory instruments

Jen Ferguson • Northeastern University • Boston MA

Introduction

Science labs are filled with laboratory instruments that gather and store data. These data sets are neither glamorous nor polished, and very little of this data will ultimately be published. However, they represent the reality that science faculty, staff, and students are attempting to manage daily.

A faculty member granted access to instruments used in her molecular biology teaching lab.

Experimental data files generated by these instruments were examined in order to:

- Gain insight into lab data management practices
- Evaluate possible curation/preservation challenges posed by this data

Materials and methods

Xplorer² file management software was used to flatten directory structures and capture files and metadata on hard drives dedicated to these laboratory instruments:

- Agilent Technologies atomic force microscope
- LI-COR NEN DNA analyzer
- Bio-Rad Gel Doc XR
- Hitachi fluorescence spectrophotometer

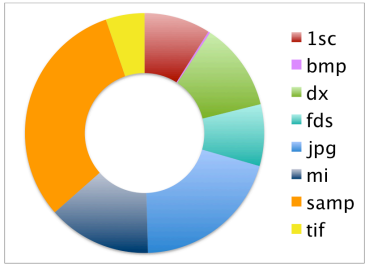
Xplorer² and Microsoft Excel were used to sort and analyze experimental data files. Formats were categorized as proprietary or open/standard.

Informal discussions with the faculty member & lab staff during the course of this work also informed the findings.

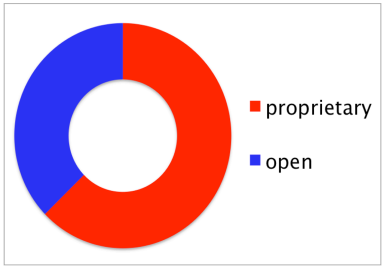
Results

Quantitative analysis of data files

Eight most prevalent file formats



62% of files were in proprietary formats



Additional findings

Image files in various formats were the most prevalent type of experimental data found, accounting for 65% of the data files on the lab instrument hard drives.

Data from these instruments can be stored as packages comprised of several different file types. For example, DNA sequence data can consist of sequence text file (.txt) + curve file (.scf) + gel image file (.samp).

One instrument (the LI-COR DNA analyzer) had very few data files stored on its hard drive. This instrument uses drive space as temporary storage only; the oldest data sets are continuously overwritten. Some laboratory instruments, including the LI-COR, can be configured to save data to cloud storage supplied by the instrument's vendor.

Faculty & staff perspectives

Teaching lab faculty and staff noted challenges in organizing files so students could find them, due to:

- Inconsistent file storage locations & naming conventions
- Inconsistent data management practices from one lab instrument to the next

They observed that students found working with experimental data frustrating, largely as a result of barriers imposed by proprietary file formats. Challenges also arose due to issues such as difficulty manipulating 32-bit vs. 16-bit vs. 8-bit images.

Conclusions

Experimental data files in both proprietary and open formats were found in analysis of lab equipment hard drives.

Image files made up the bulk of the experimental data and were usually in open formats.

Other experimental data files tended to be in proprietary formats.

Several factors pose potential challenges to data use, sharing, management, and preservation:

- Automatic overwriting and/or cloud storage of data sets
- Data consisting of bundles of different file types
- Proprietary file formats

Acknowledgments

Thanks to M. Kosinski-Collins & D. Bordne of Brandeis University for generously providing insights and access to their laboratory equipment.

Sources

Cornell University Library/Research Department. (2003). Common image file formats. In *Moving theory into practice: Digital imaging tutorial.* Retrieved from http://www.library.cornell.edu/preservation/tutorial/presentation/table7-1.html

File-extensions.org - File extension library. (2012). Retrieved from www.file-extensions.org

LI-COR, Inc. (2006). *Operator's manual: NEN model 4300 DNA analyzer.* Retrieved from http://biosupport.licor.com/docs/4300_OpMan_08591.pdf

Zabkat Software. (2011). Xplorer² portable edition (Version 1.8.1) [Software]. Available from http://zabkat.com/x2port.htm