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

Video Article

Touring a Data Curation Network Primer: A Focus on Neuroimaging Data

Sara Samuel[1], Michael Moore[2], Helenmary Sheridan[3],
Chris Sorensen[4], and Brandon Patterson[5]

[1] University of Michigan, Ann Arbor, MI, USA
[2] University of Washington, Seattle, WA, USA
[3] University of Pittsburgh, Pittsburgh, PA, USA
[4] Washington University in St. Louis, St. Louis, MO, USA
[5] University of Utah, Salt Lake City, UT, USA

Abstract

This video article provides an introduction to a data primer which leads data curators through the process of preparing a neuroimaging dataset for submission into a repository. A team of health sciences librarians and informationists created the primer, which is focused on data from functional magnetic resonance images that are saved in either DICOM or NIfTI formats. The video walks through a flowchart discussing the process of preparing data sets to be deposited into a repository, key curatorial questions to ask for data that is highly sensitive, and how to suggest edits to this and other primers. The primer grew out of a data curation workshop hosted by the Data Curation Network.

Hi! My name is Sara Samuel. I am an informationist at the University of Michigan and I'm a member of a team who put together a data curation primer to help data specialists understand neuroimaging data.

The primer grew out of a data curation workshop held in November of 2019, which was hosted by the Data Curation Network or DCN. Since 2017, the DCN has spearheaded the creation of data curation primers, which are open-access documents that provide detailed information about different types of data, their technical requirements, and curatorial considerations for the data's long-term preservation, access, and use. Data curation primers are aimed at helping curators prepare a dataset for submission into a repository, but they can be helpful for anyone looking to have a greater understanding of specific data types and formats.

Our team focused our primer on neuroimaging data from functional magnetic resonance imaging, or fMRI. Specifically, our primer is focused on curating fMRI images that are saved in either DICOM or NIfTI formats. These are the most common formats for fMRI images. DICOM images tend to be very large and have a lot of metadata embedded within the file. If you are working with imaging data that came straight off the instrument, it will probably be in DICOM. The NIfTI format is newer, and tends to be more commonly used in research settings as opposed to clinical. It is more interoperable than DICOM, with smaller file sizes that make it better suited for analysis, sharing, and uploading in a data repository. Both DICOM and NIfTI formats can include headers with information about the files.

Our primer begins with a flowchart which we'll walk through now. Let's imagine you are the curator and you've just received DICOM or NIfTI files to prepare to be deposited into a repository. The general recommendation is to convert DICOM files to NIfTI and ensure that there is good documentation and metadata to accompany the files. There are several openly available tools to help data specialists with opening, converting, and organizing the files during each step of the curation process. These programs are linked to in the primer. Let's walk through the process together.

After a data set has been received, first check to see if it can be opened. If not, contact the user and request re-submission. If it can be opened, check to see if header information is included. If not, contact the user and request re-submission with the missing information. The next steps will depend if the dataset is in DICOM with a .dcm extension or already in NIfTI with a .nii extension. If the files are in DICOM format and the header information is included, the next step is to anonymize the data, then convert it to NIfTI. If the files are in NIfTI format and the header is included, review the data header manually for personal identifiers. Following the conversion and review, organize the data structure using the Brain Imaging Data Structures, or BIDS, by using the BIDS starter kit, then validate it using the BIDS-Validator. After a final review of the files, the data set can be uploaded into the repository.

Curating medical data adds an additional challenge not faced in other domains: datasets may include highly sensitive and identifying patient health information, a large portion of which is protected by strict federal laws like HIPAA. Researchers should only be submitting datasets that have already been anonymized, but data curators have an ethical responsibility to ensure that the process has been done correctly before hosting the files. Protected health information may be found both in the image header and within the images themselves. That information can cover a wide range of areas, from textual data like names, locations, and dates, to detailed and recognizable facial features from high-resolution scans. Tools that convert the images from DICOM to NIfTI can be set to automatically strip out that data. But data curators should manually review the final dataset before uploading it into their institutional repository to ensure no health information slipped by.

Key curatorial questions you might ask include:

- Has patient data been removed from the header of a DICOM file?
- For high-resolution structural images, have facial features been removed from the images?
- Has any "burned-in" text been removed if it is protected health information?
- Are data in raw format? If not, has the researcher provided documentation of processing procedures?

Other highlights of the primer include:

- Detailed descriptions of the formats, and their respective pros and cons
- Example datasets
- Software for viewing data
- Suggested repositories for sharing neuroimaging data
- Key curatorial considerations and preservation actions
- And finally, our primer addresses how to help the data meet the FAIR Principles of improving the findability, accessibility, interoperability, and reuse of digital data assets. These principles are published by FORCE11, a community to help facilitate the change toward improving knowledge creation and sharing.

Our primer is available for anyone to use. The original version has been deposited into the Digital Conservancy at the University of Minnesota, and there is also a version stored in GitHub. The version on GitHub is a living document which can be updated with new information in the future. You can suggest edits to this and other primers by visiting the "Contributors Guide" link on the DCN primers' GitHub page.

Visit datacurationnetwork.org to learn more about the DCN and access the different data primers.

Please contact a team member if you have any questions about the neuroimaging data primer.

Our team consists of Michael Moore from the University of Washington; Brandon Patterson from the University of Utah; myself, Sara Samuel from the University of Michigan; Helenmary Sheridan from the University of Pittsburgh; and Chris Sorensen from Washington University in St. Louis. Special thanks to our DCN mentor, Joel Herndon from Duke University, and the others who taught at the Data Curation Workshop in November 2019.

Acknowledgements

Data Availability

Moore, Michael, Brandon Patterson, Sara Samuel, Helenmary Sheridan, and Chris Sorensen. 2020. "Neuroimaging DICOM and NIfTI Data Curation Primer." Data Curation Network. University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/216582

References

Brain Imaging Data Structure. n.d. "Brain Imaging Data Structure." Accessed June 11, 2021. https://bids.neuroimaging.io

Gao, Chuanji, Christine E. Weber, Douglas H. Wedell, and Svetlana V. Shinkareva. 2021. "SUB-01_T1W.NII." In *fMRI: Audiovisual Valence Congruence*. OpenNeuro. https://doi.org/10.18112/openneuro.ds003507.v1.0.1

González-Vivas, Carlos, Gracián García-Martí, Pau Soldevila-Matías, Roberto Sanz-Requena, Eduardo J. Aguilar, María José Castro-Bleda, Luis Martí-Bonmatí, and Julio Sanjuan. 2020. "First-Episode Psychotic Patients Showed Longitudinal Brain Changes Using FMRI With an Emotional Auditory Paradigm." *Frontiers in Psychiatry* 11: 1374. Figure 1. https://doi.org/10.3389/fpsyt.2020.593042

Moore, Michael, Brandon Patterson, Sara Samuel, Helenmary Sheridan, and Chris Sorensen. 2020. "Neuroimaging DICOM and NIfTI Data Curation Primer." Data Curation Network. University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/216582