

A FAIR Principle Data Model for Focused Ion Beam Scanning Electron Microscopy (FIB-SEM) and the Frederick National Laboratory Data Coordinating Center

Paul Aiyetan^{1,2*}, Kedar Narayan^{1,3}, David Mott¹, Rajani Kuchipudi¹, Corinne Zeitler¹, Debra Hope¹, Uma Mudunuri¹ and Andrew Quong¹

¹ Frederick National Laboratory for Cancer Research, Frederick, MD, USA.

² George Mason University School of Systems Biology, Manassas, VA, USA.

³ Center for Molecular Microscopy, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA.

* Corresponding author: paul.aiyetan@nih.gov

The need to represent biomedical data in manners that facilitate ease of reuse has never been more apparent [1][2]. Beyond such datasets being well-archived is the importance of the richness of the associated metadata. Well-annotated metadata provides not only insights into the experiment that generated the data but also the processes employed. The ISA framework [3], in addition to satisfying the FAIR (Findable, Accessible, Interoperable, Reproducible) principle [4], provides a rich description of experimental metadata that is agnostic to sample characteristic, technology or measurement type. It provides clear and simple sample-to-data relationships that enable data and discoveries to be reproducible and reusable. It describes a scientific investigation, its study or studies, and each study's assay(s). We employed this to describe a typical Focused Ion Beam Scanning Electron Microscopy (FIB-SEM) experiment.

First, we illustrated the processes employed with an experimental graph (Figure 1). Thereafter identifying processes and materials (the primitive units of the ISA model), we derived a single investigation, a single study, and a three or four assay type model representation (Figure 2).

Our in-house generated data are hosted at the Frederick National Laboratory for Cancer Research Data Coordinating Center (FNLCCR-DCC) (*see sample archive here* <https://cssi-dcc.nci.nih.gov/cssiportal/view/5ac3e62d37384e051c7ab310/> [5]). The FNLCCR-DCC and Toolset (the Metadata Editor plus Validator) embodies a scalable, next-generation biological and cancer research data repository that is flexible, intuitive, and adaptive through implementation of the ISA (Investigation-Study-Assay) framework. It provides integrated management of datasets across all deposited projects making its data more accessible and easily reusable by the cancer research community. It stores and manages access to data, enabling researchers or data depositors to grant controlled access only to specific collaborators while maintaining a user-specified embargo on deposited datasets. It enables a data access and sharing capability aimed to facilitate the development of new biological insights from deposited datasets including FIB-SEM experiment datasets.

In conclusion, the ISA model together with the FNLCCR-DCC provides a much-needed solution to processing, storing and distributing large image data to the research community.

References:

- [1] JPA Ioannidis. PLoS Med. **2** (2005), p. e124.
- [2] V Stodden, F Leisch and RD Peng, (CRC Press).
- [3] S-A Sansone et al., Nat Genet. **44** (2012), p. 121.
- [4] MD Wilkinson et al., Sci Data. **3** (2016), p. 160018.
- [5] L Balagopalan et al., Nat Comm. **9** (2018), p. 2013.

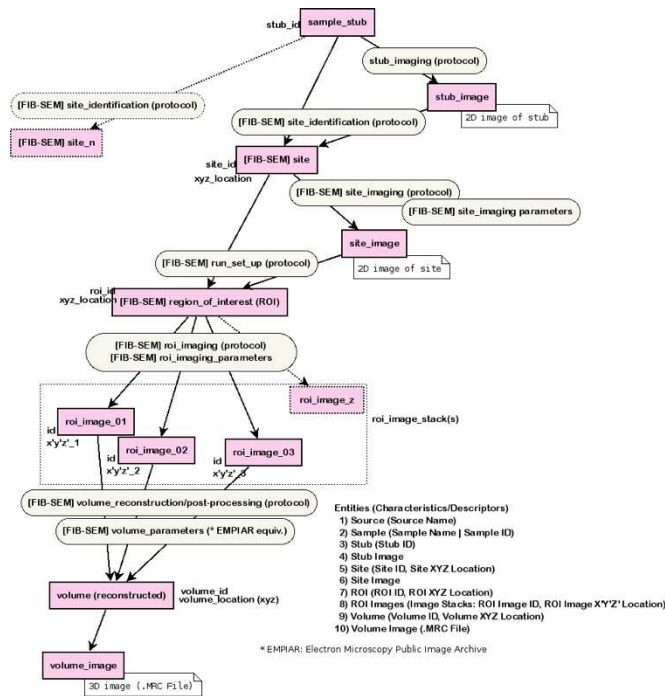


Figure 1. Sample experimental graph of a typical FIB-SEM imaging experiment

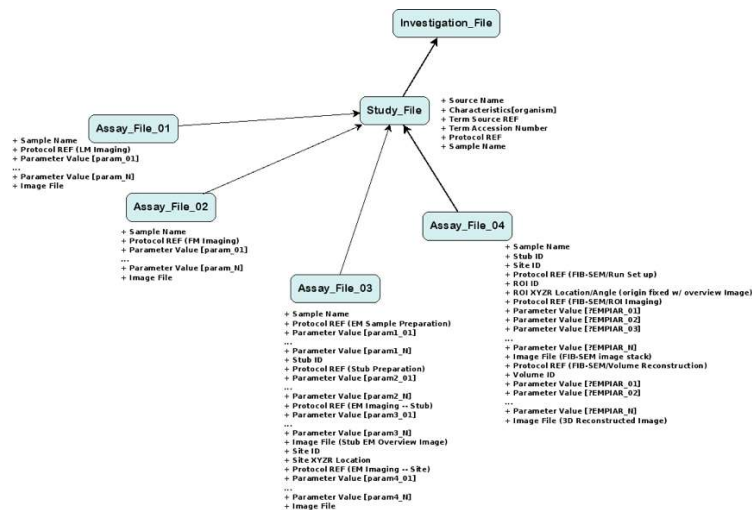


Figure 2. An ISA model representation of a typical FIB-SEM imaging experiment