

A Novel New Approach to Compose, Develop and Implement Cell Image Ontology and Application to Semantic Web Databasing

Authors: Talapady N. Bhat, John T. Elliott, Anne L. Plant,
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Cell images are a mainstay of cell biology data because of the vast amount of information that they can contain. In addition to the information on the explicit analytes of interest, images contain information such as spatial and intensity relations that can provide insight to the experimentalist, even when the information is not formally extracted by image analysis. Ideally, cellular images are reusable because new image analysis routines that quantify different features of cell images can be used to provide new information about an earlier experiment. With the advent of modern automated instrumentation, vast amounts of image data are being collected in many laboratories. Acquiring, archiving, annotating and distributing the diverse types and vast amount of data generated by these experiments in a federated system are major challenges for the community. To facilitate effective use of data from distributed cell image related research, several infrastructure problems must be solved. Capturing experimental data and metadata in a way that is convenient for experimentalists and yet suitable for seeking succinct answers in use-case based applications is essential. Also needed are domain-dependent user-friendly techniques and tools to support data analysis and concept building in a multi-investigator environment.

We have developed¹ an approach to cell image databasing that uses open source technology such as Oracle, Flex, Excel, and Perl that is extensible for large number of entries and can be maintained in a Federated environment. Flexible, experiment driven, schema-less, extensible vocabulary and ontology are built on 'root' term concepts. A prototype database is constructed in a layered approach in that the metadata and experimental terms are stored independently of the image data in order to allow expansion and flexibility in semantic searching. The schema-less and database-friendly approach to organize metadata terms eliminates the need to commit to specific data formats or software tools. This approach relies on data tables of metadata, which can be semantically organized during a user-defined search query. By assembling metadata terms in a data table, image data can be searched across different datasets and different federated sites. A file-naming rule keeps image data, metadata, protocol, and other files unambiguously linked. In our current implementation, logical operations are performed on metadata terms to identify differences between experimental details of different images series. The 'root'-based approach demonstrated here is compatible with most data models and data-acquisition tools.

Keywords: cell imaging, image database, semantic web, cell image metadata, image data search tools.

¹ http://xpdb.nist.gov/image/cell_image.html