

Open Microscopy and Open Semantics

Eric Neumann, Clinical Semantics Group

The Digital Revolution in research is growing at an ever-increasing pace. Issues of storage and accessibility are being addressed, and will result with unprecedented image data availability for the research community. Efforts such as OME are opening the use of digital microscopy imaging, and allowing academic as well as commercial groups to develop and utilize common data models with various applications. Yet extracting the content and meaning in collections of images is still lagging, not because of image analytics and computation, but rather due to minimal standards in metadata definitions and formatting.

One of the promises of the W3C Semantic Web standards is in representing both data and metadata in powerful and flexible ways that will offer humans as well as machines efficient strategies for mining and enhancing research information. Image analysis especially requires the application of semantics since the general content in an image is still best interpreted by researchers, and therefore is dependent on the knowledge around a specific area of study by the researcher.

While computational tools are excellent at extracting specific features efficiently from images series, the meaning behind a measurement, such as how to interpret “the red fluorescence found in treated and stained cells” requires experimental metadata to be associated to the images. Since the space of all possible biological microscopy experiments is virtually infinite, the metadata structures must be based on expressive languages that can be extended, yet easy to parse with no need for code-revisions. Consequently, they should not be based on coded schema models such as XML-Schema (or DTDs), since this would result in a brittle solution to costly to maintain.

The Triple structure of RDF (Resource Description Framework) is ideal for image metadata as well as image interpretations, and can be based on a combination of controlled vocabulary from microscopy and experimental designs. Additionally, all inferences from a given set of observations can also be linked directly to the images they originated from, for example:

Using in vivo images of Pancreatic cancer cells transfected with DsRed-2 (RFP)-Lentovirus, the pattern and degree of fluorescence is a measure of the spread of cancer metastases in animals. Hence the amount of red-fluorescence protein in a field can be interpreted as degree of metastasis in the focused anatomical region.

The experimental metadata for the images can be stated in RDF N3 code as:

```
<Q35-PCC> <isa> <Collection of Images> .  
<Q35-PCC> <containing> <Tx-PCC> .  
<Tx-PCC> <isa> < Pancreatic cancer cells > ;  
<Tx-PCC> <transduced_by> <RFP-Lent> .
```

<RFP-Lent> <contain> <RFP-Fluor> .

One of many findings (per field) can be described in N3 as:

<Field FX37><amount of fluor> "48.6 ru" .

<Field FX37> <interpretation> <Moderate_Metastasis> .

The use of the RFP, the transducer virus, and the targeted pancreatic cancer cells now all searchable entities when a researcher is looking for images using similar technology or working on similar cancer research. Using Triples, an image or a collection of images can have triples that "link" such metadata using appropriate semantic relations. Anyone accessing the system can extract and navigate around such information, *even if their application has never seen the metadata vocabulary used*. In other words, the system would be completely data interoperable (based on open standards).

If such metadata is shared, the simple structure of RDF allows trivial parsing, and the use of <*> identifiers (URI) for the various triple components allow all definitions and vocabulary to be retrievable and examined over the Internet. Hence new vocabulary can be defined, used, published, and shared by the community without lengthy debates onto their meaning. More practically, a vocabulary can be created as an extension to an already existing common vocabulary. This model is akin to a semi-controlled tagging approach, offering both common semantics and fine-grain extensibility.

Basic strategies of how RDF can be tied to OME standards and structures will be discussed, and a Roadmap to create a core minimal ontology (using the OWL standard) for linking any kind of metadata and data interpretation onto images will also be outlined. This paper recommends the launch of a productive and collaborative open initiative to effectively develop of a general, extensible semantic model based on RDF and OWL for Open Microscopy efforts.