Imaging Ontology, contributing to "*reasonable*" semantics for biomedical image repositories

Joris Slob, Fons J. Verbeek
Imaging and Bioinformatics
LIACS, Leiden University
2333 CA Leiden, The Netherlands
{j.slob, f.j.verbeek}@liacs.leidenuniv.nl

Abstract

Ontologies are required for precise description of image resources in repositories. The Imaging Ontology describes image acquisition instruments as used in the life sciences and complements ontologies that are used for describing the content. We show how we can reason over optical resolution to assist in experiment planning or knowledge discovery. Use of adjacent classes helps the overall structure of the ontology and its extendibility.

1 Introduction

Images are an important output for scientific research. They are used for pure phenomenological description as well as for measurements. In both cases, good description contributes to longevity of the images as well as possibilities to connect to images with similar or complementary content so as to extend on the application domain of the images.

Ontologies terms are a successful vocabulary for annotating data in the life sciences. Using existing ontologies helps with integration and validation of data. Algorithms can assist the user to annotate their data accurately and sufficiently.

A lot of research has been done on the (automated) annotation of biological and medical texts (Rebholz-Schuhmann et al., 2012). Recent studies also included images found in text resources, however the studies have focussed on a single modality or on image captions (Kuhn et al., 2012).

In previous research we have developed LSIDx (Leiden Scientific Image Database for Exchange). The goal of this database is to link its content by means of standardized terms; this requires precise annotation with such terms and these terms are drawn from a range of different ontologies. Consequently much attention has been given to the user interaction so that the expert is well supported in the annotation as well as in the retrieval phase (Bei et al., 2007) (Kallergi et al., 2009). Thus it becomes possible to do reasoning between images and allows extra data to be pulled from external databases. Instead of reconstructing the semantics of images, we propose a method to preserve as much information about the images as possible from the moment of acquisition.

The LSIDx database is a canonical database, it aims at connecting phenomena at different resolutions and from different imaging modalities as well as connect complementary phenomena to one "story" that is told by images. In this manner a virtual microscope is built based on a semantic model (Kallergi et al., 2009). For this purpose custom user interfaces were developed (Dmitrieva et al., 2010) (Kallergi et al., 2013).

Previously in LSIDx the microscope information was recorded using templates that describe common microscope modalities (e.g. confocal microscopy). The templates consisted of multiple attribute-value pairs, that the user could customize before the actual annotation process. This template was extendible to capture more information. This metadata was captured using an Entity-Attribute-Value (EAV) model. While simple queries on modality were available, the system could not offer the flexibility to express queries with aspects of integration and higher levels of complexity and granularity. In this paper we describe what can be gained from moving the microscope acquisition templates from an EAV model to an ontology based system that integrates with the semantic models we have for content.

2 Image Annotation

Within LSIDx the Image Annotation is divided into five categories:

- Dublin Core
- File information
- Content
- Microscope
- External links

Dublin Core metadata (Weibel, 1997) is needed to provide access rights and provenance information. Scientific images are often used as evidence for hypotheses. As such, it should conform to the high standards we come to expect from evidence.

File information metadata is important to ensure computers can parse the images and visualize its content. To check the data integrity checksums and file size are stored. For visualization we have metadata on image geometry and file type.

Content metadata describes the subject matter that is depicted in the image. Content annotation is based on a range of ontology terms that are preselected from 37 different but standard ontologies. Annotation is facilitated by providing semantic context by showing the concept neighbourhood in a graphical user interface.

We changed the original concept of templates, so that microscope information can now be captured using relations from a newly developed Imaging Ontology (cf. section 3) in addition to the ontology terms used for content annotation.

External links provide the images with more context. The annotation model of LSIDx allows including, or linking to, information from standard reference databases, like UniProt (Apweiler et al., 2004) and PDB (Berman et al., 2003).

3 Imaging Ontology

The header file of microscope images holds a great deal of information about the acquisition process. The headers are often stored in proprietary formats and lack standardization across different microscopy techniques. We want to encode our microscopy data using an ontology with the following requirements:

- **Range**: The ontology should include concepts close to the following list: Fluorescence, Confocal, Transmission Electron, Scanning Electron and Atomic Force.
- **Instantiable**: Microscope should be a class that is an endurant, i.e. it should model a physical object. Microscope types should be subclasses of this class.
- **Class Semantics**: All microscope subclasses should have more axioms than just the assertion that it is a subclass of Microscope.
- **Properties**: Ontology should have properties for numerical aperture and (optical) resolution.

In Table 1 we present the results of this checklist for the ontologies we evaluated using BioPortal (Noy et al., 2009).

None of the ontologies had the concept of numerical aperture or a property that can hold this data. In some of the imaging pipelines deconvolution (McNally et al., 1999) is an important step and numerical aperture is critical for this operation. Only two ontologies: FBbi (Orloff et al., 2012) and QIBO (Buckler et al., 2013), had extra semantics to validate or classify microscope subclasses. From our evaluation we concluded that the existing ontologies are not sufficiently generic. We have engineered a generic ontology for microscope acquisition, the Imaging Ontology.

For maintainability and reasoning extra semantics on every subclass of Microscope are needed. These additional semantics require extra classes for physical interactions and imaging modes. The microscopy classes are made equivalent to relations to these additional adjacent classes.

3.1 Adjacent classes

At the heart of imaging techniques is the interaction with the sample that will be measured by some form of detector and the mode of operation (e.g. Bright Field Imaging). These interactions and modes are modelled using additional classes that are not meant to be directly instantiated, but only serve to provide context to the microscope classes. The adjacent classes are linked to appropriate DBPedia articles (Auer et al., 2007).

	Requirements			
Ontology	Range	Instantiable	Class Semantics	Properties
BAO	✓	✓	✗	✗
BIRNLEX	✗	✓	✗	✗
BRO	✗	✓	✗	✗
EP	✗	✓	✗	✗
ERO	✓	✓	✗	✗
FBbi	✓	✗	✓	✗
GALEN	✗	✓	✗	✗
NPO	✗	✓	✗	✗
OBI	✗	✓	✗	✗
QIBO	✓	✓	✓	✗
RH-MESH	✓	✗	✗	✗

Table 1: Existing ontologies and their results on our Checklist. BAO = BioAssay Ontology, BIRNLEX = Biomedical Informatics Research Network Project Lexicon, BRO = Biomedical Resource Ontology, EP = Cardiac Electrophysiology Ontology, ERO = Eagle-I Research Resource Ontology, FBbi = Biological Imaging Methods Ontology, NPO = NanoParticle Ontology, OBI = Ontology for Biomedical Investigations, QIBO = Quantitative Imaging Biomarker Ontology and RH-MESH = Robert Hoehndorf Version of MeSH.

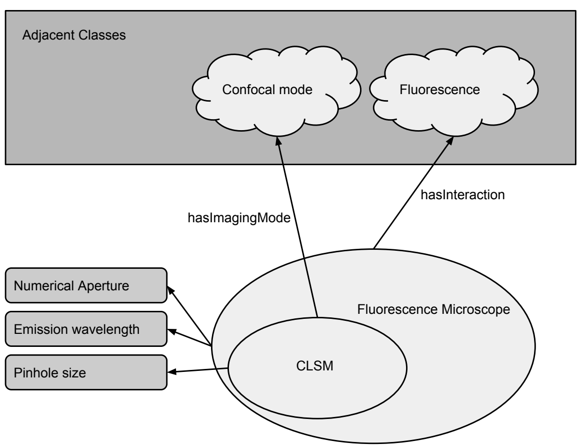

Figure 1: Relation of the classes within the Imaging Ontology

The hierarchical structure of the Imaging Ontology can be inferred from the relations the microscopes have with the adjacent classes. See Figure 1 for an example.

With these semantics a reasoner can assist the user to infer the microscope type from the evidence it gathers from different means, like:

- Pixel information
- Header information
- Content annotations from the user

3.2 Relation to domain ontologies

In LSIDx we aim at storing images of the same phenomenon in multiple modalities with varying resolutions. Of special importance in the Imaging Ontology is the (optical) resolution of the imaging device. The resolution allows the computer to reason about the biological structures that might be visible in an image. This is illustrated in Figure 2. The units have to be represented using an ontology of units (Rijgersberg et al., 2013), or the reasoner has to be able to cope with the different representations of the has_optical_resolution property.

If the domain ontology lacks information on the general size of its physical objects, then we can at least get an ordering using the mereology that most ontologies have. For these ontologies some extra facts have to be asserted to allow the reasoner to validate what can be observed with what microscope.

3.3 Query possibilities

See Figure 3 and Figure 4 for illustrations of queries with the Image Ontology and resulting image sets. The user interface depends on the application at hand.

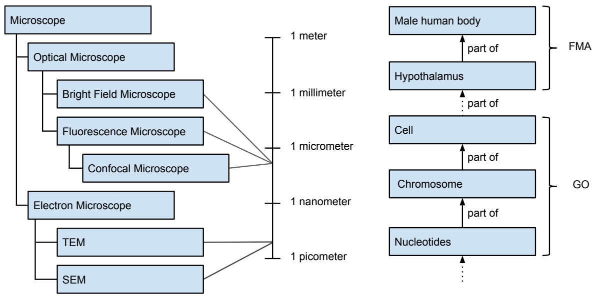

Figure 2: Imaging Ontology relation to domain ontologies

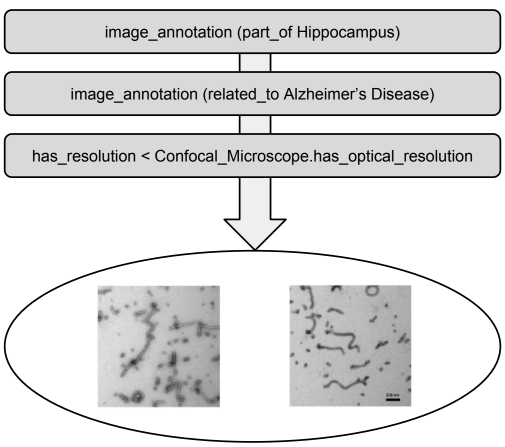

Figure 3: Example resultset from query 1

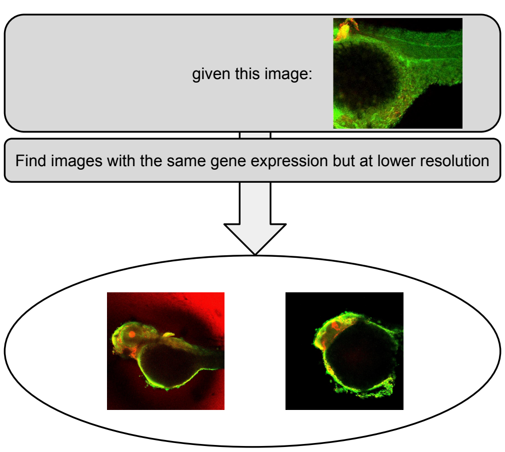

Figure 4: Example resultset from query 2

4 Results & Discussion

In this paper we have explained the need for an imaging ontology to complement image annotation. We present the Imaging Ontology and demonstrate the possibilities such an ontology can offer to a biomedical expert. The ontology requires constant curation as new imaging techniques and improvements to techniques appear. With the addition of interaction subclasses we have provided a framework that will help the ontology to grow naturally while remaining internally consistent.

We have included our ontology to the canonical database LSIDx, but the same ontological infrastructure is used for a dedicated application database CytomicsDB (Larios et al., 2012) (Larios et al., 2014) in which the focus is put on the management of large volumes of image time-lapse data and their analysis.

The next step we are working on are personalized agents that use the information from the ontologies we have developed to help annotate, retrieve and connect data.

5 Conclusions

The Imaging Ontology is a valuable asset in an image knowledge base. It can help validate, assist annotation and enhance retrieval of images. The ontology also increases the internal connectivity of the data, creating a richer dataset. Using adjacent terms helped structuring the ontology and making it easier to extend.

Acknowledgements

This work was funded by Cyttron Consortium II (LSH framework: FES0908)

References

Rolf Apweiler, Amos Bairoch, Cathy H. Wu, Winona C. Barker, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Hongzhan Huang, Rodrigo Lopez, Michele Magrane, Maria J. Martin, Darren A. Natale, Claire O'Donovan, Nicole Redaschi and Lai-Su L. Yeh 2004. UniProt: the universal protein knowledgebase *Nucleic acids research*, 32 (Suppl 1): D115-D119.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak and Zachary Ives 2007.

Dbpedia: A nucleus for a web of open data *The Semantic Web*, (Vol 4825) Lecture Notes in Computer Science, Springer Berlin Heidelberg.

Yun Bei, Julia Dmitrieva, Mounia Belmamoune and Fons J. Verbeek. 2007. Ontology driven image search engine *Electronic Imaging*, International Society for Optics and Photonics

Helen Berman, Kim Henrick, and Haruki Nakamura. 2003. Announcing the worldwide protein data bank *Nature Structural & Molecular Biology* 10.12: 980-980.

Andrew J. Buckler, M. Ouellette, J. Danagoulian, G. Wernsing, Tiffany Ting Liu, Erica Savig, Baris E. Suzek, Daniel L. Rubin, David Paik 2013. Quantitative Imaging Biomarker Ontology (QIBO) for Knowledge Representation of Biomedical Imaging Biomarkers *Journal of digital imaging* 26.4: 630-641.

Julia Dmitrieva and Fons J. Verbeek. 2010. Different geometries in ontology visualization *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics

Amalia Kallergi, Yun Bei, Peter Kok, Jouke Dijkstra, Jan-Pieter Abrahams, Fons J. Verbeek 2009. Cyttron: A Virtualized Microscope supporting image integration and knowledge discovery *Proteins Killing Tumour Cells*, Cell Death and Disease Series (Eds.Backendorf,Noteborn,Tavassoli): ResearchSignPost, Kerala India, 291-315

Amalia Kallergi and Fons J. Verbeek. 2009. OntoFrogger: Playing with Semantic Structure *Proceedings Workshop on Semantic Web Applications and Tools for Life Sciences (SWAT4LS)* CEUR-Workshop Proceedings ISSN 1613-0073, (Vol 559) (Eds. Marshall, S.M.,Burger A.,Romano P.,Paschke A.,Splendiani A.) Amsterdam, The Netherlands, November 20

Amalia Kallergi and Fons J. Verbeek 2013. Playfull Interfaces for Scientific Image data *Emerging Research and Trends in Interactivity and the Human-Computer Interface* (Eds.Katherine Blashki & Pedro Isaias) IGI Global, Hershey PA, USA. pp. 471-489

Tobias Kuhn and Michael Krauthammer 2012. Image Mining from Gel Diagrams in Biomedical Publications *Proceedings of the 5th International Symposium on Semantic Mining in Biomedicine (SMBM 2012)*. Edited by: Ananiadou, Sophia; Pyysalo, Sampo; Rebholz-Schuhmann, Dietrich; Rinaldi, Fabio; Salakoski, Tapio. Zurich. ISBN 978-3-033-03823-3.

James G. McNally, Tatiana Karpova, John Cooper, José Angel Conchello 1999. Three-dimensional imaging by deconvolution microscopy *Methods* 19.3: 373-385.

Natalya F. Noy, Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L. Rubin, Margaret-Anne Storey, Christopher G. Chute and Mark A. Musen 2009. BioPortal: ontologies and integrated data resources at the click of a mouse *Nucleic acids research* 37. (Suppl 2): W170-W173.

David N. Orloff, Janet H. Iwasa, Maryann E. Martone, Mark H. Ellisman and Caroline M. Kane 2012. The cell: an image library-CCDB: a curated repository of microscopy data *Nucleic acids research*: gks1257.

Dietrich Rebholz-Schuhmann, Anika Oellrich and Robert Hoehndorf 2012. Text-mining solutions for biomedical research: enabling integrative biology *Nature Reviews Genetics* 13.12: 829-839.

Hajo Rijgersberg, Mark van Assem, and Jan Top. 2013. Ontology of units of measure and related concepts *Semantic Web* 4.1: 3-13.

Enrique Larios Vargas, Ying Zhang, Lu Cao and Fons J. Verbeek. 2014. CytomicsDB: A Metadata-Based Storage and Retrieval Approach for High-Throughput Screening Experiments *Proceedings PRIB 2014* (M. Comin et al. (Eds.)). Springer Int. Publ. Switzerland; LNBI 8626, pp. 72-84, (In Press)

Enrique Larios Vargas, Ying Zhang, Kuan Yan, Zi Di, Sylvia E. Le Dévédec, Fabian Groffen and Fons J. Verbeek. 2012. Automation in Cytomics: A Modern RDBMS Based Platform for Image Analysis and Management in High-Throughput Screening Experiments *Proceedings of 1st Int. Conf. on Health Information Science* (Eds. Jing He, Xiaohui Liu, Elizabeth A. Krupinski, Guandong Xu). Beijing, China April. Springer (Heidelberg-Berlin), LNCS Vol. 7231, 76-87

Stuart Weibel 1997. The Dublin Core: a simple content description model for electronic resources *Bulletin of the American Society for Information Science and Technology* 24.1: 9-11.