Connecting the dots : playful interaction with scientific image data in repositories

Kallergi, A.

Chapter 2

The Cyttron Scientific Image Database for eXchange

Based on:

A. Kallergi, Y. Bei, P. Kok, J. Dijkstra, J. P. Abrahams, and F. J. Verbeek. Cyttron: A virtualized microscope supporting image integration and knowledge discovery. In C. Backendorf, M. Noteborn, and M. Tavassoli, editors, Proteins killing tumour cells, Cell Death and Disease Series, pages 291–315. ResearchSignPost, 2009a

A. Kallergi, Y. Bei, and F. J. Verbeek. The ontology viewer: Facilitating image annotation with ontology terms in the CSIDx imaging database. In Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009), February 2009b

Abstract: This chapter introduces the Cyttron Scientific Image Database for eXchange (CSIDx), a multi-modal imaging database for images produced in the life sciences. The database was developed within the framework of the Cyttron project, a consortium intended to create an integrated infrastructure for bio-imaging. As the CSIDx database and community provide the use case of this thesis, the aims of both the Cyttron and the CSIDx initiatives are relevant. Moreover, some familiarity with the data involved in this thesis and their organization in the repository should be beneficial for the reader. Of particular interest is the extensive use of ontology annotations for the management of image data: Semantic annotation is proposed as a requirement for the management, sharing and integration of images. Ultimately, and with respect to the notion of playful interaction, we elaborate on information visualization features introduced in the CSIDx interface in order to better exemplify the CSIDx data and collection.

2.1 Introduction

"An image is worth a thousand words". Surely, but does this conventional wisdom apply to scientific images produced in the life science as well? In one hand, we need to consider whether biological images are straightforward enough to speak for themselves. On the other hand, we need to consider if a single image is enough to tell the whole story. And finally, we need to ask ourselves what the point of a thousand words is, if none is listening anyway. The answers to these seemingly innocent questions have implications on how we envision and design digital systems for the management of scientific image data in biology and the life sciences.

Imaging is an indispensable practice for biological research. With researchers producing proliferating numbers of image data, data management soon becomes an issue. What is more, there is a wealth of image research material that could benefit other researchers but is never shared. Assuming that image provenance is well-tended, the benefits from sharing one another's "treasure chests of images" (Marx, 2002) should be self-explanatory. That said, delivering systems to support image-based biological research, systems that will better the management, sharing, remote access, interoperability and integration of image data, as advocated by Swedlow et al. (2009), is not without its challenges. Consider, for example, the content of biological images: The subject depicted is rarely evident and may depend on the experiment design (e.g. which gene is being tagged). To quote Goldberg et al. (2005), "deriving information from images is completely dependent on contextual information that may vary from experiment to experiment". A system or repository will need to assure sufficient metadata, i.e. data about the data, for the stored images to be of any use. Furthermore, well-described content is not sufficient without utility. The system will need to provide useful services and user-friendly interfaces for the researcher to benefit from the maintained images. To put it differently, we need good data to be put in good use by both machines and humans; the latter, we argue, can be empowered by means of an interface. Well-designed interfaces, effective displays and interactive tools can empower the human researcher by both easing her tasks and amplifying her cognitive capacities. During the development of the CSIDx image database, we soon realized that, one, an image can not "speak for itself" and, two, we need to pay attention to a neglected aspect of scientific management systems, i.e. the user interface.

Image-based biological research is often multi-modal, i.e. involving various imaging modalities. The study of a biological phenomenon frequently requires combining observations of various levels of abstraction and such observations

will be recorded by instruments operating at various levels of resolution. Multimodal imaging is typically performed on the same biological sample using imaging instruments that support multiple imaging techniques (Bonnet, 2004). Other comparative microscopic studies such as imaging in both 2D and 3D dimensions, i.e. multi-dimensional imaging (Bonnet, 2004), are also frequent. Eventually and to fully comprehend a biological phenomenon, we need contributions from various domains of biology, operating at the organism/cell/protein/molecular level; such contributions may not always be derived from the same sample. Consider as an illustration the concept of apoptosis, i.e. programmed cell death. The phenomenon can be studied at the cell level, by examining the phenotype of an apoptotic cell, at the protein level, by locating proteins responsible for apoptosis and at the molecular level, by identifying the structure of the proteins involved. The imaging techniques employed could be light microscopy, electron microscopy and crystallography respectively. In short, a single image/imaging technique is not always enough to tell the whole story: Biological research needs an integration of observations/images from the cellular level down to the molecular level. This realization lies at the heart of the Cyttron project, the host project underlying the development of the CSIDx database.

The structure of this chapter is as follows: Sessions 2.2 and 2.3 present the aims of the Cyttron and CSIDx initiatives respectively, and summarize their visions about biological research practice, data management and data integration. The Cyttron context and the nature of the CSIDx data have impregnated our understanding of what useful interactions with images may be. Section 2.4 describes in depth the CSIDx system and web interface, paying particular attention to the process of image annotation. Having established our context and background, we shift our attention to the capacities of the interface as a means to better understand and further explore the CSIDx data. In particular, section 2.5 elaborates on information visualization principles introduced in CSIDx in order to improve on some of the user tasks and foster a better understanding of the CSIDx data and collection.

2.2 Aims and scope of the Cyttron project

The Cyttron consortium was a publicly funded project initialized in 2004. The project brought together a consortium of academic and industrial partners with the goal to implement an integrated infrastructure for bio-imaging. Such infrastructure could facilitate biological research by providing a generic platform for studying and modelling biological phenomena over various resolutions, i.e. from

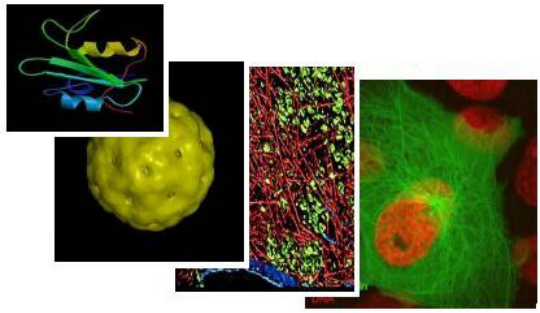


Figure 2.1: The Cyttron visual rhetoric: 'Zooming in' the cell. From right to left, a cell with a tagged protein, the ultrastructure of a cell, a virus molecule and a protein. Implied in the visual is the ability to move across imaging modalities, from the micro scale down to the atomic level.

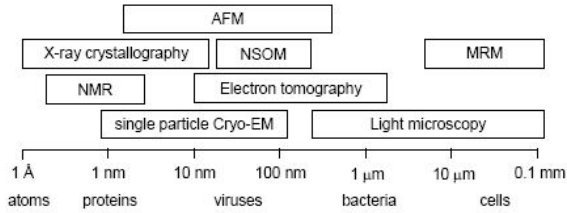


Figure 2.2: Overview of imaging modalities addressed in the Cyttron platform. Each modality is represented by a rectangle with its specific range of resolution and overlap with other modalities.

the cell level down to the molecular level. As said, implicit in the project is the realization that the study of a biological phenomenon often spans across multiple imaging techniques. Table 2.1 summarizes some of the modalities that were deemed relevant for the Cyttron platform based on their relevance for contemporary life science practice. From the perspective of the user, an integration of images depicting a phenomenon at different levels of detail is highly desirable. During Cyttron meetings and symposia, the metaphor of a virtualized microscope that would allow 'zooming in' at various resolution levels was developed. Figure 2.1 became iconic of the aims of the project.

Being an interdisciplinary consortium, the Cyttron project addressed the challenge of image integration from a multiplicity of approaches. As shown in Figure 2.2, imaging modalities often overlap in resolving power. By means of image registration, this overlap in resolution can be exploited to establish connections between images of the same sample. A significant part of the Cyttron project engaged in bettering the overlap between microscopy techniques by extending the resolution, range or throughput of existing technologies. This approach was particularly prominent in partners involved with hardware and/or imaging protocols.

Another approach, shared by the software groups of the project, focused on the development of a common software platform that would provide seamless and continuous access to diverse image data. The notion of a common platform was conceived at both the visualization level (images should be visualized in the same viewer) and the data model level (data about the images should comply to the same data model). As a result, the common visualization platform was implemented as two distinct but complimentary and co-operating software products, namely the CVP (Common Visualization Platform) viewer and the CSIDx image database. In effect, the integration of images was attempted at both the pixel and the metadata level.

2.3 Aims and scope of the CSIDx system

The Cyttron Scientific Image Database for Exchange (CSIDx) is a multi-modal imaging database for images produced in the life sciences. CSIDx was built to be the backbone repository of the Cyttron project; as such, CSIDx was designed to support a wide range of imaging modalities and techniques, as well as a variety of users and research practices. With respect to the aims of Cyttron, CSIDx proposes an integration of images based on their metadata and semantic content. Multi-modality and semantic image annotation are the two core requirements for CSIDx.

CSIDx was also designed to be a web-based resource and community for researchers from various institutes to share image resources. It is highly unlikely that the same scientist or research group can produce data across the full range of existing modalities. Therefore, a seamless integration across the resolution range is only possible via principles of sharing and collaboration. CSIDx is a manifestation of a community of scientists wanting to learn from connections established across one another's images.

2.3.1 Multi-modality

From the impressive number of databases publishing data from the life sciences[1], only a small portion provide microscopy images. Table 2.2 enlists several examples of biological image database systems; the list is not exhaustive but is a compilation of systems and repositories discussed by Walter et al. (2010); Linkert et al. (2010); Swedlow et al. (2009); Lindek et al. (2006). Genomics databases

[1] Consider as an illustration the 'Nucleic Acids Research' inventory of molecular biology databases containing 1330 resources (Galperin and Cochrane, 2011).

Table 2.1: Some of the imaging modalities addressed in the Cyttron platform.

Modality	Description	Level of Resolution
Magnetic Resonance Microscopy (MRM)	Can potentially visualize the interior of a cell inside an organism, permitting *in vivo* imaging.	small organisms, organs, cells (up to 10 μm)
Light Microscopy	Can visualize live cells and can trace the movements of single molecules through a cell, provided that they carry a fluorescent label. Three- and four dimensional imaging allows capturing dynamic information.	tissues, cells (typically around 1 to 10 μm)
Electron Tomography	Allows visualization of cellular ultrastructure in 3D. Cells need to be fixed, so dynamic information must be obtained via comparative studies.	cells, organelles/subcellular structures (typically around 0.5 to 10 nm)
Atomic (Scanning) Force Microscopy (AFM)	Can visualize non-crystalline proteins and protein complexes. It can potentially image cellular surfaces and may be used to visualize live cells.	subcellular structures, bacteria, ribosomes, protein complexes (up to 1nm)
Single particle cryo-EM	Can visualize very large bio-molecular complexes, not necessarily crystalline or purified. Samples are frozen at (at least) liquid nitrogen temperatures, therefore dynamic information cannot be obtained from single molecules but must be inferred via comparative studies.	proteins (up to 1 nm)
Nuclear Magnetic Resonance (NMR)	Can visualize non-crystalline proteins, in purified form and labelled with stable isotopes with a limited size range. Provides dynamic information about the mobility of the protein and interactions between proteins.	membranes, proteins, atoms
X-ray crystallography	Can determine atomic structure but the sample must be crystalline, and so the throughput of the technique is limited.	proteins, atoms (up to 1 Å)

often provide fluorescence images to verify or illustrate gene expression data, while proteomics databases frequently provide visualizations, i.e. models of protein structure, to render their content. In both cases, however, one can argue that the unit exchanged, i.e. stored, annotated, queried and retrieved, is the genomics or proteomics data rather than the microscopy images or models. Resources dedicated to model organisms also provide microscopy images on the anatomy, phenotype and gene expression of the organism under study. Finally, when microscopy images are the focus of the repository, the system is often originally oriented towards a particular imaging technique or a particular field of biological research. CCDB (cf. Table 2.2) was initially designed for electron microscopy data but is now expanded to various light microscopy techniques. A couple of projects, namely Bioimage and SIDB (cf. Table 2.2), were originally designed to host images from various modalities. Table 2.2 also includes two significant software initiatives, namely OME and Bisque, that do not host research data but deliver open source platforms for image management.

Obviously, the design requirements of CSIDx dictate that a wide range of imaging techniques is supported. Supporting a multiplicity of imaging techniques is a challenge that manifests at various levels. On the file format level, the system needs to cater for a variety of image formats with often proprietary characteristics. Regarding the data model, some flexibility in metadata must be afforded as microscopy techniques vary considerably in their material and methods. Finally, the interaction design needs to accommodate users with diverse methodologies, workflows and requirements.

2.3.2 Semantic image annotation

In the core of CSIDx is the need for an extensive annotation of the image data. Thorough and unambiguous annotation is proposed as both the means to link modalities and a necessary requirement for image sharing and reuse. It should be repeated here that the content of biological images is rarely self-explanatory but must be supplied by the image producer or by researchers in the corresponding research field; lack of metadata will simply result in meaningless and unusable data. Linking images will also require metadata. To begin with, linking images by exploiting the resolution overlap between modalities will require metadata about the production of an image. Depending on the image format, microscopy settings may be available as header information but information about the experiment design, e.g. sample preparation, must be provided by the producer of the image. Most importantly, CSIDx proposes an integration of images via biological

Table 2.2: Selected list of image databases developed for the life sciences as compiled from Walter et al. (2010); Linkert et al. (2010); Swedlow et al. (2009); Lindek et al. (2006).

Name	URL	Short Description	Status
BDGP insitu	http://insitu.fruitfly.org	Patterns of gene expression in Drosophila embryogenesis, generated by high-throughput RNA in situ hybridization.	Active
4dxpress	http://ani.embl.de/4DXpress	Gene expression data acquired through whole mount in situ hybridisation.	Active
EMDatabank	http://emdatabank.org/	Global deposition and retrieval network for cryoEM map, model and associated metadata.	Active
Allen Brain Atlas	http://www.brain-map.org	Gene expression and neuroanatomical data.	Active
The e-Mouse Atlas Project	http://www.emouseatlas.org	EMA provides histology sections and 3D reconstructions of mouse embryos, EMAP provides gene expressions generated by various in situ assays.	Active
Flybase	http://flybase.org	Genetic and genomic data on Drosophila melanogaster and the insect family Drosophilidae, provides graphics and microscopy images to illustrate anatomy and development.	Active
zfin	http://zfin.org	Zebrafish genetic, genomic, phenotypic and developmental data, compiles images of anatomical structures, phenotype images and gene expressions from various assays.	Active
Protein Subcellular Location Image Database (PSLID)	http://murphylab.web.cmu.edu/services/PSLID	2D to 5D fluorescence microscopy images depicting subcellular location proteins, focus is on image processing.	Unspecified
Mitocheck	http://www.mitocheck.org	Regulation of mitosis by phosphorylation, high-throughput live cell imaging of a genome-wide RNA interference (RNAi) screen, focus is on image processing.	Ended
The Cell Centered Database (CCDB)	http://ccdb.ucsd.edu	2D, 3D and 4D data from light and electron microscopy, including correlated imaging.	Active
Bioimage	-	Multidimensional microscopic images of biological samples, broad scope.	Ended, offline
Scientific Image DataBase (SIDB)	http://sidb.sourceforge.net	2D, 3D images, time series, broad scope.	Ended
JCB data viewer	http://jcb-dataviewer.rupress.org	Visualization tool for original image data files associated with JCB (Journal of Cell Biology) articles, various image formats, mostly light microscopy, but also autoradiographs, (scans of) immunoblots, and histology and electron micrographs.	Active
Bisque	http://www.bioimage.ucsb.edu/bisque	Open source platform for data management in the life sciences.	Active
Open Microscopy Environment (OME)	http://www.openmicroscopy.org	Open source platform for data management in the life sciences.	Active

concepts, i.e. based on the phenomenon or sample under study. To that end, consistent and accurate metadata on the content of an image is essential.

To capture the content and interpretation of an image, CSIDx utilizes ontology terms from numerous ontologies relevant to the domain of the life sciences. Ontologies are formal specifications of a domain, modelling the domain concepts and relations, and have description logic underpinnings. For our present discussion, it suffices to describe ontologies as collections of concepts and their relations. Used for annotation instead of user generated keywords or tags, ontologies provide a consistent and unambiguous vocabulary across images, modalities and researchers. Domain specific, curated ontologies capture the nomenclature of the particular field. Finally, the relations between concepts can be further exploited to organize and link the annotated image data. The initiators of CSIDx envisioned a semantically enhanced integration of images achieved by reasoning over the ontologies. As a matter of fact, one of the solutions envisioned would output a graph of images that are linked together not only explicitly, based on the assigned metadata, but also implicitly, based on derived relationships (Bei et al., 2007).

2.4 The CSIDx system

The CSIDx system consists of a relational database management system to store images and their metadata, a relational database management system to store ontologies and a web interface for registered users to access and manage their images. The system is developed, maintained and hosted by the Imaging and Bioinformatics group in Leiden University.

2.4.1 Data management in CSIDx

An image entry in CSIDx is a composite entity of both raw pixel data and user generated metadata (annotations). Bei et al. (2006) put forth the requirements for a sufficient annotation that respects the particularities of scientific image production. It is useful to repeat here that, as regards biological image data, a sufficient annotation should record the following aspects of an imaging experiment:

- *Who* did the experiment, including information about the image owner and the research facility the image was produced in.
- *What* the experiment is about, including information about the biological phenomenon the image depicts or attempts to study.
- *How* the experiment was done, including information about the instruments utilized and the experiment design.

In response, CSIDx stores the following categories of metadata:

- Administrative metadata, capturing information about the owner of the image and the access rights to an image entry (the *who*). Users are organized under research groups and access to resources can be granted group-wide or community-wide.
- Ontology metadata, capturing information about the content of an image entry (the *what*). To safeguard accuracy, ontology metadata are expressed in ontology terms derived from life science related ontologies. The system incorporates 37 ontologies, the majority of which were retrieved from the Open Biomedical Ontologies (OBO) Foundry (Smith et al., 2007).
- Microscope metadata, capturing information about the experiment setup (the *how*). To maximize flexibility, microscope metadata are stored as pairs of settings-values. Generic microscope templates, i.e. lists of representative fields for an instrument or imaging technique, are provided to guide the annotation. For example, the confocal microscope template includes fields such as laser line, laser type, emission wavelength, objective lens and others. Groups can create their own group-specific microscope templates to be used instead of or in conjunction with the generic templates.
- Image metadata, capturing basic features of the image file e.g. file size. Such metadata are usually extracted automatically.
- External links, maintaining links to existing pieces of knowledge in external resources[2].

An entry from the CSIDx database, illustrating the various metadata stored together with the raw pixel data, can be seen in Figure 2.3. As thoroughly explained in subsection 2.3.2, semantic annotation by ontology terms derived from life science related ontologies lies at the core of CSIDx and our data management approach.

2.4.2 The CSIDx web interface

The CSIDx web interface allows registered users to access and manage their images. The interface supports annotation of one's own images, administration of resources, such as terms and templates, and search and retrieval of all visible images. Kallergi et al. (2009a) provide an extensive description of the CSIDx web interface; this subsection will only briefly discuss the nuts and bolts of the CSIDx web interface regarding the image annotation process.

[2]While not part of the initial specification by Bei et al. (2006), external links were suggested by subsequent user studies.

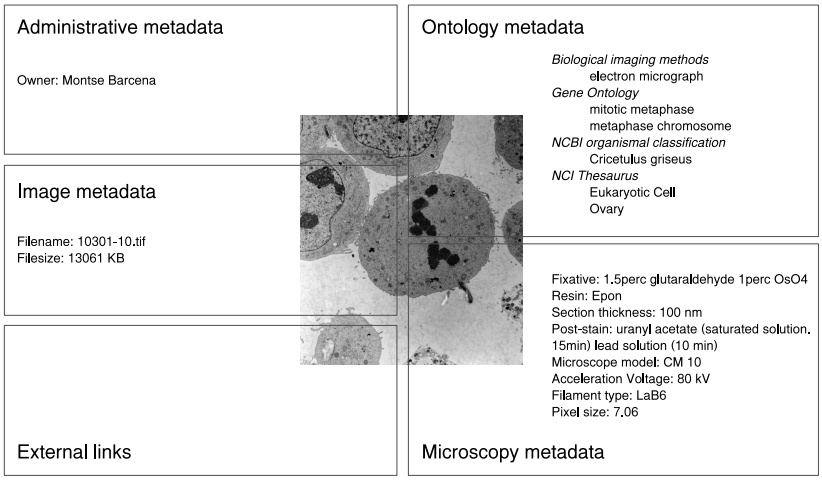


Figure 2.3: Image entry in CSIDx. Image entries in CSIDx are composite entities of both raw pixel data and user generated metadata (annotations). Image courtesy of M. Barcena, Department of Molecular Cell Biology (MCB), section Electron Microscopy, Leiden University Medical Center (LUMC).

Image annotation is a central task in CSIDx. An overview of tasks involved in the workflow of image annotation is given in Figure 2.4. In a simple use case scenario, the user uploads her images and annotates them with ontology terms she collected beforehand ('My terms') and with (generic and/or specific) microscope templates. Some of the graphical user interface (GUI) elements used during an annotation are illustrated in Figure 2.5. Generic microscope templates are curated centrally while group specific microscope templates are to be maintained by the group administrator; both types of microscope templates should be readily available to the user during annotation. Annotation by ontology terms utilizes terms out of a subset of terms the researcher has compiled beforehand ('My terms'); once such as set is compiled, it is available to the user for all subsequent annotation sessions.

Thorough image annotation by a human expert is a time-consuming and demanding task. In order to reduce overhead, the CSIDx interface provides additional options such as the possibility to annotate similar images simultaneously and the possibility to store an annotation into an annotation template to be applied on subsequent images (cf. Figure 2.4). Specific to the annotation with ontology terms is the above-mentioned strategy of 'My terms', i.e. the preselected collection of terms that are of interest to the researcher. Introducing ontologies for annotation was not without its challenges for our users and coping with the vast amount of available terms was one of them (Kallergi et al., 2009b). The strategy of 'My

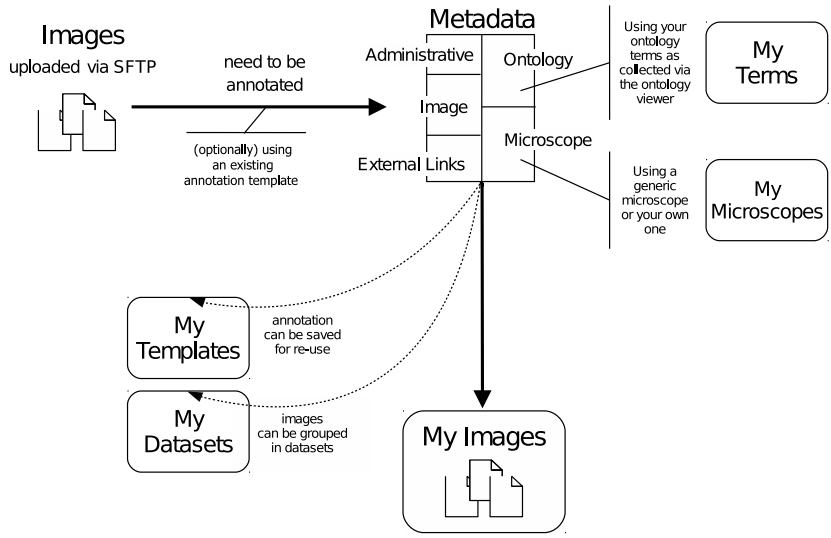


Figure 2.4: Overview of the tasks involved in image annotation.

terms' requires a preparation step but, in effect, minimizes the effort of searching and identifying relevant terms (search once, use in all subsequent annotations). It also reduces the overwhelming amount of ontology terms to a meaningful and manageable subset. Collected ontology terms can be shared among members of a group who are likely to engage with the same research subject. On the whole, effort has been expended that a proper annotation is reused whenever possible.

It should be remarked that the development of the CSIDx interface has been a dynamic and evolutionary process. New features were implemented in collaboration with users and via case studies on biological topics suggested by the Cyttron community. For obvious reasons, a major part of our development effort was dedicated to improving the process of image annotation. We are thus particularly aware of the challenge of acquiring quality metadata in a user-friendly way. That said, the remainder of this thesis will, for the most part, consider semantically annotated images as a given. In a sense, we will shift attention from data acquisition to data access and 'consumption'.

2.5 Visualization for interaction

As the backbone repository of Cyttron, CSIDx is also meant to provide data access to various software clients. For example, the CVP viewer retrieves data and metadata from CSIDx to be further visualized in a dedicated visualization environment.

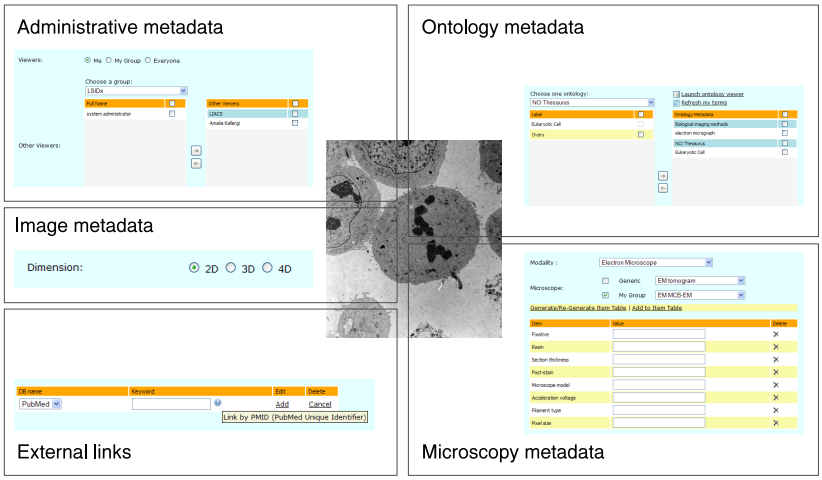


Figure 2.5: Various GUI elements used in the CSIDx web interface for submitting the various types of metadata during an annotation. Image courtesy of M. Barcena, Department of Molecular Cell Biology (MCB), section Electron Microscopy, Leiden University Medical Center (LUMC).

Potentially useful algorithms or services, e.g. image convulsion, feature extraction etc., could be executed on the available data. Seen from this perspective, the web interface is yet another client using the CSIDx database and one that ensures that well-annotated data are available for other services.

Of course, there is more to be supported in an interface than only populating the database. Effective displays and interactive tools can engage the human researcher in examining and exploring the available data. The remainder of this thesis investigates these ideas in depth but efforts to support exploration have already been attempted within the web interface. More specifically, we consider and apply ideas from the field of information visualization to better exemplify our metadata or to provide an overview of the collection. Information visualization is a well-established field exploiting the visual capacities of humans in order to reinforce understanding of data. Note that in this context, visualization refers to the visualization of the metadata of the images and not of the raw data.

2.5.1 Visualization of ontologies

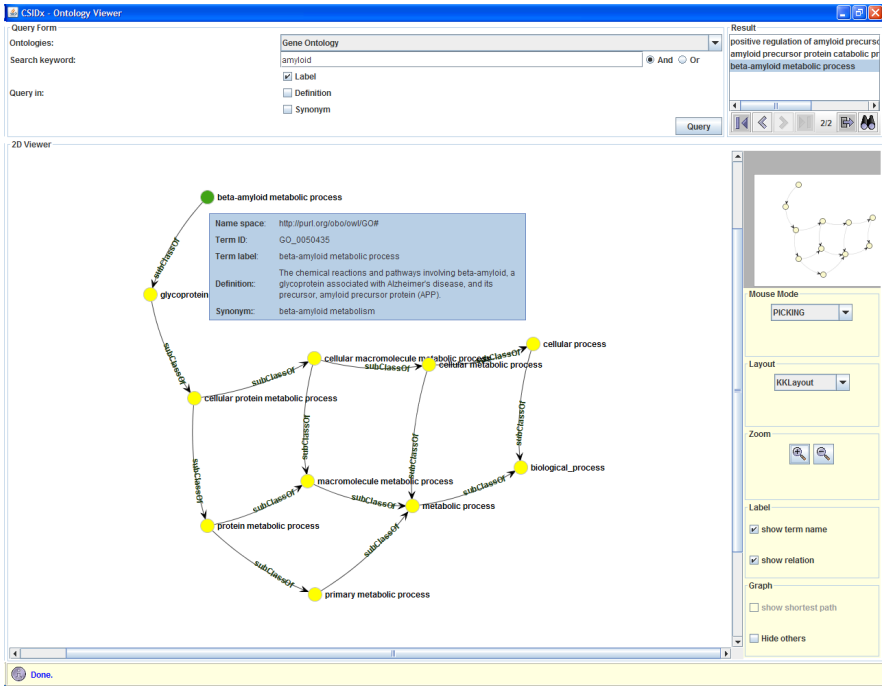
Visualization of ontologies was introduced in the CSIDx system as a means to assist the process of image annotation with ontology terms. While part of the annotation process, the graphical tool implemented is worth mentioning for it provides a visual aid on a core aspect of CSIDx, i.e. the ontology metadata. Noticeably, introducing ontologies for image annotation was a considerable challenge

for our users who were hindered from unfamiliarity with ontologies and from the overwhelming amount of ontology terms. Kallergi et al. (2009b) fully document the difficulties observed and the actions taken in response. Here, we focus our discussion on the employment of a graph visualization to better exemplify ontology metadata.

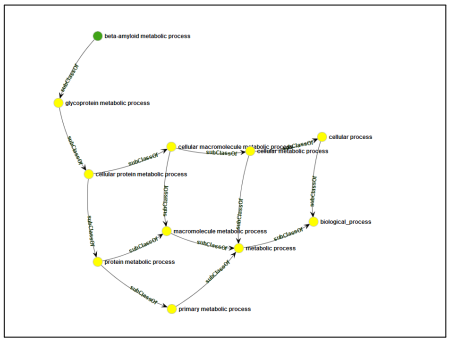
The ontology viewer (cf. Figure 2.6) is a web application that accesses the ontology database of the CSIDx system. It consists of a query form for querying ontology terms and a 2D viewer for visualizing ontology structure. In the 2D viewer, the ontology structure is shown as a graph: Terms are graph nodes and relations are graph edges. Selected ontology terms, as collected from a query, are used to extract a sub-graph of the ontology graph. This sub-graph provides the local context for the selected nodes which are highlighted green to distinguish from their connected terms. A short description with information on any given term can be obtained by hovering over the corresponding graph node. Regular graphical manipulations are supported and the ontology graph can be zoomed, panned, rotated and sheared. For optimal spatial arrangement, the user can switch between a number of different graph layouts. The graph drawing and manipulation was implemented by means of the Java Universal Network/Graph (JUNG) framework (O'Madadhain et al., 2005). The ontology viewer is developed as a Java Web Start application and is available to registered users of the CSIDx database.

For most CSIDx users, the ontology viewer was their first impression on biological ontologies. As such, the ontology viewer functions as an introductory tool for users to familiarizing themselves with the concept, structure and content of ontologies. The tool was reported to improve the mapping of expert knowledge to ontology terms by eliminating ambiguities and by enriching the vocabulary of the user. What is more, the ontology viewer made the concept of ontologies and of semantic annotation concrete and explicit to our users. The graph representation, although an over-simplification from a formalist's perspective, sufficiently communicated the added value of ontologies, i.e. semantic relations. In a sense, and by making the connection between concepts explicit by visualization, the vision of CSIDx as a semantically rich repository was better comprehended and valued.

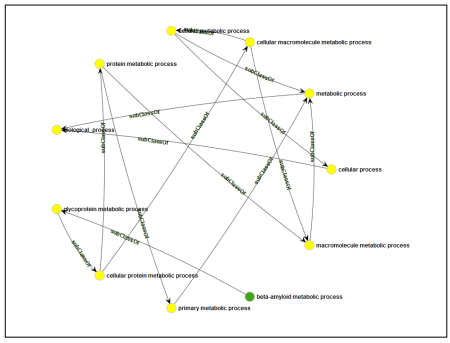
As users familiarize themselves with the ontology structure, they express the wish to further interact with the ontology graph. Compared to a full-scale visualization tool, the ontology viewer will be found lacking in functionality; after all, the tool was initially conceived to provide some context for the queried ontology terms. Nonetheless, the educational potential of the tool is of considerable interest. Our



(a) The ontology viewer interface



(b) KKLayout



(c) CircleLayout

Figure 2.6: The ontology viewer. The query form (Figure 2.6a, top panel) is used to search for ontology terms which can be then visualized in the 2D viewer. Here the subgraph of 'beta-amyloid metabolic process' (Gene Ontology). Figures 2.6b, 2.6c show two different renderings of the same subgraph in the KKLayout and CircleLayout respectively.

research group has continued investigation on ontology visualization resulting in full-featured visualization tools for ontologies (Dmitrieva, 2011).

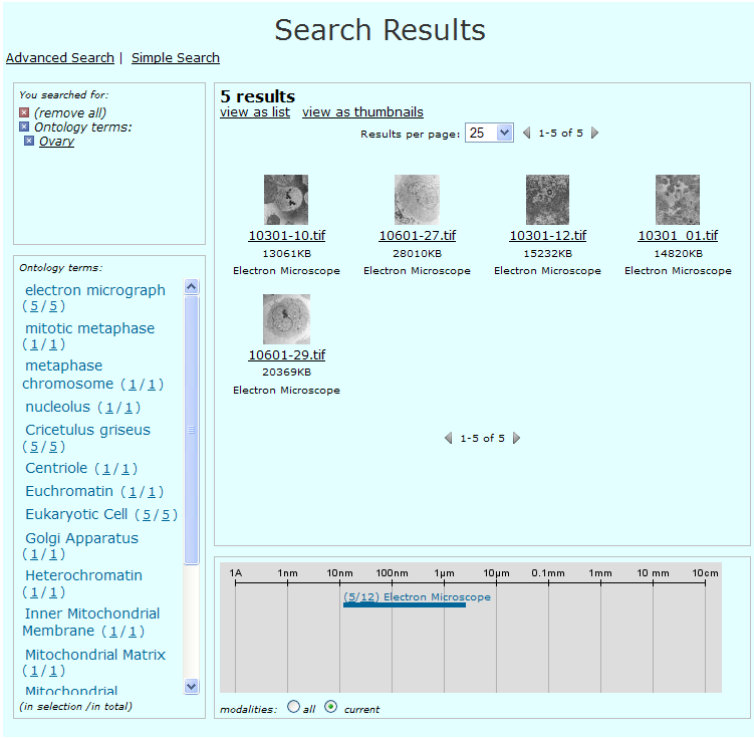
2.5.2 Visualization of search results

From plain lists of hits to more elaborate visualizations, the presentation of search results is a central component of any search interface to a data repository. By means of visual encodings, visualizations can potentially provide a better overview of the result set. More importantly, search results visualizations can be combined with interactive elements to allow query reformulation and/or navigation. Strategies with a proven exploratory value such as dynamic queries, i.e. the instantaneous update of a visual representation by manual adjustment of filters (Shneiderman, 1994), and, most importantly, faceted search, i.e. the use of (hierarchical) categorical metadata for search (Yee et al., 2003), have further motivated our interest in an interactive presentation of search results.

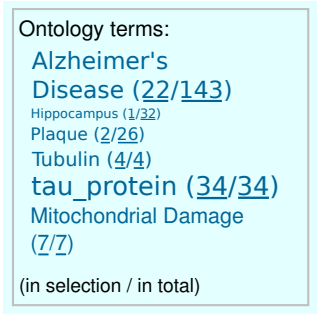
Straightforward information visualization elements were introduced in the CSIDx web interface to facilitate user interaction with the search results. Initially, search results were displayed as a list, in a sortable table of filenames or in a grid of thumbnails. While a standard design pattern, the list view lacks any overview of the results and provides no flexibility in refining the query. In response, the result page was redesigned and the standard result list has been updated with two new panels, the ontology term cloud and the modality graph (cf. Figure 2.7).

The ontology term cloud (cf. Figure 2.7b) is a list of all ontology annotations appearing in the given result set, presented in the popular tag cloud metaphor. Tag clouds, i.e. weighted lists of words, are typically related with folksonomies, i.e. sets of freely generated tags, but not with ontologies, which are highly structured. Yet, the tag cloud representation has some qualities that are of particular interest to our purposes. According to Sinclair and Cardew-Hall (2008), tag clouds can provide a visual summary of the corresponding dataset, while Rivadeneira et al. (2007) believe tag clouds to be suitable for impression formation, i.e. gaining a general understanding on a presented dataset. Furthermore, interactive tag clouds[3] can provide an easy way to browse, particularly in a non-directional, exploratory way (Sinclair and Cardew-Hall, 2008; Rivadeneira et al., 2007). Our ontology term cloud supports both browsing and query refinement by functioning as a combination of a tag cloud, in its popular form, and a drill cloud, as proposed by Newton (2008). More specifically, the user can either refine the current search

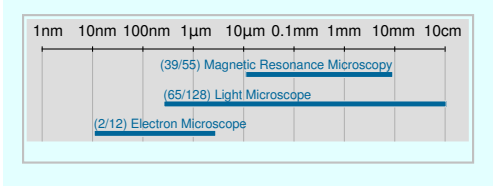
[3]When a tag cloud consists of hyperlinks instead of static words, clicking on a word stereotypically corresponds with a search for all items tagged with this word.



(a) The search results web page. Here, the results of a query for the ontology term 'Ovary' (NCI thesaurus). The web page provides a (paginated) list of results augmented with the ontology term cloud and the modality graph.



(b) Ontology term cloud



(c) Modality graph

Figure 2.7: Search results in CSIDx.

by adding another term to the current search or start a new search on a term. Terms in the ontology term cloud are weighted based on their frequency in the current result set. Per term, both the relative frequency, i.e. number of hits in the current result set, and the global frequency, i.e. number of hits in the entire (visible to the user) collection, are indicated.

The modality graph (cf. Figure 2.7c) is a dynamic graphic of the resolution range supported in the database. In the modality graph, all modalities present in a given result set are drawn at their approximate resolving ranges. Per modality, the number of hits in the current result set and the number of hits in the entire (visible to the user) collection are indicated. The modality graph allows queries on modality in a way similar to the ontology term cloud: The user can either add a modality restriction to the current search or start a new search on modality. While the resolution ranges are drawn only in approximation, the use of a resolution axis is a design choice motivated by the context and domain of the collection. On a sidenote, the modality graph is a lightweight interactive graphical representation generated entirely via CSS (Cascade Style Sheet) styles. As such, it was an early experiment towards web-native information visualization. At present, with significant advances in graphical support from within the browser, a number of toolkits deliver appealing, interactive, browser-rendered visualizations.

The above-mentioned panels emphasize the two major types of metadata in CSIDx, namely ontology and microscopy metadata. They visualize aspects of the data and result set that were previously out of sight and provide a means for the user to browse the result set and collection. Using simple interactive and visual elements, we are able to support a richer context for a given search. Eventually, the user is invited to examine the results and the collection as a whole rather than as an enumeration of individual entries.

2.6 Conclusion

CSIDx has addressed the challenges of biological image management and sharing by introducing semantic web technologies, i.e. ontologies. In effect, CSIDx proposes that a semantic image annotation that is both machine readable and human understandable will improve the management, sharing and integration of image data. Such tasks are dictated by the needs of biological research practice in general and the aims of the Cyttron project in particular. The emphasis given on metadata propagated into our understanding of (scientific) images: Images are not self-sustaining entities but composite entities of pixel data and metadata; the one (pixel data) should not be considered without the other (metadata). In

fact, most of the interactions we design are centred around (communicating or exploring) metadata. What is more, emphasis on ontology metadata resulted in a mental model of the CSIDx collection as a structure of interconnected entities that are linked together based on annotation. Such a vision was already prominent in the conception of the CSIDx and enforced by the nature and semantic potential of ontologies.

A thorough annotation by an expert, as required by CSIDx, is a task of considerable effort. During the development of CSIDx, it was apparent that we need to seriously support our users in the process of annotation. As the task can not be fully automated yet, a significant part of our support must be delivered in the form of a reasonable interface. The challenge of acquiring quality metadata in a user-friendly way remains formidable and will require the attention of HCI practitioners. Recent work on games for image annotation (von Ahn and Dabbish, 2004; Russell et al., 2008; Goh et al., 2011) indicates that costly operations need not always be painful; the applicability of similar ideas in CSIDx is still to be considered. Then again, it may be time for the CSIDx system to consider aspects other than data annotation, namely the access, retrieval and exploration of the available data. After all, the CSIDx system was conceived as a system for biologists to benefit from one anothers' research. To this end, more opportunities for 'consuming' the available data are needed. To date, we have provided simple, interactive visualization aids to better illustrate the CSIDx data (cf. ontology viewer) and collection (cf. search results visualization).

All in all, the context and background of Cyttron and CSIDx have greatly influenced our understanding of biological research practice and its needs. Evidently, both projects have impressed the significance of connectivity across image research material. As said, the Cyttron community is a manifestation of researchers wanting to learn from connections established across one another's images. Establishing connections by crossing modalities, subject matters or researchers is a highly desirable act. We extrapolate this idea to suggest that integration as linking, connecting, associating research material to other material or to new material is also a highly desirable cognitive act.