

# Imaging Informatics: Toward Capturing and Processing Semantic Information in Radiology Images

D. L. Rubin, S. Napel

Department of Radiology, Stanford University, Stanford, CA, USA

## Summary

**Objectives:** To identify challenges and opportunities in imaging informatics that can lead to the use of images for discovery, and that can potentially improve the diagnostic accuracy of imaging professionals.

**Methods:** Recent articles on imaging informatics and related articles from PubMed were reviewed and analyzed. Some new developments and challenges that recent research in imaging informatics will meet are identified and discussed.

**Results:** While much literature continues to be devoted to traditional imaging informatics topics of image processing, visualization, and computerized detection, three new trends are emerging: (1) development of ontologies to describe radiology reports and images, (2) structured reporting and image annotation methods to make image semantics explicit and machine-accessible, and (3) applications that use semantic image information for decision support to improve radiologist interpretation performance. The informatics methods being developed have similarities and synergies with recent work in the biomedical informatics community that leverage large high-throughput data sets, and future research in imaging informatics will build on these advances to enable discovery by mining large image databases.

**Conclusions:** Imaging informatics is beginning to develop and apply knowledge representation and analysis methods to image datasets. This type of work, already commonplace in biomedical research with large scale molecular and clinical datasets, will lead to new ways for computers to work with image data. The new advances hold promise for integrating imaging with the rest of the patient record as well as molecular data, for new data-driven discoveries in imaging analogous to that in bioinformatics, and for improved quality of radiology practice.

## Keywords

Imaging informatics ontologies, content based image retrieval, structured reporting

Yearb Med Inform 2010;34:42

## Introduction

Imaging informatics research has been propelled by the unique and evolving information needs of radiologists. Initial work was motivated by the advent of digital imaging and the need for radiologists to interpret images in a filmless environment. The focus of early imaging informatics research was thus on image management, communication, and visualization of 2D (and more recently of 3D) data sets, resulting in development of Picture Archiving and Communications Systems (PACS) [1]. This replacement of film-based radiology systems with an electronic infrastructure for the storage and retrieval of image data created opportunities for more efficient workflow [2]. Research has also involved the images themselves, with image processing methods developed for image segmentation, computer assisted detection, to reduce noise and distortion, and to provide advanced visualization to enhance the conspicuity of structures and abnormalities within images [3].

As radiology has evolved from images recorded and viewed on film to being stored, retrieved, and manipulated electronically, radiologists have been confronted with challenges that have created new imperatives for imaging informatics. Technical advances in CT and MRI are resulting in an explosion in the number of images acquired in each study, necessitating new informatics techniques to manage the growing numbers, size, and complexity of images generated by these and

other modern imaging technologies [4]. The advent of digital images brought expectations of greater portability and accessibility as patients move within diverse health care systems, resulting in efforts to define protocols to enable institutions to share images [5].

In reviewing the recent literature, some trends are emerging that are pushing imaging informatics beyond the pragmatic issues of improving the management and display of large collections of images to helping radiologists and researchers to access and use the *information content* in images and reports for discovery and to improve clinical care. For example, new informatics methods could use large collections of images in the PACS with the associated reports in the RIS and other clinical information (e.g., lab results, demographics, pathology) [6], to discover patterns in imaging results, diseases, and practice patterns for research discovery, quality improvement programs, and teaching activities [7, 8]. The clinical performance of radiologists in detecting abnormalities in images might be improved through computer methods that analyze the image contents to identify potential suspicious image regions [9].

The goal of this paper is to describe these recent developments and new directions of research in imaging informatics that aim to permit clinical radiologists and researchers to leverage the information in images to improve clinical practice and enable discovery. We performed PubMed searches to retrieve recent articles related to imaging

informatics, looking in particular at papers from the past two years. Much of the literature focuses on the common topics of image processing, visualization, and computerized detection. We discuss three new trends that involve capturing and using structured semantic image content: (1) standard terminologies and ontologies for describing images, (2) structured representation of image content, and (3) retrieval of image content for decision support. These trends have close parallels to related research occurring in the biomedical informatics community, which we will also highlight in this review and use to point to directions for future research in this exciting field.

## 1. Standard Terminologies and Ontologies for Describing Images

Radiologists record in their reports a variety of information pertaining to their interpretation of images, such as clinical indications, technique of imaging, anatomy, visual features of abnormalities, diagnostic interpretations, and recommended management. The language radiologists use to describe images can vary greatly [10], which limits clear communication in radiology reports and the ability to consistently index and use radiology information in applications. Standard terminologies (and ontologies) have thus been recently advocated to improve radiology reporting [11].

The need for standard terminology in radiology is similar to the need that motivated creation of controlled terminologies in other domains, such as MeSH, SNOMED, and other terminologies in the UMLS [12, 13]. The primary goal is to provide a means for radiologists to communicate clear, concise, and orderly descriptions of imaging findings in understandable, unambiguous language. Another goal is to promote an orderly thought process and logical assessments and recommendations based on observed imaging features. In fact, the diagnostic thought

process relating terminology-based description of radiology images to likely diagnoses has been modeled in a few recent decision support systems [14, 15]. A third goal is to enable radiology research; data mining is facilitated by the use of standard terms to code large collections of reports and images [16]. Finally, ontologies can enhance the ability of image data to be organized and analyzed in content-based image retrieval applications. To support these and other goals, research is being undertaken to incorporate controlled terminology into radiology reports [17, 18].

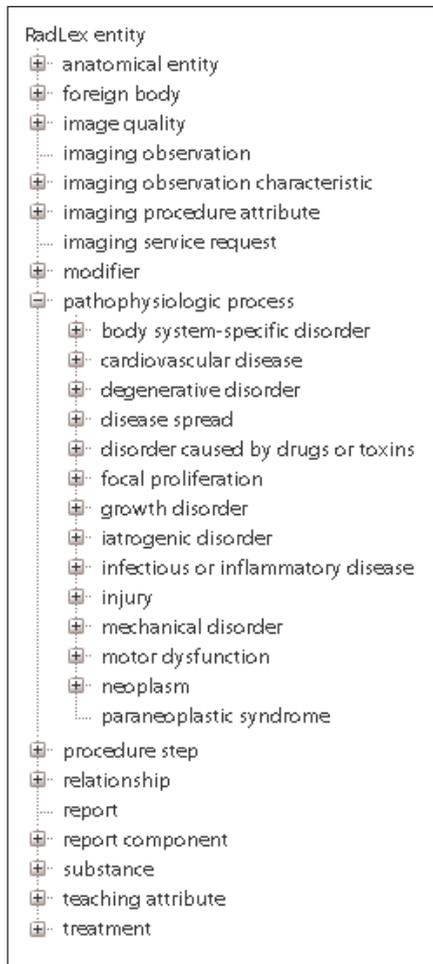
A variety of terminologies for imaging are being developed. The Breast Imaging Reporting and Data System (BI-RADS) is a lexicon of descriptors and a reporting structure comprising assessment categories and management recommendations created by the American College of Radiology [19]. Terminologies are also being created in other radiology imaging domains, including the Fleischner Society Glossary of terms for thoracic imaging [20], the Reporting Terminology for Brain Arteriovenous Malformations [21], the Nomenclature of Lumbar Disc Pathology [22], terminologies for image guided tumor ablation [23] and transcatheter therapy for hepatic malignancy [24], and the CT Colonography Reporting and Data System [25].

Some specialty groups in radiology have been developing “classification systems” for particular diseases to improve the consistency in interpretation, diagnosis, and management. Classification systems are similar to controlled terminologies in that they enumerate standard lists of imaging features (sometimes also giving them a preferred name), although they do not necessarily provide standard names for those features. They also include categorizations that guide radiologists in interpreting the diagnostic significance of combinations of imaging features. Such classification systems have been developed for evaluation of thyroid

nodules [26], renal cysts [27], and pancreatic cysts [28]. Classification systems generally lack information contained in controlled terminologies, such as synonyms, term identifiers, and relationships to other terms.

In general, the existing radiology terminologies comprise lists of terms and definitions, but not other information, such as relationships to other terms (e.g., synonymy) or taxonomic structure. On the other hand, *ontologies* can specify the relationships that exist among terms, and there is much work in developing ontologies in several biomedical domains [29-31]. Recently, the radiology community has begun creating an ontology to unify existing terminology efforts and to provide a single overarching terminology structure. RadLex [32, 33] is a comprehensive ontology providing controlled terms for uniform indexing and retrieval of radiology information. Subspecialty groups that are part of the RadLex project have developed component terminologies to cover their respective domains [34]. RadLex includes thousands of descriptors of visual observations and characteristics for describing imaging abnormalities, as well as terms for naming anatomic structures, radiology imaging procedures, and diseases (Figure 1). Each term in RadLex contains a unique identifier as well as a variety of attributes such as definition, synonyms, and foreign language equivalents. The latter attribute provides a means for RadLex to become a multi-lingual terminology, and translations of RadLex are being created [35].

In addition to a lexicon of standard terms, the RadLex ontology includes *term relationships*—links between terms to relate them in various ways to encode radiological knowledge. For example, the *is-a* relationship records subsumption (e.g., astrocytoma is a type of glioma; Figure 2). Other relationships include *part-of* (e.g., “caudate lobe of liver” is *part-of* “liver”), connectivity, blood supply, and other relationships to encode anatomic knowledge



**Fig. 1** Top-level term hierarchy of RadLex. RadLex includes term hierarchies for describing anatomy ("anatomical entity"), imaging observations ("imaging observation") and characteristics ("imaging observation characteristic"), imaging procedures and procedure steps ("procedure step"), diseases ("pathophysiologic process"), treatments ("treatment"), and components of radiology reports ("report"). All terms can be expanded to reveal a hidden sub-hierarchy; the sub-hierarchy under pathophysiologic process is shown as an example.

[36]. These relationships are enabling computer-reasoning applications to process image-related data annotated with RadLex. For example, the RadLex taxonomies can be used for query expansion, in which a query for a general term (e.g., glioma) is expanded into comprehensive, granular terms (e.g., astrocytoma, ependymal tumor, etc; Figure 2). Controlled terminology can greatly improve search for radiology information; it has been shown that RadLex-based indexing of radiology

journal figure captions achieved very high precision and recall, and significantly improved image retrieval over keyword-based search [37].

A number of imaging informatics projects are using RadLex to index radiology reports [38] and imaging literature [37]. Indexed texts can then be searched using RadLex terms or synonyms (which map to the preferred RadLex term). Work is also underway to introduce RadLex controlled terms into radiology reports to reduce radiologist variation in use of terms for describing images [17].

Ultimately, in order for controlled terminology such as RadLex to impact radiology research and patient care, it will need to be incorporated into the clinical workflow. Future imaging informatics research will need to develop methods and user interfaces to efficiently translate radiologist verbal descriptions of images into controlled terms, or prompt them for those terms while dictating. In the clinical research setting, tools are appearing to enable radiologists to annotate and query image databases using RadLex and other controlled terminologies [16, 39].

## 2. Structured Representation of Image Content

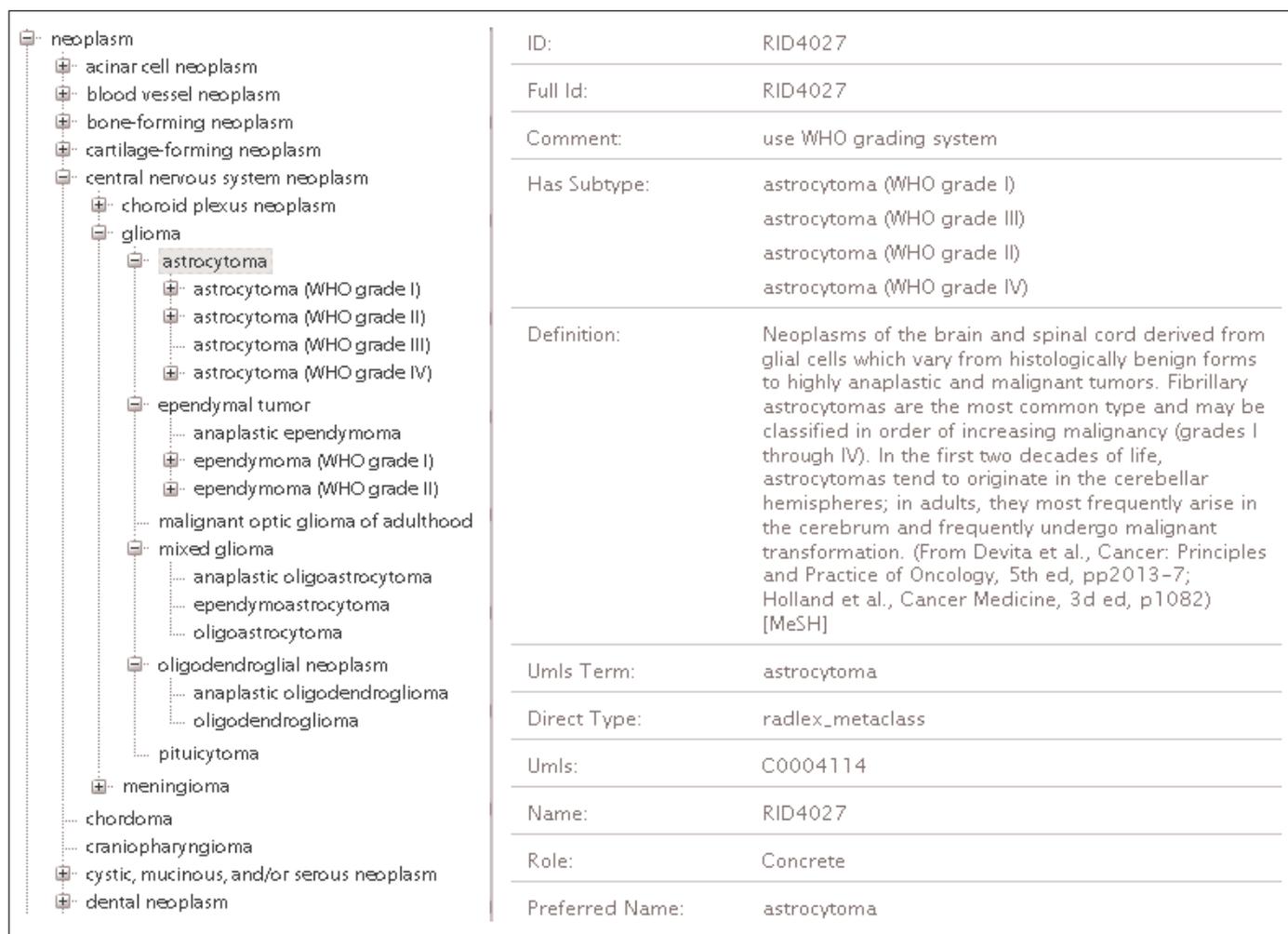
Image information is recorded in two ways: (1) as reports dictated by the radiologist evaluating the images, and (2) as the raw image itself, which can be subsequently processed by machines to extract features. Because the raw image and text report are not well suited to query and analysis directly, recent work in imaging informatics is focusing on structuring the content of images and reports, and unifying this information into a single searchable data structure. This is accomplished by applying knowledge representation approaches to the images and text, generally using ontologies and annotation schemes, to provide a human-readable and machine-interpretable representation of the image contents. Two ap-

proaches are being pursued: *structured reporting* and *semantic annotation of images*. As described subsequently, the structured representations of image content can then be used for data-intensive applications such as mining images and related data, and for decision support.

### 2.1 Structured Reporting

The radiology report documents the imaging study, including the indications, technique, and the interpreting physician's analysis of the imaging findings. The content of these reports is valuable for research and teaching, but the unstructured free-text format of reports makes it challenging for machines to access this information. Natural language processing (NLP) techniques have been applied to radiology reports [40-42] to extract a variety of types of information. NLP has also been applied to generate structured reports from the free text report [42, 43]. However, since the source reports may be incomplete, the completeness of structured reports produced by NLP depends on the quality of the source reports. Recent informatics research is thus focusing on methods to enable radiologists to create complete structured reports as part of the reporting workflow [17].

Structured reporting is the process of using predefined formats and terms to create reports, and covers a spectrum of structured representations that ranges from *capturing basic report structure* to *capturing the complete report content*. The recent work on structured reporting is defining and implementing templates or checklists (e.g., imaging procedure descriptions, clinical indications, imaging findings, and summary information [44]) to capture basic information about reports [17]. The goal of this approach is to improve quality and efficiency of radiology reporting operations. It is expected that future work will develop systems to capture more complete information about the imaging procedure and its results (e.g., clinical data, technical parameters,



**Fig 2** RadLex sub-hierarchy for neoplasms. Panel on the left shows the terms representing subtypes of neoplasms, expanded to show types of gliomas (shown indented under each parent term). Panel on the right shows the attributes for a selected term (astrocytoma in this example).

measurements, annotations, and key images) [45]. To that end, an extension to the DICOM standard called DICOM Structured Reporting (DICOM-SR), was recently created to enable storing a diversity of types of metadata about images, including structured information in radiology reports, in DICOM objects [46]. The Integrating the Healthcare Enterprise (IHE) effort has been developing profiles to enable acquisition of structured radiology information as it pertains to teaching cases, called the Teaching Files and Clinical Trial Export (TCE) radiology technical framework extension [47]. As other important use cases beyond teaching emerge, additional standards and im-

plementing tools for capturing and leveraging structured image information will emerge.

There are clinical needs for detailed, structured representation of the complete report content. For example, recent studies have found that current unstructured radiology reports are inconsistent [48] and incomplete; they often do not provide critical information needed for decision making, and there are substantial variations in the quality of reports in the community [49]. Fully structured reports could capture all image information using controlled terminologies such as RadLex and enforce constraints on the valid values for particular data collec-

tion fields in the reporting templates. Algorithms can be developed to automatically or semi-automatically abstract the report results and provide feedback to the radiologist about potential omissions or inconsistencies in the report. The abstracted data can then be used for data mining and computerized clinical decision support.

Two areas of active work in developing structured reporting approaches are: (1) creating user-friendly interfaces and efficient workflow, and (2) defining and implementing minimal information requirements for reporting. User-friendly interfaces for structured reporting are crucial because producing a structured report can be more

time-consuming than dictating a free-text report. New paradigms for recording structured representation of the radiologist's interpretation as they interact with the image have been proposed, such as gesture-based reporting [50]. Research efforts have also been undertaken to incorporate structured reporting into voice recognition systems [51].

A second area of work in structured reporting is defining "minimal information requirements," which describes the information that needs to be reported for imaging studies, depending on modality, clinical indication, and abnormalities observed. For example, a CT of the chest performed to evaluate a lung nodule should describe its size, composition, presence of calcification, anatomic location and image location [52]. Minimal information content requirements for radiology imaging reports are beginning to be specified, such as exemplified in cardiovascular imaging [53]. The Radiological Society of North America recently began an initiative in structured reporting [17], one of the goals of which is to define minimal information reporting requirements for each radiology imaging procedure, including the use of controlled terminology (RadLex) for conveying imaging results.

Imaging informatics work in the future will be able to leverage the content made explicit in structured reporting. For example, having the report information in structured format will permit image mining in large databases as well as linking image data to other clinical data. In addition, quality assurance systems will be able to evaluate the content of reports as the radiologist generates them, alerting the radiologist if the report is incomplete or inconsistent.

## 2.2 Semantic Annotation of Images

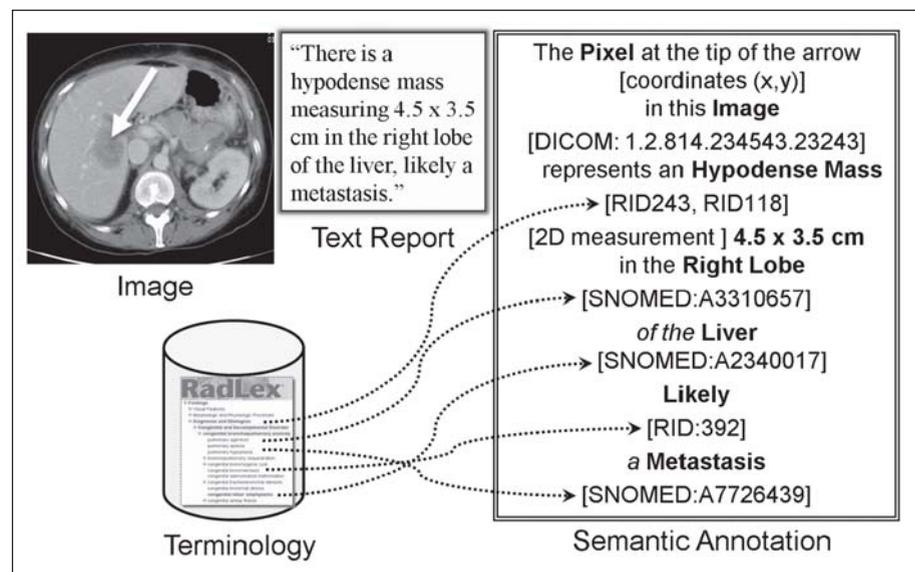
Radiology images and the corresponding radiologist report are currently disconnected; e.g., the report may describe a mass in an organ, and the image may

contain a region of interest (ROI) measuring the lesion, but there is no information directly linking the description of the lesion in the report with the ROI in the image. Such linkage could enable applications such as content-based image retrieval, as described below. In addition, the annotations or measurements in images are usually recorded as graphical overlays and not directly accessible to query and analysis. For example, radiologists usually measure the size of tumors to help physicians assess the response to treatment, but it is not possible to query the images to produce a summary of lesion sizes directly from the annotations.

So-called "semantic annotation" methods are being developed to make information content in image annotations explicit. These methods use ontologies to convey the pertinent image information explicitly and in human-readable and machine accessible format. For example, a semantic annotation might record the coordinates of the tip of an arrow and indicate the organ (anatomic location) and imaging observations (e.g., mass) in that organ

(Figure 3). These annotations can be recorded in a standard, searchable format, such as the Annotation and Image Markup (AIM) schema, recently developed by the National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG) initiative [16, 54]. AIM captures a variety of information about image annotations, e.g., regions of interest, lesion identification, location, measurements, method of measurement, and other qualitative and quantitative features [55].

The AIM schema includes use of controlled terms as semantic descriptors of lesions (e.g., RadLex). It also provides a syntax associating a region of interest (ROI) in an image with the aforementioned information, enabling raw image data to be linked with semantic information, and thus bridges the current disconnect between semantic terms and the lesions in images being described. In conjunction with RadLex, the AIM information model provides a standard structure within which image query and analysis applications can be built. In addition to image metadata, AIM also can also cap-



**Fig. 3** Semantic annotation of images. The radiologist's image annotation (left) and interpretation (middle) associated with the annotation are not represented in a form such that the detailed content is directly accessible. The same information can be put into a structured representation as a semantic annotation (right), comprising terms from controlled terminologies (Systematized Nomenclature of Medicine (SNOMED) and RadLex) as well as numeric values (coordinates and measurements).

ture a subset of clinical data and other non-image data pertinent to the image being evaluated, such as patient gender and age.

Work to create tools for semantic annotation of images is underway [39]. Automated semantic image annotation methods are also being pursued [56-58]. As semantic annotation of images becomes a part of the routine evaluation of radiological images, new research and applications will leverage this information to help radiologists analyze data in large databases of images. For example, cancer patients often have many serial imaging studies in which a set of lesions is evaluated at each time point. Automated methods will be able to use semantic image annotations to identify the measurable lesions at each time point and produce a summary of, and automatically reason about, the total tumor burden over time, helping physicians to determine how well patients are responding to treatment [59].

The semantic annotation methods for images could even be applied to other resources such as the literature. An increasing amount of the medical scientific literature is available for search (including images), and several search interfaces are being developed, such as Goldminer (<http://goldminer.arrs.org/>), MedSearch (<http://medgift.unige.ch:8080/MedSearch/faces/Search.jsp>), and BioText (<http://biosearch.berkeley.edu/?action=about>). Automated methods to map ontology terms to such radiology resources to accomplish semantic annotation have recently been in development [60]. Such annotated information resources will permit more accurate searches, as well as enable integrating other imaging-related resources.

### 3. Retrieval of Image Content for Decision Support

A major challenge for practicing radiologists is integrating the multitude of imaging data, clinical data, and soon

molecular data, to formulate a diagnosis and management plan for the patient. Decision support applications are needed to help radiologists understand the biomedical import of this information and to provide guidance [61].

Since a key aspect of radiological interpretation is recognizing the biomedical meaning of patterns of imaging features (the “imaging phenotype”), searching databases for similar images with known diagnoses could be an effective strategy to provide decision support. Searching databases for similar images is referred to as content-based image retrieval (CBIR); such methods have been well studied and are already useful in non-medical applications such as consumer imaging and on the Web [62, 63], and have even been employed for automatic image annotation [62].

A number of groups are developing CBIR methods in radiology, with the goal to retrieve similar images from a database to provide decision support. Several studies have shown that CBIR approaches can improve radiologist diagnostic performance [64-66]. The approach to CBIR in radiology generally is based on deriving quantitative characteristics from the images (e.g., pixel statistics, spatial frequency content, etc.), followed by application of similarity metrics to search databases for similar images [62, 67]. In current work, the focus is generally *entire images*, describing them with sets of numerical features [68, 69], with the goal of retrieving similar images from medical collections [70] that provide benchmarks for image retrieval. However, in many cases only a particular *region* of the image is of interest when seeking similar images (e.g., finding images containing similar-appearing lesions to those in the query image). More recently, “localized” CBIR methods are being developed in which a part of the image containing a region of interest is analyzed [68, 71, 72].

A limitation of the recent work on retrieval of images based on image content is that it focuses on query of *single* images. Challenges yet to be tackled in

image retrieval is developing features and methods for 3D retrieval, and integrating images with non-image clinical data to permit retrieval based on *entire cases* and not single images.

Another limitation of current work is that *image semantics* is not routinely included. The information reported by the radiologist (“semantic features”), is complementary to the information represented by the image pixels. One approach to capturing image semantics is analyzing and processing “visual words” in images (captured as image patches or codebooks [73-75]); these techniques have shown to perform well in CBIR applications [76]. Another approach to capture image semantics is to use the radiologist observations as image features. Several studies have found that combining the semantic information obtained from radiologists’ imaging reports or annotations with the pixel-level features can enhance performance of CBIR systems [72, 77, 78]. The knowledge representation methods described earlier make it possible to combine these types of information in large datasets. These methods will also enable integrating imaging data with molecular data, an approach that shows promise for discovering imaging biomarkers for molecular phenotypes [79].

### 4. Parallels to Recent Trends in Biomedical Informatics

The recent trends in imaging informatics research are similar to a number of directions underway in biomedical informatics research outside of imaging. As in radiology, there is a data explosion in biomedicine; large databases of biological and non-image medical data are increasingly prevalent, and numerous techniques have been developed for searching and mining biomedical databases [80]. These techniques include *knowledge representation* (ontologies), *annotation* of datasets, and *analysis* of annotated data to generate hypotheses for experimental validation [81].

Biologists, like radiologists, use a variety of terms to describe experiments and results; the naming of biological entities in the literature is inconsistent [82]. To combat this, standard terminologies and ontologies are being created for naming anatomic structures, genes, diseases, and phenotypes [12, 29-31]. Use of ontologies in biomedicine has exploded, and they are now becoming commonplace for providing a “domain of discourse” for reporting experimental results and to annotate biomedical data [83-87].

Specifying how much information should be captured in ontology-based annotation, so-called “minimal information reporting requirements,” is also common in biomedicine. Biomedical researchers use such requirements to define how to describe the context, methods, data, and conclusions that pertain to experiments [88]. Such minimum information reporting standards ensure transparency and interoperability and facilitate integration and exchange of data from different sources, and have motivated similar work recently in the imaging domain.

Ultimately, the techniques of biomedical informatics research will improve data access, interoperability, and the ability to integrate related biological datasets. Benefits of applying similar methods to large image datasets can be expected in the future. For example, one group recently integrated ontology-based description of tissue microarray images with ontological description of molecular data identified in tissues, enabling correlations between pathology and commonly related biological processes [89]. Likewise, integrating imaging data with molecular data could enable discovery of image biomarkers of molecular signatures of disease, leading to the use image-based phenotyping to guide personalized care in the future.

The developments of ontologies, semantic annotation, and image analysis methods in radiology imaging will enable new paradigms for radiology discovery and improved diagnostic capa-

bilities for practicing radiologists. In terms of research, ontology-annotated image data will facilitate exploratory data mining in large image databases to, say, correlate imaging observations, pathology, and clinical outcome. Such analyses are already commonplace in prospective radiology research, but these imaging informatics developments will enable these analyses to be performed in large-historical image datasets to generate new hypotheses. Quantitative processing of image data for techniques such as CBIR will improve clinical radiology practice by enabling radiologists to find images containing similar-appearing lesions; by integrating clinical data in decision support of content-based image retrieval, the radiologists will glean valuable insights into the current cases being interpreted.

## Conclusion

Research in imaging informatics is evolving. While there is continued work on the display and visualization of abnormalities in large imaging datasets, new efforts are focusing on representing and structuring the content that radiologists observe or annotate in images. This work will eventually result in making the full spectrum of image content—from the pixel values to the radiologists’ interpretations—accessible for query and analysis. Developing and applying ontologies and annotation schemes in radiology will enable researchers to leverage large collections of images collected in research and practice to augment discovery in biomedicine and improve clinical practice. This trend is analogous to that in bioinformatics in which similar approaches applied to biological data are enabling computer-assisted discovery.

Controlled terminologies and ontologies in radiology such as RadLex are expanding to provide comprehensive descriptors of the visual features that radiologists see in images. Infor-

mation models such as AIM that use these terminologies will provide well-defined, precise, and computable reference to image annotation information. These methods will ultimately enable computer applications to process large image collections, connect them with other clinical and molecular data, and deliver decision support to radiologists as they evaluate images at the PACS workstation.

## References

1. Ratib O. Imaging informatics: from image management to image navigation. *Yearb Med Inform* 2009;167-72.
2. Branstetter BF. Basics of imaging informatics. Part 1. *Radiology* 2007 Jun;243(3):656-67.
3. Armato SG, 3rd, van Ginneken B. Anniversary paper: image processing and manipulation through the pages of *Medical Physics*. *Med Phys* 2008 Oct;35(10):4488-500.
4. Andriole KP, Morin RL. Transforming medical imaging: the first SCAR TRIP conference a position paper from the SCAR TRIP subcommittee of the SCAR research and development committee. *J Digit Imaging* 2006 Mar;19(1):6-16.
5. Mendelson DS, Bak PR, Menschik E, Siegel E. Informatics in radiology: image exchange: IHE and the evolution of image sharing. *Radiographics* 2008 Nov-Dec;28(7):1817-33.
6. Bashyam V, Hsu W, Watt E, Bui AA, Kangaroo H, Taira RK. Problem-centric organization and visualization of patient imaging and clinical data. *Radiographics* 2009 Mar-Apr;29(2):331-43.
7. Reiner BI, Siegel EL. The clinical imperative of medical imaging informatics. *J Digit Imaging* 2009 Aug;22(4):345-7.
8. Thrall JH. Reinventing radiology in the digital age. Part II. New directions and new stakeholder value. *Radiology* 2005 Oct;237(1):15-8.
9. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 2007 Jun-Jul;31(4-5):198-211.
10. Sobel JL, Pearson ML, Gross K, Desmond KA, Harrison ER, Rubenstein LV, et al. Information content and clarity of radiologists’ reports for chest radiography. *Acad Radiol* 1996 Sep;3(9):709-17.
11. Rubin DL. Informatics methods to enable patient-centered radiology. *Acad Radiol* 2009 May;16(5):524-34.
12. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform* 2008:67-79.
13. Cimino JJ. Review paper: coding systems in health care. *Methods Inf Med* 1996 Dec;35(4-5):273-84.
14. Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE, Jr. Breast cancer: prediction with artificial neural network based on BI-RADS standard-

- ized lexicon. *Radiology* 1995 Sep;196(3):817-22.
15. Burnside ES, Davis J, Chhatwal J, Alagoz O, Lindstrom MJ, Geller BM, et al. Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. *Radiology* 2009 Jun;251(3):663-72.
  16. Channin DS, Mongkolwat P, Kleper V, Rubin DL. Computing human image annotation. *Conf Proc IEEE Eng Med Biol Soc* 2009;1:7065-8.
  17. Kahn CE, Jr., Langlotz CP, Burnside ES, Carrino JA, Channin DS, Hovsepian DM, et al. Toward best practices in radiology reporting. *Radiology* 2009 Sep;252(3):852-6.
  18. Sistrom CL, Langlotz CP. A framework for improving radiology reporting. *J Am Coll Radiol* 2005 Feb;2(2):159-67.
  19. D'Orsi CJ, Newell MS. BI-RADS decoded: detailed guidance on potentially confusing issues. *Radiol Clin North Am* 2007 Sep;45(5):751-63, v.
  20. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008 Mar;246(3):697-722.
  21. Reporting terminology for brain arteriovenous malformation clinical and radiographic features for use in clinical trials. *Stroke* 2001 Jun;32(6):1430-42.
  22. Appel B. Nomenclature and classification of lumbar disc pathology. *Neuroradiology* 2001 Dec;43(12):1124-5.
  23. Goldberg SN, Grassi CJ, Cardella JF, Charboneau JW, Dodd GD, 3rd, Dupuy DE, et al. Image-guided tumor ablation: standardization of terminology and reporting criteria. *J Vasc Interv Radiol* 2009 Jul;20(7 Suppl):S377-90.
  24. Brown DB, Gould JE, Gervais DA, Goldberg SN, Murthy R, Millward SF, et al. Transcatheter Therapy for Hepatic Malignancy: Standardization of Terminology and Reporting Criteria (Reprinted from *J Vasc Interv Radiol*, vol 18, pg 1469-1478, 2007). *Journal of Vascular and Interventional Radiology* 2009 Jul;20(7):S425-S434.
  25. Zalis ME, Barish MA, Choi JR, Dachman AH, Fenlon HM, Ferrucci JT, et al. CT colonography reporting and data system: a consensus proposal. *Radiology* 2005 Jul;236(1):3-9.
  26. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab* 2009 May;94(5):1748-51.
  27. Bosniak MA. The current radiological approach to renal cysts. *Radiology* 1986 Jan;158(1):1-10.
  28. Sahani DV, Kadavigere R, Saokar A, Fernandez-del Castillo C, Brugge WR, Hahn PF. Cystic pancreatic lesions: a simple imaging-based classification system for guiding management. *Radiographics* 2005 Nov-Dec;25(6):1471-84.
  29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium Nat Genet.* 2000 May;25(1):25-9.
  30. Robinson PN, Kohler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet* 2008 Nov;83(5):610-5.
  31. Rosse C, Mejino JL, Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform* 2003 Dec;36(6):478-500.
  32. Rubin DL. Creating and curating a terminology for radiology: ontology modeling and analysis. *J Digit Imaging* 2008 Dec;21(4):355-62.
  33. Langlotz CP. RadLex: a new method for indexing online educational materials. *Radiographics* 2006 Nov-Dec;26(6):1595-7.
  34. Kundu S, Itkin M, Gervais DA, Krishnamurthy VN, Wallace MJ, Cardella JF, et al. The IR RadLex project: an interventional radiology lexicon—a collaborative project of the Radiological Society of North America and the Society of Interventional Radiology. *J Vasc Interv Radiol* 2009 Apr;20(4):433-5.
  35. Marwede D, Daumke P, Marko K, Lobsien D, Schulz S, Kahn T. [RadLex - German version: a radiological lexicon for indexing image and report information]. *Rofo* 2009 Jan;181(1):38-44.
  36. Mejino JL, Rubin DL, Brinkley JF. FMA-RadLex: An application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. *AMIA Annu Symp Proc* 2008:465-9.
  37. Kahn CE, Rubin DL. Automated Semantic Indexing of Figure Captions to Improve Radiology Image Retrieval. *J Am Med Inform Assoc* 2009 May-Jun;16(3):380-6.
  38. Marwede D, Schulz T, Kahn T. Indexing thoracic CT reports using a preliminary version of a standardized radiological lexicon (RadLex). *J Digit Imaging* 2008 Dec;21(4):363-70.
  39. Rubin DL, Rodriguez C, Shah P, Beaulieu C. iPad: Semantic annotation and markup of radiological images. *AMIA Annu Symp Proc* 2008:626-30.
  40. Sistrom CL, Dreyer KJ, Dang PP, Weilburg JB, Boland GW, Rosenthal DI, et al. Recommendations for additional imaging in radiology reports: multifactorial analysis of 5.9 million examinations. *Radiology* 2009 Nov;253(2):453-61.
  41. Cheng LT, Zheng J, Savova GK, Erickson BJ. Discerning Tumor Status from Unstructured MRI Reports-Completeness of Information in Existing Reports and Utility of Automated Natural Language Processing. *J Digit Imaging* 2009 May 30.
  42. Taira RK, Soderland SG, Jakobovits RM. Automatic structuring of radiology free-text reports. *Radiographics* 2001 Jan-Feb;21(1):237-45.
  43. Fujii H, Yamagishi H, Ando Y, Tsukamoto N, Kawaguchi O, Kasamatsu T, et al. Structuring of free-text diagnostic report. *Stud Health Technol Inform* 2007;129(Pt 1):669-73.
  44. Kushner DC, Lucey LL. Diagnostic radiology reporting and communication: the ACR guideline. *J Am Coll Radiol* 2005 Jan;2(1):15-21.
  45. Reiner BI, Knight N, Siegel EL. Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol* 2007 May;4(5):313-9.
  46. Hussein R, Engelmann U, Schroeter A, Meinzer HP. DICOM structured reporting: Part 1. Overview and characteristics. *Radiographics* 2004 May-Jun;24(3):891-6.
  47. Kamaau AW, DuVall SL, Liimatta AP, Wiggins RH, 3rd, Avrin DE. Translating the IHE Teaching File and Clinical Trial Export (TCE) profile document templates into functional DICOM structured report objects. *J Digit Imaging* 2008 Dec;21(4):390-407.
  48. Naik SS, Hanbidge A, Wilson SR. Radiology reports: examining radiologist and clinician preferences regarding style and content. *AJR Am J Roentgenol* 2001 Mar;176(3):591-8.
  49. Houssami N, Boyages J, Stuart K, Brennan M. Quality of breast imaging reports falls short of recommended standards. *Breast* 2007 Jun;16(3):271-9.
  50. Reiner B, Siegel E. Radiology reporting: returning to our image-centric roots. *AJR Am J Roentgenol* 2006 Nov;187(5):1151-5.
  51. Rosenthal DF, Bos JM, Sokolowski RA, Mayo JB, Quigley KA, Powell RA, et al. A voice-enabled, structured medical reporting system. *J Am Med Inform Assoc* 1997 Nov-Dec;4(6):436-41.
  52. Radiological Society of North America. CT Lung Nodule Structured Reporting Template. <http://reportingwikirsna.org/index.php?title=Templates> (Accessed 12/28/09).
  53. Douglas PS, Hendel RC, Cummings JE, Dent JM, Hodgson JM, Hoffmann U, et al. ACCF/ACR/AHA/ASE/ASNC/HRS/NASCI/RSNA/SAIP/SCAI/SCCT/SCMR 2008 Health Policy Statement on Structured Reporting in Cardiovascular Imaging. *J Am Coll Cardiol* 2009 Jan 6;53(1):76-90.
  54. Rubin DL, Supekar K, Mongkolwat P, Kleper V, Channin DS. Annotation and Image Markup: Accessing and Interoperating with the Semantic Content in Medical Imaging. *IEEE Intelligent Systems* 2009 Jan-Feb;24(1):57-65.
  55. Channin DS, Mongkolwat P, Kleper V, Sepukar K, Rubin DL. The caBIG Annotation and Image Markup Project. *J Digit Imaging* 2009 Mar 18.
  56. Carneiro G, Chan AB, Moreno PJ, Vasconcelos N. Supervised learning of semantic classes for image annotation and retrieval. *IEEE Trans Pattern Anal Mach Intell* 2007 Mar;29(3):394-410.
  57. Yu F, Ip HH. Semantic content analysis and annotation of histological images. *Comput Biol Med* 2008 Jun;38(6):635-49.
  58. Mechouche A, Golbreich C, Morandi X, Gibaud B. Ontology-based annotation of brain MRI images. *AMIA Annu Symp Proc* 2008:460-4.
  59. Levy MA, Rubin DL. Tool support to enable evaluation of the clinical response to treatment. *AMIA Annu Symp Proc* 2008:399-403.
  60. Kahn CE, Jr., Rubin DL. Automated semantic indexing of figure captions to improve radiology image retrieval. *J Am Med Inform Assoc* 2009 May-Jun;16(3):380-6.
  61. Hudson DL, Cohen ME. Multidimensional medical decision making. *Conf Proc IEEE Eng Med Biol Soc* 2009;1:3405-8.
  62. Datta R, Joshi D, Li J, Wang JZ. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys* 2008 Apr;40(2):-.
  63. Wang JZ, Wiederhold G, Firschein O, Wei SX. Content-based image indexing and searching using Daubechies' wavelets. *Int J on Digital Libraries* 1997;1(4):311-28.
  64. Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, et al. Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. *Radiology* 2003 Jul;228(1):265-70.
  65. Giger ML, Huo Z, Vyborny CJ, Lan L, Bonta IR, Horsch K, et al. Intelligent CAD workstation for

- breast imaging using similarity to known lesions and multiple visual prompt aids. In: Sonka M, Fitzpatrick JM, editors. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series; 2002. p. 768-73.
66. Zheng B, Mello-Thoms C, Wang XH, Abrams GS, Sumkin JH, Chough DM, et al. Interactive computer-aided diagnosis of breast masses: computerized selection of visually similar image sets from a reference library. *Acad Radiol* 2007 Aug;14(8):917-27.
67. Greenspan H, Pinhas AT. Medical image categorization and retrieval for PACS using the GMM-KL framework. *IEEE Trans Inf Technol Biomed* 2007 Mar;11(2):190-202.
68. Deselaers T, Muller H, Clough P, Ney H, Lehmann TM. The CLEF 2005 automatic medical image annotation task. *Int J Comput Vis* 2007 Aug;74(1):51-8.
69. Lehmann TM, Guld MO, Deselaers T, Keysers D, Schubert H, Spitzer K, et al. Automatic categorization of medical images for content-based retrieval and data mining. *Comput Med Imaging Graph* 2005 Mar-Apr;29(2-3):143-55.
70. Hersh W, Muller H, Kalpathy-Cramer J. The ImageCLEFmed Medical Image Retrieval Task Test Collection. *J Digit Imaging*. 2009 Dec;22(6):648-55.
71. Rahmani R, Goldman SA, Zhang H, Cholleti SR, Fritts JE. Localized content-based image retrieval. *IEEE Trans Pattern Anal Mach Intell* 2008 Nov;30(11):1902-12.
72. Napel SA, Beaulieu CF, Rodriguez C, Cui J, Xu J, Gupta A, et al. Automated Retrieval of CT Images of Liver Lesions Based on Image Similarity: Method and Preliminary Results. *Radiology* 2010 (in press).
73. Avni. Addressing the ImageClef 2009 Challenge Using a Patch-based Visual Words Representation %U [http://www.clef-campaign.org/2009/working\\_notes/avni-paperCLEF2009.pdf](http://www.clef-campaign.org/2009/working_notes/avni-paperCLEF2009.pdf). Working Notes CLEF2009. 2009.
74. Jurie F, Triggs B. Creating Efficient Codebooks for Visual Recognition. Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 - Volume 01: IEEE Computer Society; 2005. p. 604-10 %@ 0-7695-2334-X-01.
75. Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification. *Computer Vision - Eccv 2006, Pt 4, Proceedings 2006*;3954:490-503.
76. Qiu G. Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition* 2002 Aug;35(8):1675-86.
77. Ruiz ME. Combining image features, case descriptions and UMLS concepts to improve retrieval of medical images. *AMIA Annu Symp Proc* 2006:674-8.
78. Zhenyu H, Yanjie Z, Tonghui L, Jianguo Z, editors. Combining text retrieval and content-based image retrieval for searching a large-scale medical image database in an integrated RIS/PACS environment; 2009. SPIE.
79. Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *Eur J Radiol* 2009 May;70(2):232-41.
80. Kann MG. Advances in translational bioinformatics: computational approaches for the hunting of disease genes. *Brief Bioinform*. 2009 Dec 10.
81. Blake JA, Bult CJ. Beyond the data deluge: data integration and bio-ontologies. *J Biomed Inform* 2006 Jun;39(3):314-20.
82. Ananiadou S, Friedman C, Tsujii J. Introduction: named entity recognition in biomedicine. *J Biomed Inform* 2004 Dec;37(6):393-5.
83. Shah NH, Jonquet C, Chiang AP, Butte AJ, Chen R, Musen MA. Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinformatics* 2009;10 Suppl 2:S1.
84. Bug W, Astahkov V, Boline J, Fennema-Notestine C, Grethe JS, Gupta A, et al. Data federation in the Biomedical Informatics Research Network: tools for semantic annotation and query of distributed multiscale brain data. *AMIA Annu Symp Proc*. 2008:1220.
85. French L, Lane S, Law T, Xu L, Pavlidis P. Application and evaluation of automated semantic annotation of gene expression experiments. *Bioinformatics* 2009 Jun 15;25(12):1543-9.
86. Roberts A, Gaizauskas R, Hepple M, Davis N, Demetriou G, Guo Y, et al. The CLEF corpus: semantic annotation of clinical text. *AMIA Annu Symp Proc* 2007:625-9.
87. Larson SD, Martone ME. Ontologies for Neuroscience: What are they and What are they Good for? *Front Neurosci* 2009 May;3(1):60-7.
88. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol* 2008 Aug;26(8):889-96.
89. Viti F, Merelli I, Caprera A, Lazzari B, Stella A, Milanesi L. Ontology-based, Tissue MicroArray oriented, image centered tissue bank. *BMC Bioinformatics* 2008;9 Suppl 4:S4.

**Correspondence to:**

Daniel L. Rubin, MD, MS  
 Assistant Professor of Radiology  
 Richard M. Lucas Center  
 1201 Welch Road, Office P285  
 Stanford, CA 94305-5488  
 USA  
 Tel: +1650 725 4936  
 Fax: +1 650 725 7944  
 E-mail: dlrubin@stanford.edu