Ontology-Enhanced Representations of Non-image Data in The Cancer Imaging Archive

Jonathan P. Bona, Tracy S. Nolan, Mathias Brochhausen
Department of Biomedical Informatics
University of Arkansas for Medical Sciences
Little Rock, Arkansas, United States
jpbona@uams.edu, tnolan@uams.edu, mbrochhausen@uams.edu

***Abstract*—The Cancer Imaging Archive (TCIA) hosts over 11 million de-identified medical images related to cancer for research reuse. These are organized around DICOM-format radiological collections that are grouped by disease type, modality, or research focus. Many collections also include diverse non-image datasets in a variety of formats without a common approach to representing the entities that the data are about. This paper describes work to make these diverse non-image data more accessible and usable by transforming them into integrated semantic representations using Open Biomedical Ontologies, highlights obstacles encountered in the data, and presents detailed representations data found in select collections.**

Keywords—cancer; imaging, ontology development; semantics

I. Introduction

Since 2011 the Cancer Imaging Archive [1] has been NCI's primary resource for acquiring, curating, managing and distributing images and related data to support Cancer Research. TCIA hosts over 11 million de-identified medical images of cancer for research reuse, organized around DICOM-format radiological collections related by disease type, modality, or research focus. The PRISM (Platform for Imaging in Precision Medicine) initiative seeks to sustain and expand TCIA's capabilities to meet the rapidly evolving requirements of cancer Precision Medicine research. Through discussions with investigators in the imaging and cancer research communities, and through review of TCIA helpdesk tickets, we have identified a number of near-term goals and challenges. These include enhanced support for reproducible research and data publication capabilities; expanded support for additional data types, including pathology data, and radiomics and pathomics feature sets; uniform management of non-image data; semantic query mechanisms and enhanced data exploration; and automatic curation of current and new data types.

Many TCIA collections include non-image data in a variety of formats, often as downloadable spreadsheet files without a common representation scheme. These include patient demographics, diagnoses, treatments, outcomes, TNM staging, gene assays and other test results, etc. Some collections provide data dictionaries or other documentation that aid the human reader in interpreting these data. However, these are not machine-interpretable, and hence are difficult to query. Complicating this is the use of different representations schemes in different collections to encode the same or similar information.

This paper describes work to make these diverse non-image data more accessible and usable. Our immediate aims are to: 1) make these data sets queryable; 2) make them computer-interpretable, and hence available for automated reasoning and more amenable to exploration and analysis; and 3) establish links between related data across collections and across data types.

To support these aims we are converting these data into common, semantically-enhanced representations using Open Biomedical Ontologies Foundry [2] resources, and integrating the results in a single repository with this shared representation, to facilitate queries such as "*Which patients in lung cancer collection have been diagnosed with metastatic colon cancer, and how was that diagnosis obtained?*", or "*Which patients across multiple head and neck cancer collections have tumors specifically in their oropharynx, and have been diagnosed with human papillomavirus, and how were those diagnoses obtained?*"

By using ontologies and semantic web technology we are making these data more readily available for query, automated reasoning, exploration, and analysis. TCIA users in general are not familiar enough with biomedical ontologies and semantic web technology to write SPARQL queries to access data. This semantic repository with transformed non-image data will serve as the back-end data store for user-friendly tools that support search and exploration of the data.

II. Non-image Data in TCIA Collections

A. Overview

The Cancer Imaging Archive currently has 74 publicly-available collections. We reviewed and compared the de-identified non-image data provided with these collections as a first step toward crafting a semantic representation useable to represent the bulk of non-image data in TCIA collections.

A large group of 18 of the public collections is provided by The Cancer Genome Atlas [3]. These collections, whose names all start with "TCGA", structure their data using a common, standardized representation scheme that is published as an RDF file in Turtle format [4]. TCGA linked data has also been exposed as a SPARQL endpoint [5]. Our work to integrate non-

image data from all TCIA collections into a single repository will be made much easier in the case of TCGA collections because of their use of a standard scheme and semantic web technology. Throughout this paper we discuss and describe only non-TCGA public collections, because those are the collections that best illustrate the diversity of available data and data representations, and the need for improved representations.

Of the 56 non-TCGA public collections, 17 include downloadable non-image data (often labeled "clinical data"). This is in addition to, and separate from, the image metadata present in many collections. We have manually reviewed each of the files provided with these collections. This section provides a summary and discussion that illustrate the richness and diversity of the data available, and the diversity of representation schemes currently used. This diversity poses significant challenges to integration of the non-image data, but also poses an unique opportunity to vastly improve the usability of this data with semantic web technology and biomedical ontologies.

The current submission process for TCIA non-image data does not specify the use of any common data model or schema, or require adherence to any specified semantics. This leads to some of the submitted data being ambiguous, or difficult to interpret. The semantically-rich representations that we are designing for this data, as presented in this paper, will become part of new submission tools TCIA that automate this curation as much as possible.

The non-image data contained in these 17 collections can be placed into 7 major categories: diagnosis, histology, genetic testing, demographics, treatment, morbidity, and neurological testing. The latter is a category only found in one of the collections that currently provide non-imaging data. Most of those categories are already broken down into subcategories. E.g. "treatment" is broken down into "primary: chemo", "primary: surgery", and "primary: radiation", and "adjuvant". Table I below indicates for each of these 17 collections the presence or absence of data in each category and subcategory. The types of non-imaging data that exist for a collection is

TABLE I. DIVERSITY OF NON-IMAGE DATA IN PUBLIC TCIA COLLECTIONS

	BREAST-DIAGNOSIS	Breast-MRI-NACT-Pilot	CT COLONOGRAPHY	Head-Neck-PET-CT	ISPY1	Ivy GAP	LGG-1p19qDeletion	LIDC-IDRI	LungCT-Diagnosis	NSCLC Radiogenomics	NSCLC-Radiomics	NSCLC-Radiomics-Genomics	PROSTATE-DIAGNOSIS	QIN-Breast	QIN Breast DCE-MRI	REMBRANDT	Soft-tissue-Sarcoma	SPIE-AAPM Lung CT Challenge
Diagnosis	■	■	■	■	▨	■	■	■	■	■	■	■	■			■	■	
Primary site											■	■						
Tumor site							■				■	■					■	■
Disease/Cancer	■	■				■		■								■	■	■
Laterality	■	▨									■	■				■		
Staging				■	▨			■	■	■	■	■				■		
Recurrence		■							■	■	■					■	■	
Nodules diagnostics							■									■		
Polyps			■													■		
Measurements			▨								■	■				■		
HPV Status				■												■		
Free text	■												■			■		
Histology						■			■	■	■	■		■	■	■	■	
Grading						■				■	■					■	■	
Response														■	■			
Genetic testing	■				■	■			■	■	■							
Demographics		■		■	■					■	■					■	■	
Treatment		■		■					■	■	■		■			■	■	
Primary: Chemo		■		■					■	■	■							
Primary: Surgery		■		■														
Primary: Radiation				■					■	■	■							
Adjuvant									■	■	■							
Morbidity								■	■	■	■					■	■	
Vital status								■	■	■	■					■	■	
Survival time								■										
Neurological testing																■		

marked green in the table. Types of non-imaging data that do not exist in a collection are represented by blank cells. If a collection provides data for at least one subtype of data, the major category is marked as existing. In very few cases a major category is represented as existing, when none of their subtypes are part of the collection's non-imaging data, but some data that of that category exists. The diagonally striped cells signify data that was not described or identified sufficiently and they are represented based on an assumption of the authors. One example for this is "laterality". In the "BREAST_DIAGNOSIS" collection we find both laterality applied to the diagnosis (on which side the tumor is located) and laterality for MRI, which specified for which side an MRI was taken. Other collections did provide laterality information, but didn't specify whether that was tumor localization information or imaging localization information. We did assume that this data represented the localization of the tumor, since that was most consistent with the context.

The following sections describe three of these collections that we have examined in more detail, highlighting specific hurdles presented by the representation schemes used. We then present ontology-based representations we have designed for use in our transformation of these data into a semantic repository.

B. LIDC-IDRI Collection

The Lung Image Database Consortium image collection (LIDC-IDRI)[1] [6, 7] contains non-image data for patients, many of whose lung cancers are the result of metastasis of other cancer types from locations other than their lungs. The data is provided as a spreadsheet labeled as "patient diagnoses". The sheet has columns for a de-identified *patient ID* linking it to other data about this person (including images), a *patient-level diagnosis*, *diagnosis method*, a *primary tumor site for metastatic disease*, and similar diagnosis information about lung nodules. We use this sheet as a running example throughout this section, focusing on the patient level diagnosis, including diagnosis method, and the primary metastatic tumor site.

An immediate obstacle to querying these data is the use of a terse coding system to indicate values. This system is presented as a key within column headers in the sheet itself (shown in Fig 2), making it available to a human reader, though not necessarily easy to interpret. This key is not computer-interpretable, making the data difficult to query even if it were extracted from the spreadsheet and used to populate a database table in this form.

For example, in this representation scheme, a **3** in the *patient level diagnosis* column indicates malignant metastatic disease, while a **3** in the *diagnosis method* column indicates that the relevant diagnosis was determined by surgical resection. Similar information is provided in separate columns for each identified lung nodule. To make matters worse, files with the same type of information in other collections use different encoding schemes, further complicating integrated querying and use of the data.

Even fields in this file with more explicit entries can be unclear or ambiguous. For instance, the *tumor site* column in this file consists of short, free text (non-standardized) descriptions, as illustrated in the excerpt in Table III, which shows three

TABLE II. LIDC-IDRI PATIENT-LEVEL DIAGNOSIS KEY

Diagnosis at the Patient Level	**Diagnosis Method**
0=Unknown	0 = unknown
1=benign or non-malignant disease	1 = review of radiological images to show 2 years of stable nodule
2= malignant, primary lung cancer	2 = biopsy
3 = malignant metastatic	3 = surgical resection
	4 = progression or response

TABLE III. TEN ENTRIES FOR DIAGNOSIS, METHOD, AND TUMOR SITE

Patient ID	Diagnosis at the Patient Level	Diagnosis Method	Primary tumor site for metastatic disease
ID1	3	4	Head & Neck Cancer
ID2	3	1	Head & Neck
ID3	3	0	Uterine Cancer
ID4	2	3	NSCLC
ID5	3	4	urothelial carcinoma
ID6	3	1	Testis
ID7	3	0	Prostate
ID8	3	2	colon cancer
ID9	3	4	colon
ID10	3	4	Metastatic colon cancer

consecutive entries (ID8 - ID10) that indicate metastatic colon cancer by using three different values in the *tumor site* column: "*colon cancer*," "*colon*," and "*metastatic colon cancer*." In this case, all three contain the word "colon", so a string-based text search for that term would locate these records. However, query and integration of these data will obviously benefit from translation to a computer-interpretable, shared representation that is explicit about which entities are involved.

Some of these tumor site entries do explicitly denote anatomical locations, containing only short words like 'colon' and 'bladder.' Others are descriptions that mention cancer types mixed with information that indicates locations ('non small cell lung left lower lobe', 'uterine cancer', 'granular cell tumor of the trachea'). Some only name a disease type ('lymphoma', 'adenocarcinoma'), or use an abbreviation that may allow a person with domain knowledge to infer the location, such as 'HCC' -- hepatocellular carcinoma, which occurs in the liver, or 'NSCLC' -- non-small cell lung cancer. As discussed more in the Methods section below, we found necessary to manually curate an intermediate spreadsheet with location-denoting terms before this data could be converted to an OWL representation.

C. Two Head and Neck Cancer Collections

The Head-Neck-PET-CT collection[2] [8] contains non-image data, including diagnostic and treatment information for patients with head and neck cancer. The HNSCC (Head and Neck Squamous Cell Carcinoma)[9, 10] collection[3] contains much of the same information. These collections overlap significantly in their contents, though with some notational differences. This section compares a subset of the non-image data provided with these two collections, focusing on a few key data types in these

[1] http://dx.doi.org/10.7937/K9/TCIA.2015.LO9QL9SX

[2] http://doi.org/10.7937/K9/TCIA.2017.8oje5q00

[3] http://doi.org/10.7937/K9/TCIA.2017.umz8dv6s

collections for which we have implemented ontology-based representations, as discussed more in the Methods section.

TABLE IV. EXCERPT FROM HEAD AND NECK SQUAMOUS CELL CARCINOMA COLLECTION

Sex	Diag	Site	Grade	T	N	M	Stage	HPV status
Male	CA tonsil	Oropharynx	moderately to poorly diff.	4	2b	0	IVA	
Male	CA larynx	Glottis	poorly diff.	3	0	0	III	
Male	CA BOT	Oropharynx	moderately diff.	1	2a	0	IVA	
Male	CA tonsil	Oropharynx	poorly diff.	2	2b	0	IVA	
Male	CA BOT	Oropharynx	poorly diff.	1	2b	0	IVA	positive
Male	CA BOT	Oropharynx	poorly diff.	1	2b	0	IVA	negative
Male	CA tonsil	Oropharynx	moderately diff.	2	2a	0	IVA	
Male	CUP	CUP	well diff.	0	2b	0	IVA	
Female	NPC	Nasopharynx	poorly diff.	4	2	0	IVA	
Male	CA tonsil	Oropharynx	moderately to poorly diff.	4a	3	0	IVB	

TABLE V. EXCERPT FROM HEAD-NECK-PET-CT COLLECTION

Sex	Primary Site	T-stage	N-stage	M-stage	TNM group stage	HPV status
M	Larynx	T3	N0	M0	stage III	-
M	Nasopharynx	T1	N1	M0	stage IIB	-
M	Larynx	T3	N2b	M0	stage IVA	N/A
M	Nasopharynx	T3	N1	M0	stage III	N/A
M	Nasopharynx	T1	N1	M0	stage IIB	-
F	Nasopharynx	T1	N2b	M0	stage III	N/A
M	Oropharynx	T4	N2b	M0	stage IVA	+
M	Oropharynx	T2	N2b	M0	stage IVA	N/A
M	Larynx	T3	N0	M0	stage III	-
F	Oropharynx	T2	N2b	M0	stage IVA	-

Both of these head and neck cancer collections contain additional types of data not discussed here, many of which we also will transform into semantic representations for our integrated repository as this project progresses. As shown in Tables IV and V below, both collections include the biological sex of the patient, among other demographic data, as well as tumor staging information, HPV status, and an indication of the primary tumor location.

III. METHODS

We designed and built representations for these data using OBO Foundry ontologies, including the Human Disease Ontology [11] and The Uber Anatomy Ontology, Uberon [12]. Instances for individual entries are linked to ontology classes to explicitly represent locations, disease types, and diagnosis methods. These representations are used to transform data from spreadsheets from these three collection into OWL/RDF files that are loaded into a triple store database for reasoning and query. This section presents the details of these representations, and of the translation process.

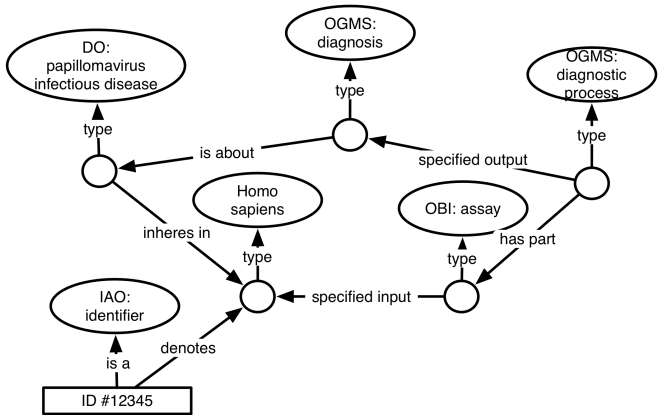

Fig. 1. Representing positive HPV status for a head and neck cancer patient

A. Ontology-based Representation

Fig. 1 shows how we represent a patient's positive HPV diagnosis in the head and neck collections. In this figure the labeled ovals stand for ontology classes. The smaller circles stand for anonymous instances, which are linked to their classes through *rdf:type* assertions. The rectangle stands for a labeled instance. HPV status is provided in these sheets without specific information about how it was determined, so we can assert only

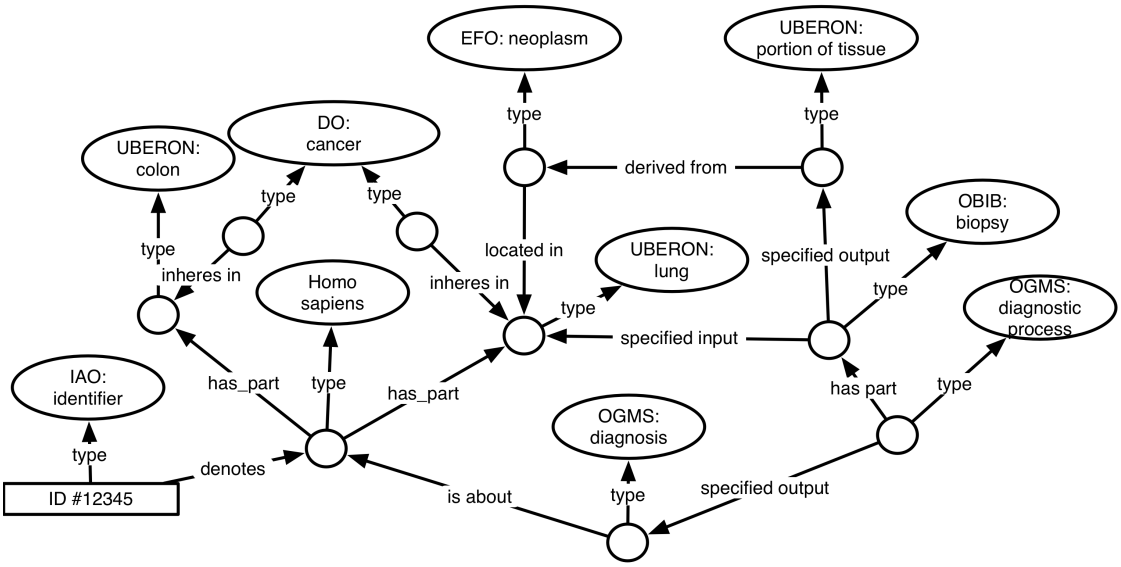

Fig. 2. Disease and diagnosis for a lung cancer patient

that an '*OGMS: diagnostic process*' with some '*OBI: assay*' has occurred (this includes physical examinations or other methods that are not strictly lab tests), and that the output was an '*OGMS: diagnosis*' *about* an instance of '*DO: papillomavirus infectious disease*' that *inheres in* the patient. Not shown in Fig. 1 is information about the patient's cancer, though a link exists, both in reality and in our representation of this data, via the patient. The representation of this information of head and neck collection data is nearly identical to the representation used for the lung collection, as shown in Fig. 4.

Fig. 2 shows our representation for a patient's disease and diagnosis using data from the LIDC collection as an example. The patient record shown here is for a person whose colon cancer has spread to their lungs, as determined by a biopsy. This patient has two instances of '*DO: cancer*', one that inheres in the patient's '*UBERON: colon*' and one that inheres in the '*UBERON: lung*'. An '*OGMS: diagnostic process*' with some '*OBIB: biopsy*' as part has produced as output an '*OGMS: diagnosis*' that *is about* the patient. The biopsy evaluated an '*UBERON: portion of tissue*' that was derived from an '*EFO: neoplasm*' that was located in the patient's lung. In this case the dataset does not contain more specific information about which type of cancer inheres in each location. Note that an OWL reasoner could infer more specific types for these instances from the assertions in Figure 2, and from logical definitions in the Disease Ontology, concluding e.g. that the instance of cancer inhering in the patient's lung is an instance of lung cancer.

B. Data Transformation and Populating Repository

As discussed above, the lung cancer collection uses some values that require manual interpretation by a human to identify which anatomical entities, if any, are specified. To facilitate the transformation of this collection into OWL, we built a spreadsheet listing all 110 unique values from the *primary tumor site* field, and used this to record and track the extent to which each value in that field indicates an anatomical location. Of these 110 *primary tumor site* entries, only 9 are short terms that precisely denote an anatomical location. 54 others explicitly mention an identifiable location, often as part of a description that also names the disease type. In total, including entries where the location can be inferred from use of a standard abbreviation, 76 out of 110 indicate a clear location for the primary tumor of the metastatic disease. For each these, we manually located and recorded the matching Uberon class in the sheet for use in our ontology-based representation of the data. This secondary sheet was then used as input to a Python script to retrieve and record the correct anatomical classes for tumor sites even for those records where the literal value stored in the source sheet did not strictly identify a location.

To transform the two head and neck collections, manual curation of a secondary sheet was unnecessary because the tumor site entries in those two sheets contain only one of a few values: 'Larynx', 'Nasopharynx', 'Hypopharynx', 'Oropharynx', 'Glottis', 'Sinus', 'Oral cavity', 'unknown', 'CUP'. A value of 'CUP' indicates *cancer of unknown primary*, so it carries similar information as the value 'unknown'. The other seven values clearly denote anatomical locations found in Uberon.

These collection data sheets, including a secondary sheet for the lung collection, were processed with a Python script using

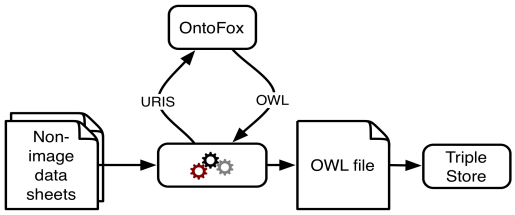

Fig. 3. Transforming non-image data

the RDFLib library to build OWL individuals from the instance data contained in each sheet, asserting the prescribed relations among these individuals, and saving the results in an OWL file. As part of this process, the script reads the spreadsheets, determines which URIs are needed, and automatically generates OntoFox [13] requests for each external ontology used. OntoFox is a web-based term extraction tool that supports ontology reuse. Our script uses OntoFox to retrieve hierarchical information and select other details only for those classes and relations that are needed to represent the data. It invokes the ROBOT command line tool [14] to convert between RDF serializations, e.g. to convert OntoFox's default RDF/XML output into turtle format for ease of use with RDFLib. The resulting OWL files are added to a triple store, making them available for reasoning and query.

```
PREFIX inheres: <http://purl.obolibrary.org/obo/RO_0000052>
PREFIX human: <http://purl.obolibrary.org/obo/NCBITaxon_9606>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX identifier: <http://purl.obolibrary.org/obo/IAO_0020000>
PREFIX denotes: <http://purl.obolibrary.org/obo/IAO_0000219>
PREFIX oroph: <http://purl.obolibrary.org/obo/UBERON_0001729>
PREFIX cancer: <http://purl.obolibrary.org/obo/DOID_162>
PREFIX has_part: <http://purl.obolibrary.org/obo/BFO_0000051>
PREFIX hpv: <http://purl.obolibrary.org/obo/DOID_11166>
PREFIX disease: <http://purl.obolibrary.org/obo/DOID_4>
select ?idl {
   # the person and identifier
   ?person rdf:type human: .
   ?id denotes: ?person .
   ?id rdf:type identifier: .
   ?id rdfs:label ?idl .

   # the person has hpv
   ?hpv rdf:type  hpv: .
   ?hpv inheres: ?person .

   # the person's oropharynx
   ?person has_part: ?o .
   ?o rdf:type oroph: .

   # cancer in the oropharynx
   ?d inheres: ?o .
   ?d rdf:type cancer: .
} limit 5
```

HNSCC-01-0050
HNSCC-01-0054
HN-HGJ-018
HNSCC-01-0098
HNSCC-01-0116

IV. RESULTS

The resulting triple store contains assertions linking patient identifiers to RDF instances representing patients, affected body parts, diagnoses, relations among those, etc. OBO Foundry Ontologies provide the types (OWL classes) for these instances and define the relations (OWL object properties).

This database can be queried using SPARQL to identify patient records matching criteria based on these fields that were previously inaccessible, as well as queries that operate across

collections. For example, the query shown above gets a list of patient identifiers for patients who have been diagnosed with HPV, and who have also been diagnosed with a cancerous tumor in their oropharynx. This query is able to retrieve results from both the Head-Neck-PET-CT collection and the Head and Neck Squamous Cell Carcinoma collection because the relevant data are now represented in the same way in the triple store. This enhanced data is immediately available for simple reasoning tasks allowed by the use of ontologies, e.g. using the partonomic information built into UBERON to support queries at different levels of anatomical granularity.

V. Discussion and Future Work

The Cancer Imaging Archive contains a wealth of diverse non-image data that is currently difficult to work with because much of it, though publicly available, is locked away in spreadsheet files that must to be downloaded and interpreted individually. As part of ongoing development work for TCIA and for the PRISM platform, we are examining the contents of these files, cataloging and characterizing the data therein, and designing realist ontology-based representations that explicitly the entities that these data are about.

The examples presented in this paper demonstrate the usefulness of ontologies and semantic web tools for knowledge representation to enable querying of otherwise opaque non-image data in these TCIA collections. We are expanding this work beyond the collections presented here to include more data from the archive. The graph-based nature of RDF stores allows us to incrementally add and link knowledge from different collections and files within them as the representation work proceeds, simplifying the task of integrating these data.

Because most users prefer not to write SPARQL queries, a next step is the development of user-friendly interfaces to help end users search, explore, and interpret these data. We also plan to provide ontology-driven submission tools that will automatically generate the same representations, allowing for seamless integration of new datasets.

VI. References

[1] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging,* vol. 26, no. 6, pp. 1045-1057, 2013/12/01 2013.

[2] B. Smith *et al.*, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature biotechnology,* vol. 25, no. 11, p. 1251, 2007.

[3] J. N. Weinstein *et al.*, "The Cancer Genome Atlas Pan-Cancer Analysis Project," *Nature genetics,* vol. 45, no. 10, pp. 1113-1120, 2013.

[4] T. C. G. Atlas. (2013). *TCGA Roadmap*. Available: https://old.datahub.io/dataset/tcga-roadmap

[5] H. F. Deus, D. F. Veiga, P. R. Freire, J. N. Weinstein, G. B. Mills, and J. S. Almeida, "Exposing the cancer genome atlas as a SPARQL endpoint," *Journal of Biomedical Informatics,* vol. 43, no. 6, pp. 998-1008, 2010/12/01/ 2010.

[6] S. G. Armato *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans," *Medical Physics,* vol. 38, no. 2, pp. 915-931.

[7] S. Armato Iii *et al.*, "Data from LIDC-IDRI. The cancer imaging archive," ed, 2015.

[8] M. Vallières *et al.*, "Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer," *Scientific Reports,* vol. 7, p. 10117.

[9] M. A. Grossberg A, Elhalawani H, Bennett W, Smith K, Nolan T, Chamchod S, Kantor M, Browne T, Hutcheson K, Gunn G, Garden A, Frank S, Rosenthal D, Freymann J, Fuller C, "Data from Head and Neck Cancer CT Atlas. ," ed. The Cancer Imaging Archive, 2017.

[10] M. A. Grossberg A, Elhalawani H, Bennett W, Smith K, Nolan T, Williams B, Chamchod S, Heukelom J, Kantor M, Browne T, Hutcheson K, Gunn G, Garden A,Morrison W, Frank S, Rosenthal D, Freymann J, Fuller C, "Imaging and Clinical Data Archive for Head and Neck Squamous Cell Carcinoma Patients Treated with Radiotherapy.," ed. The Cancer Imaging Archive, 2018.

[11] L. M. Schriml *et al.*, "Disease Ontology: a backbone for disease semantic integration," *Nucleic Acids Research,* vol. 40, no. D1, pp. D940-D946, 2012.

[12] C. J. Mungall, C. Torniai, G. V. Gkoutos, S. E. Lewis, and M. A. Haendel, "Uberon, an integrative multi-species anatomy ontology," *Genome Biology,* vol. 13, no. 1, pp. R5-R5, 01/31

[13] Z. Xiang, M. Courtot, R. R. Brinkman, A. Ruttenberg, and Y. He, "OntoFox: web-based support for ontology reuse," *BMC Research Notes,* vol. 3, no. 1, p. 175, 2010/06/22 2010.

[14] J. A. Overton, H. Dietze, S. Essaid, D. Osumi-Sutherland, and C. J. Mungall, "ROBOT: A command-line tool for ontology development," 2015, pp. 131-132.

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Under this contract the University of Arkansas is funded by Leidos Biomedical Research subcontract 16X011. Funding was also provided by U24CA215109.