PRISM Semantic Integration Approach

NCI Imaging Community Call

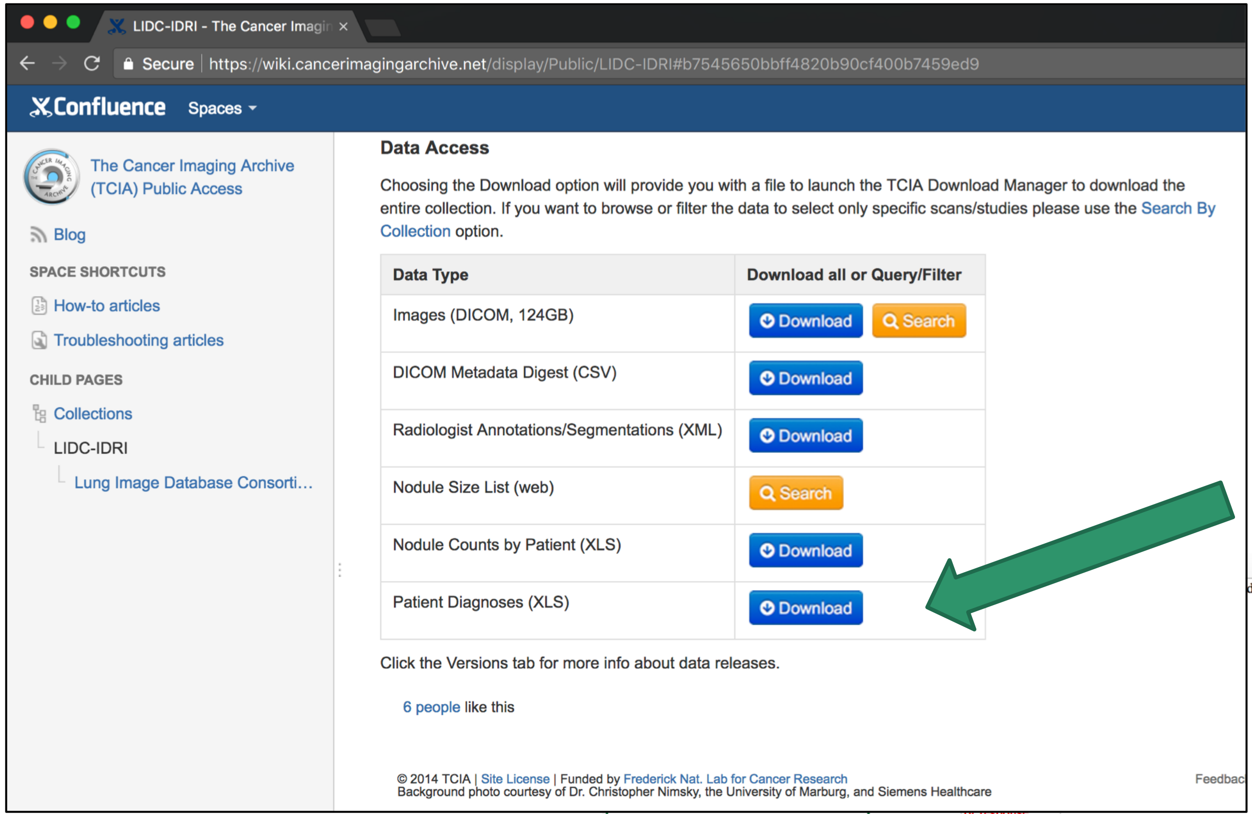
June 3, 2019

Jonathan Bona, PhD
Department of Biomedical Informatics
University of Arkansas for Medical Sciences
jpbona@uams.edu


Uniform management of all non-image data
Semantic query mechanisms to enhance data exploration
Research community guidance on priorities
Expand capacity and automate curation of current and new data types
Support radiomic and pathomic feature sets
Extend support for challenges & capture of results
Expand support for pathology data
Enable efficient deployment and integration with cloud providers
Enhance support for reproducible research
Expand existing data publication capacity

Motivation

- The Cancer Imaging Archive hosts >11 million de-identified medical images related to cancer for research reuse
- Images are in DICOM-format collections, grouped by disease type, modality, research focus, etc.
- Many collections include diverse non-image datasets
  - in a variety of formats
  - lacking a common representation
  - not discoverable/queryable
  - not integrated



LIDC-IDRI - The Cancer Imagin ×

Secure | https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI#b7545650bbff4820b90cf400b7459ed9

Confluence Spaces

The Cancer Imaging Archive (TCIA) Public Access

Blog

SPACE SHORTCUTS

How-to articles

Troubleshooting articles

CHILD PAGES

Collections

LIDC-IDRI

Lung Image Database Consorti…

Data Access

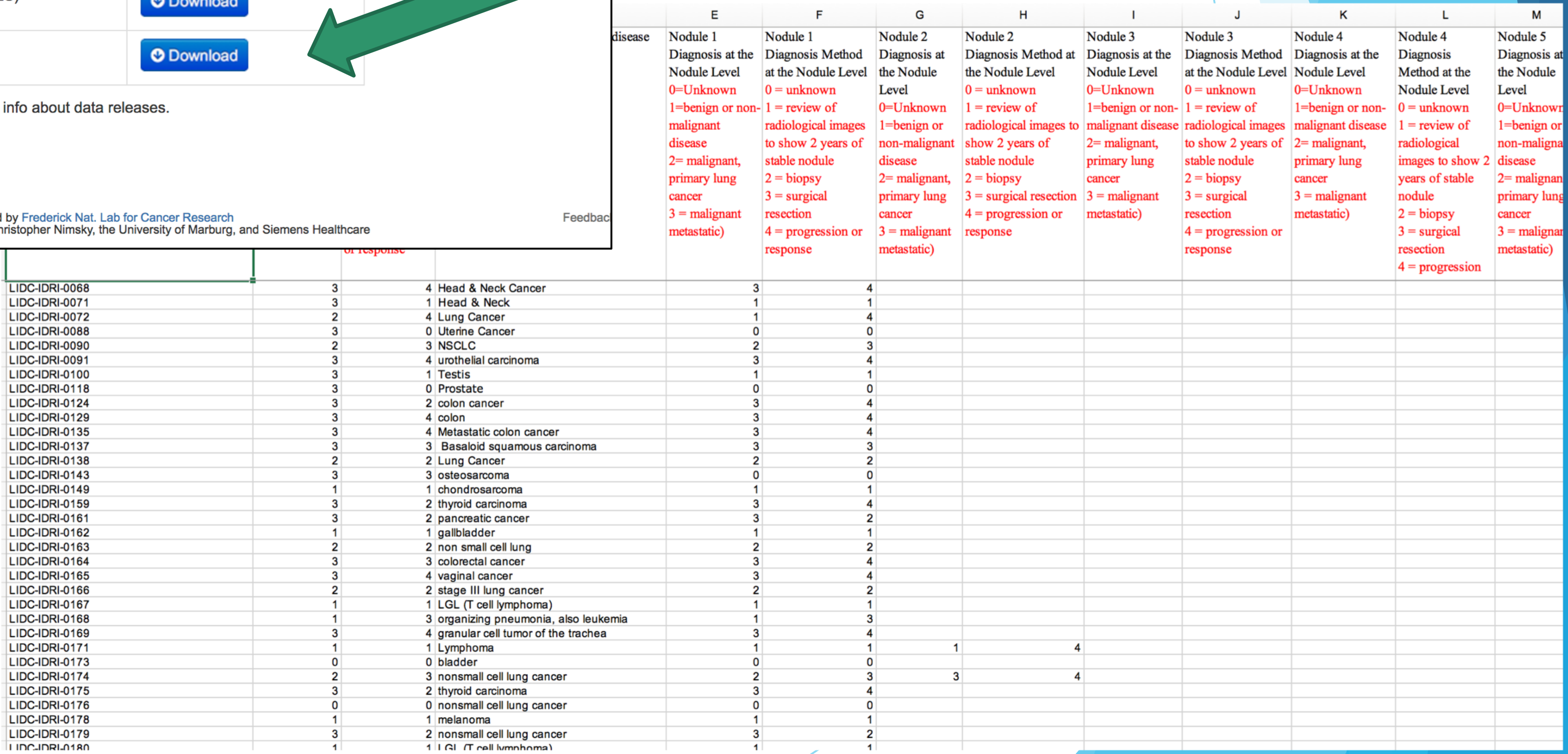
Choosing the Download option will provide you with a file to launch the TCIA Download Manager to download the entire collection. If you want to browse or filter the data to select only specific scans/studies please use the Search By Collection option.

Data Type	Download all or Query/Filter
Images (DICOM, 124GB)	Download Search
DICOM Metadata Digest (CSV)	Download
Radiologist Annotations/Segmentations (XML)	Download
Nodule Size List (web)	Search
Nodule Counts by Patient (XLS)	Download
Patient Diagnoses (XLS)	Download

Click the Versions tab for more info about data releases.

6 people like this

Feedbac



		C	D	E	F	G	H	I	J	K	L	M
		or response	disease	Nodule 1 Diagnosis at the Nodule Level 0=Unknown 1=benign or non-malignant disease 2= malignant, primary lung cancer 3 = malignant metastatic)	Nodule 1 Diagnosis Method at the Nodule Level 0 = unknown 1 = review of radiological images to show 2 years of stable nodule 2 = biopsy 3 = surgical resection 4 = progression or response	Nodule 2 Diagnosis at the Nodule Level 0=Unknown 1=benign or non-malignant disease 2= malignant, primary lung cancer 3 = malignant metastatic)	Nodule 2 Diagnosis Method at the Nodule Level 0 = unknown 1 = review of radiological images to show 2 years of stable nodule 2 = biopsy 3 = surgical resection 4 = progression or response	Nodule 3 Diagnosis at the Nodule Level 0=Unknown 1=benign or non-malignant disease 2= malignant, primary lung cancer 3 = malignant metastatic)	Nodule 3 Diagnosis Method at the Nodule Level 0 = unknown 1 = review of radiological images to show 2 years of stable nodule 2 = biopsy 3 = surgical resection 4 = progression or response	Nodule 4 Diagnosis at the Nodule Level 0=Unknown 1=benign or non-malignant disease 2= malignant, primary lung cancer 3 = malignant metastatic)	Nodule 4 Diagnosis Method at the Nodule Level 0 = unknown 1 = review of radiological images to show 2 years of stable nodule 2 = biopsy 3 = surgical resection 4 = progression	Nodule 5 Diagnosis at the Nodule Level 0=Unknown 1=benign or non-maligna disease 2= malignan primary lung cancer 3 = malignan metastatic)
LIDC-IDRI-0068	3	4	Head & Neck Cancer	3	4							
LIDC-IDRI-0071	3	1	Head & Neck	1	1							
LIDC-IDRI-0072	2	4	Lung Cancer	1	4							
LIDC-IDRI-0088	3	0	Uterine Cancer	0	0							
LIDC-IDRI-0090	2	3	NSCLC	2	3							
LIDC-IDRI-0091	3	4	urothelial carcinoma	3	4							
LIDC-IDRI-0100	3	1	Testis	1	1							
LIDC-IDRI-0118	3	0	Prostate	0	0							
LIDC-IDRI-0124	3	2	colon cancer	3	4							
LIDC-IDRI-0129	3	4	colon	3	4							
LIDC-IDRI-0135	3	4	Metastatic colon cancer	3	4							
LIDC-IDRI-0137	3	3	Basaloid squamous carcinoma	3	3							
LIDC-IDRI-0138	2	2	Lung Cancer	2	2							
LIDC-IDRI-0143	3	3	osteosarcoma	0	0							
LIDC-IDRI-0149	1	1	chondrosarcoma	1	1							
LIDC-IDRI-0159	3	2	thyroid carcinoma	3	4							
LIDC-IDRI-0161	3	2	pancreatic cancer	3	2							
LIDC-IDRI-0162	1	1	gallbladder	1	1							
LIDC-IDRI-0163	2	2	non small cell lung	2	2							
LIDC-IDRI-0164	3	3	colorectal cancer	3	4							
LIDC-IDRI-0165	3	4	vaginal cancer	3	4							
LIDC-IDRI-0166	2	2	stage III lung cancer	2	2							
LIDC-IDRI-0167	1	1	LGL (T cell lymphoma)	1	1							
LIDC-IDRI-0168	1	3	organizing pneumonia, also leukemia	1	3							
LIDC-IDRI-0169	3	4	granular cell tumor of the trachea	3	4							
LIDC-IDRI-0171	1	1	Lymphoma	1	1	1	4					
LIDC-IDRI-0173	0	0	bladder	0	0							
LIDC-IDRI-0174	2	3	nonsmall cell lung cancer	2	3	3	4					
LIDC-IDRI-0175	3	2	thyroid carcinoma	3	4							
LIDC-IDRI-0176	0	0	nonsmall cell lung cancer	0	0							
LIDC-IDRI-0178	1	1	melanoma	1	1							
LIDC-IDRI-0179	3	2	nonsmall cell lung cancer	3	2							
LIDC-IDRI-0180	1	1	LGL (T cell lymphoma)	1	1							

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

Ten entries for diagnosis, diagnosis method, and tumor site

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

LIDC-IDRI patient-level diagnosis data key

Armato III, Samuel G., McLennan, Geoffrey, Bidaut, Luc, McNitt-Gray, Michael F., Meyer, Charles R., Reeves, Anthony P., … Clarke, Laurence P. (2015). Data From LIDC-IDRI. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX

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	

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	-

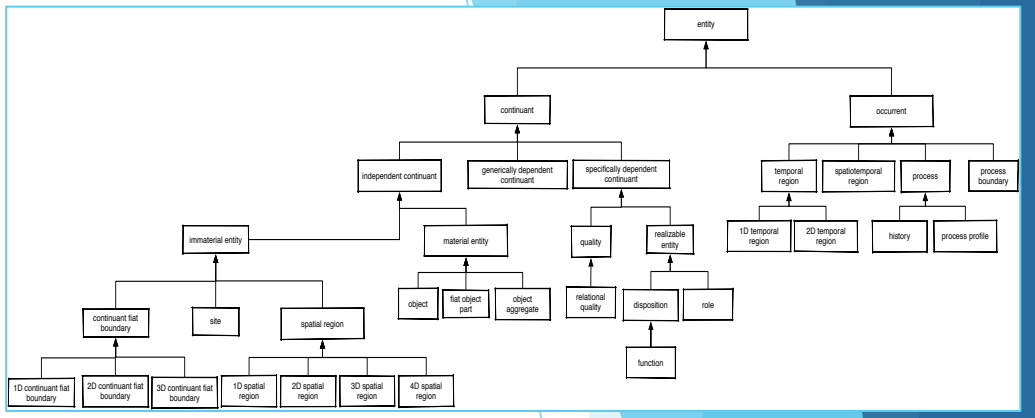
Grossberg A, Mohamed A, Elhalawani H, Bennett W, Smith K, Nolan T, Chamchod S, Kantor M, Browne T, Hutcheson K, Gunn G, Garden A, Frank S, Rosenthal D, Freymann J, Fuller C.(2017). **Data from Head and Neck Cancer CT Atlas**. The Cancer Imaging Archive. DOI: 10.7937/K9/TCIA.2017.umz8dv6s

Martin Vallières, Emily Kay-Rivest, Léo Jean Perrin, Xavier Liem, Christophe Furstoss, Nader Khaouam, Phuc Félix Nguyen-Tan, Chang-Shu Wang, Khalil Sultanem. (2017). **Data from Head-Neck-PET-CT**. The Cancer Imaging Archive. DOI: 10.7937/K9/TCIA.2017.8oje5q00

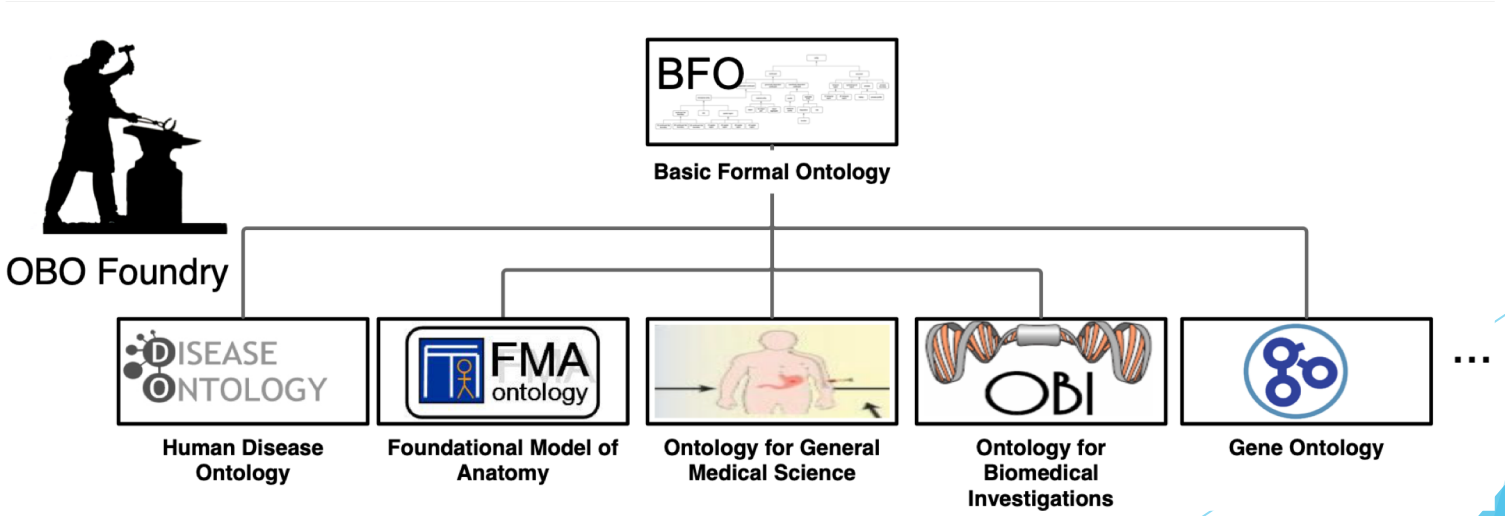
Semantic integration

- Integrate and manage data using shared representations that account for both explicit and implicit connections among the data across the source data sets
- Removing obstacles to:
  - Working with **different source representations for the same type of information**.
  - Connecting and interpreting **different types of data that are about the same phenomena**.
  - Combining **diverse data sets that are about the same individuals.**

Semantic integration with ontologies

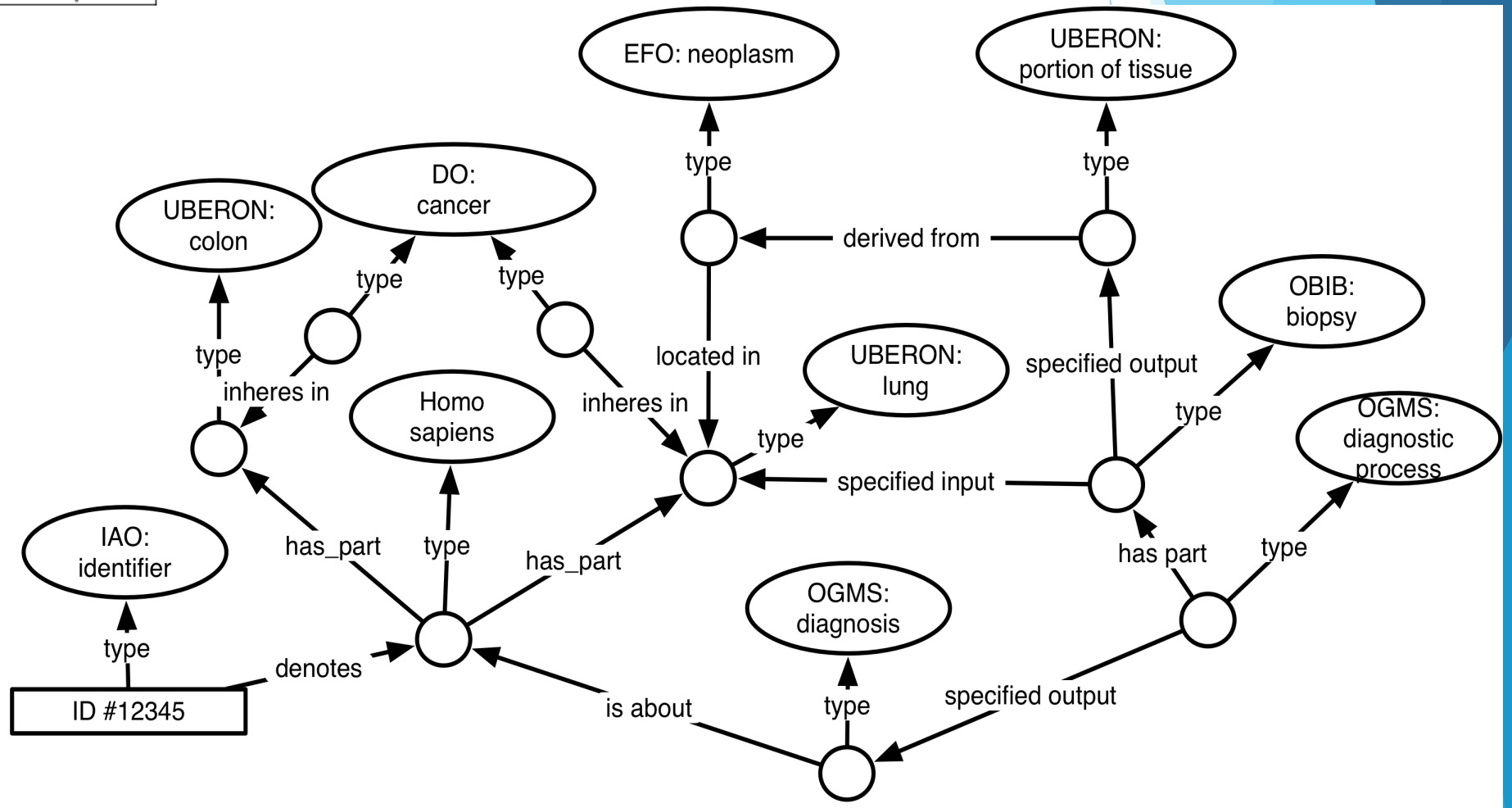


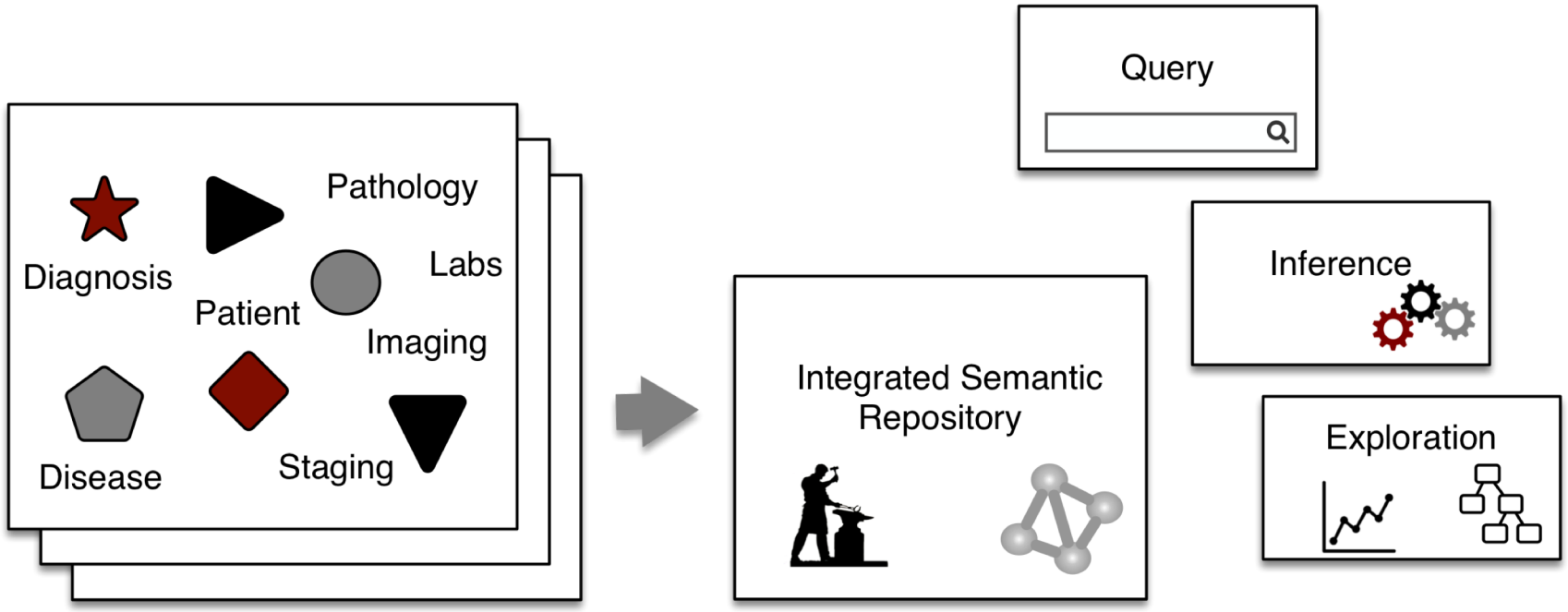
- Open Biomedical Ontologies Foundry
  - Shared design principles
  - Common upper level Basic Formal Ontology (BFO)
  - Consistent representation



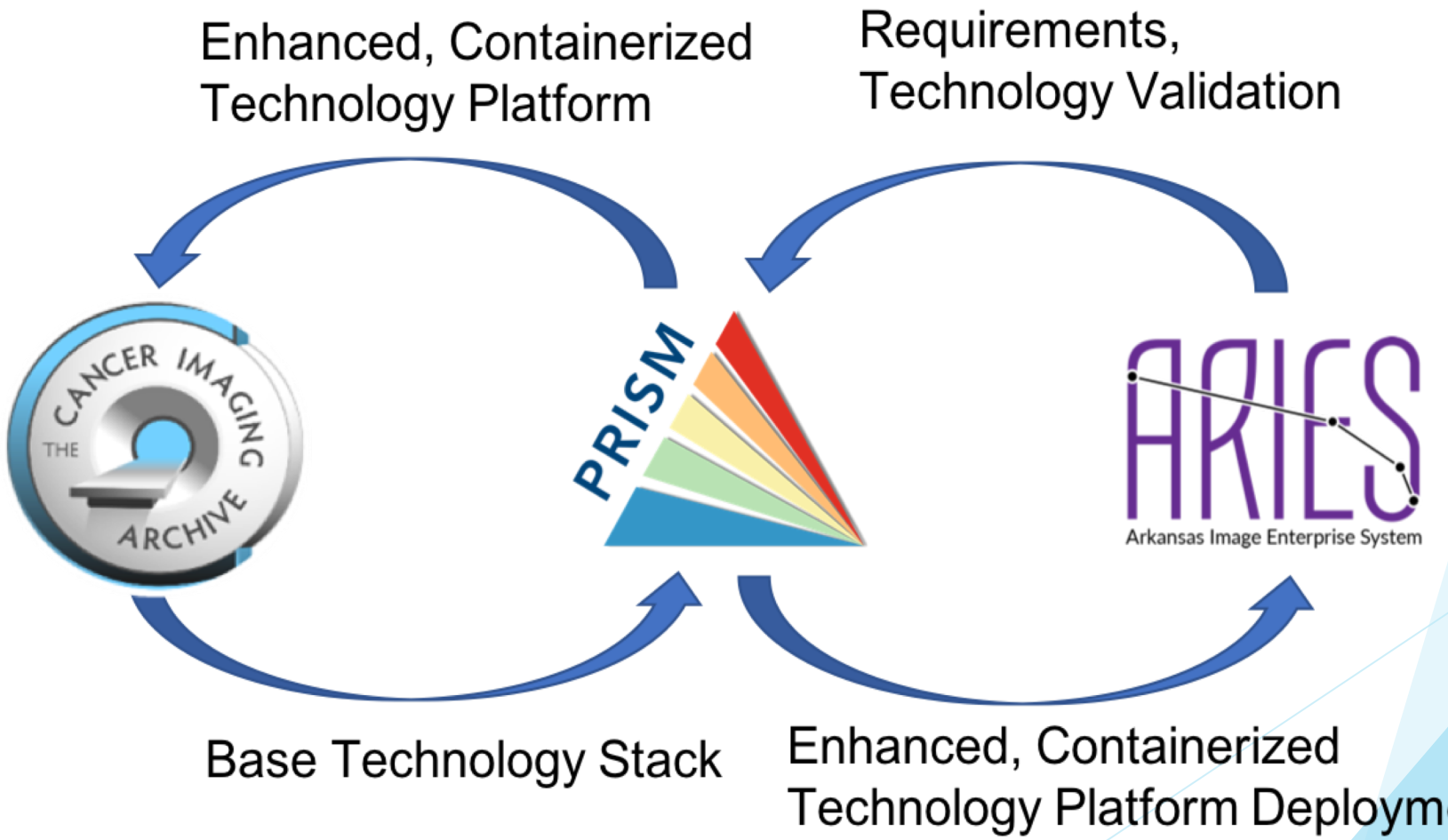
Diagnosis at the Patient Level	Diagnosis Method
0=Unknown	0 = unknown
1=benign or non-malignant disease	1 = review of radiological images to show 2 years of stable nodule
2= malignant, primary lung cancer	2 = biopsy
3 = malignant metastatic	3 = surgical resection
	4 = progression or response

12345	3	2	colon cancer



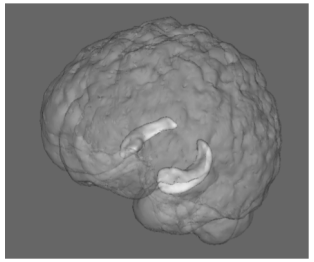

Pathology
Diagnosis
Labs
Patient
Imaging
Disease
Staging
Integrated Semantic Repository
Query
Inference
Exploration

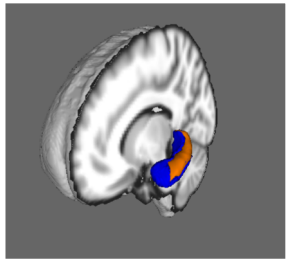
The Arkansas Image Enterprise System (ARIES) is a PRISM instance

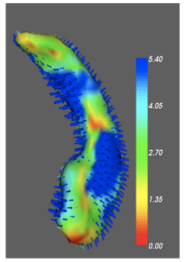


The Arkansas Image Enterprise System (ARIES) is a PRISM instance

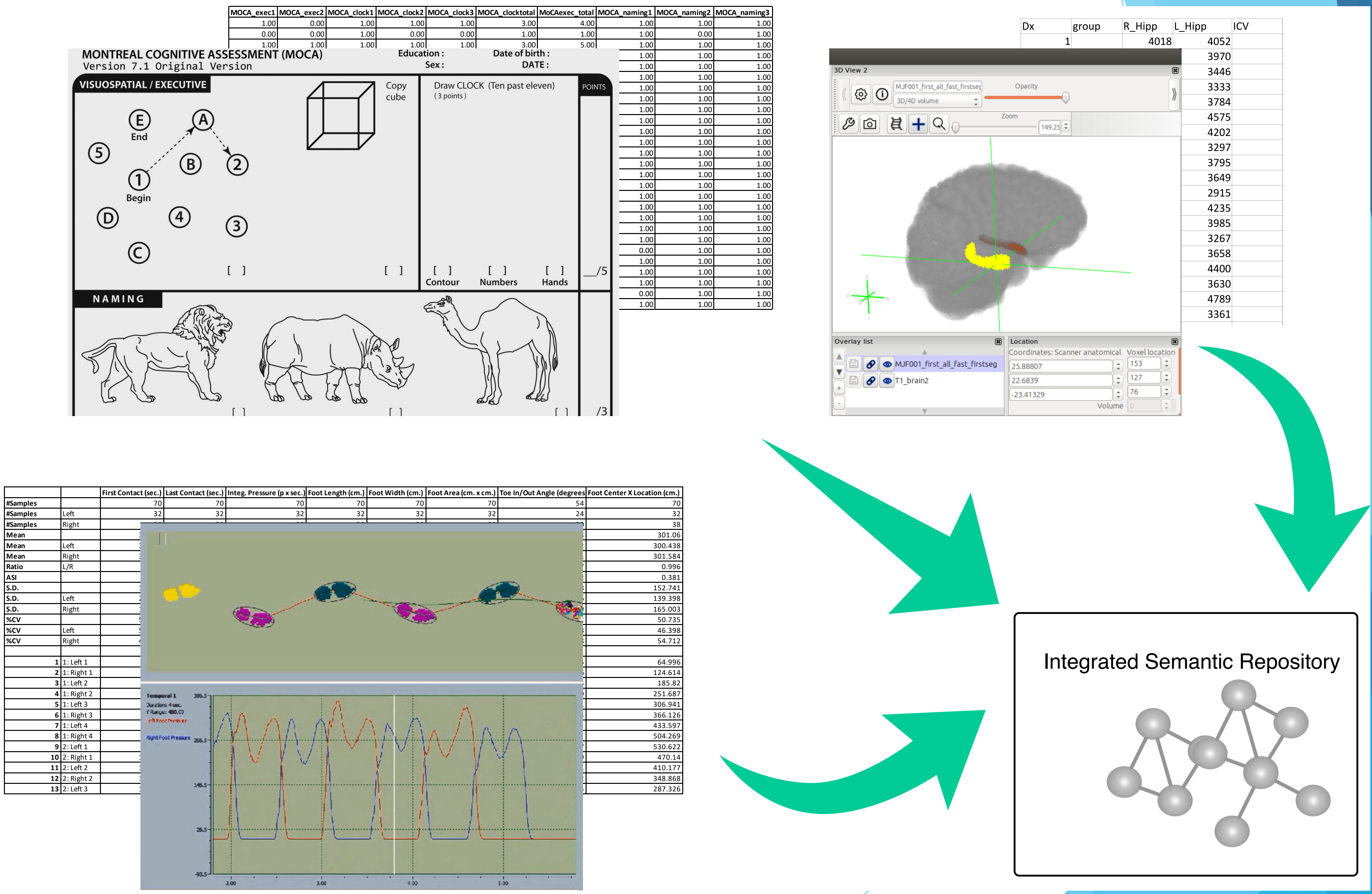
- Data from three collaborating investigative teams seeking to identify common pathways of neurodegeneration.
- Pilot data from three unique study cohorts diagnosed with Parkinson's disease (PD), Mild Cognitive Impairment (MCI), or Cancer-Related Cognitive Impairment (CRCI).
- These datasets include **images and image-derived features**, **motor assessments**, **cognitive assessments**, **clinical rating scales**, **demographics**, and **clinical data**.
- Pilot test case: linking image-derived measures of hippocampal volumes to a diverse set of cognitive assessment results.

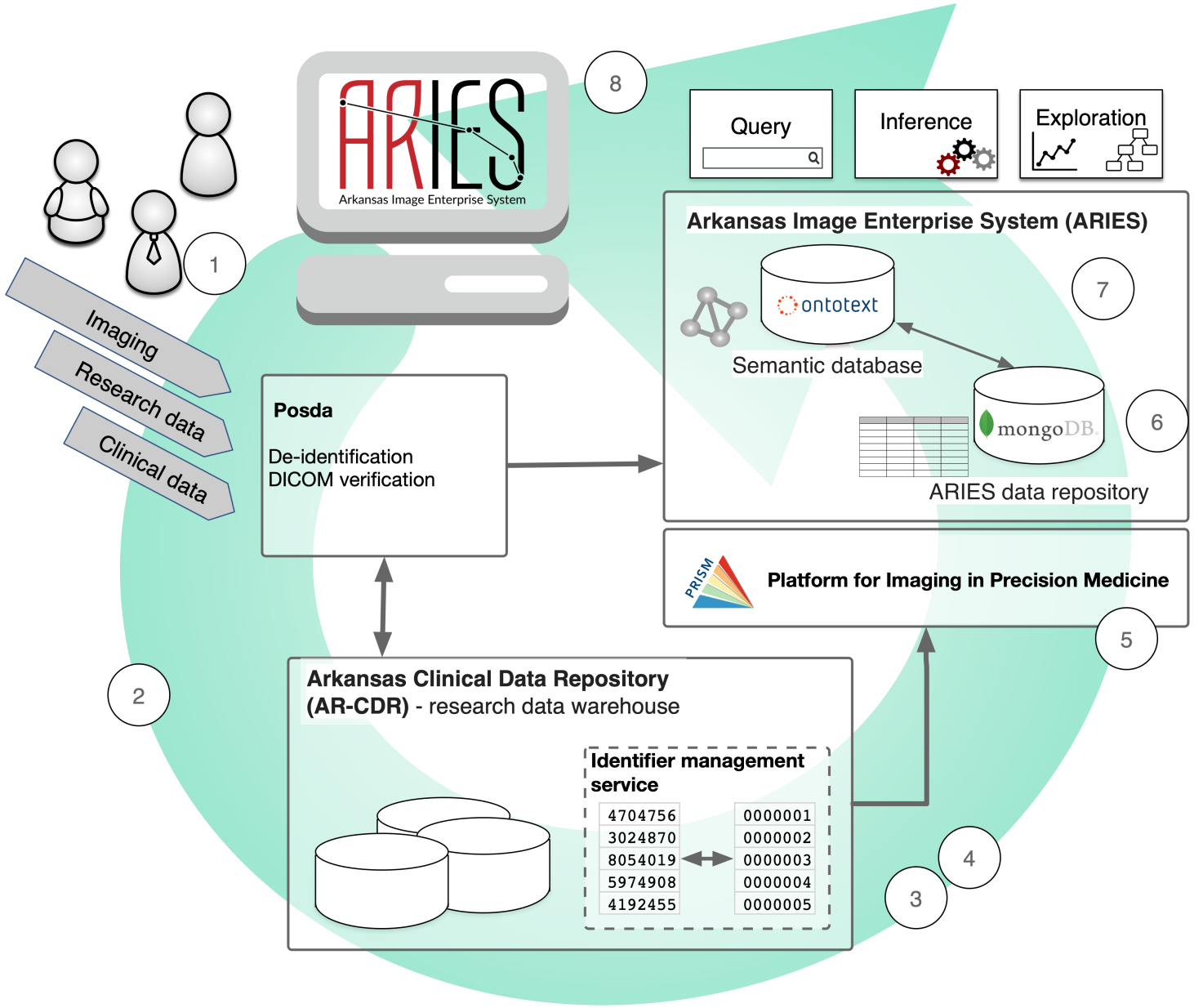


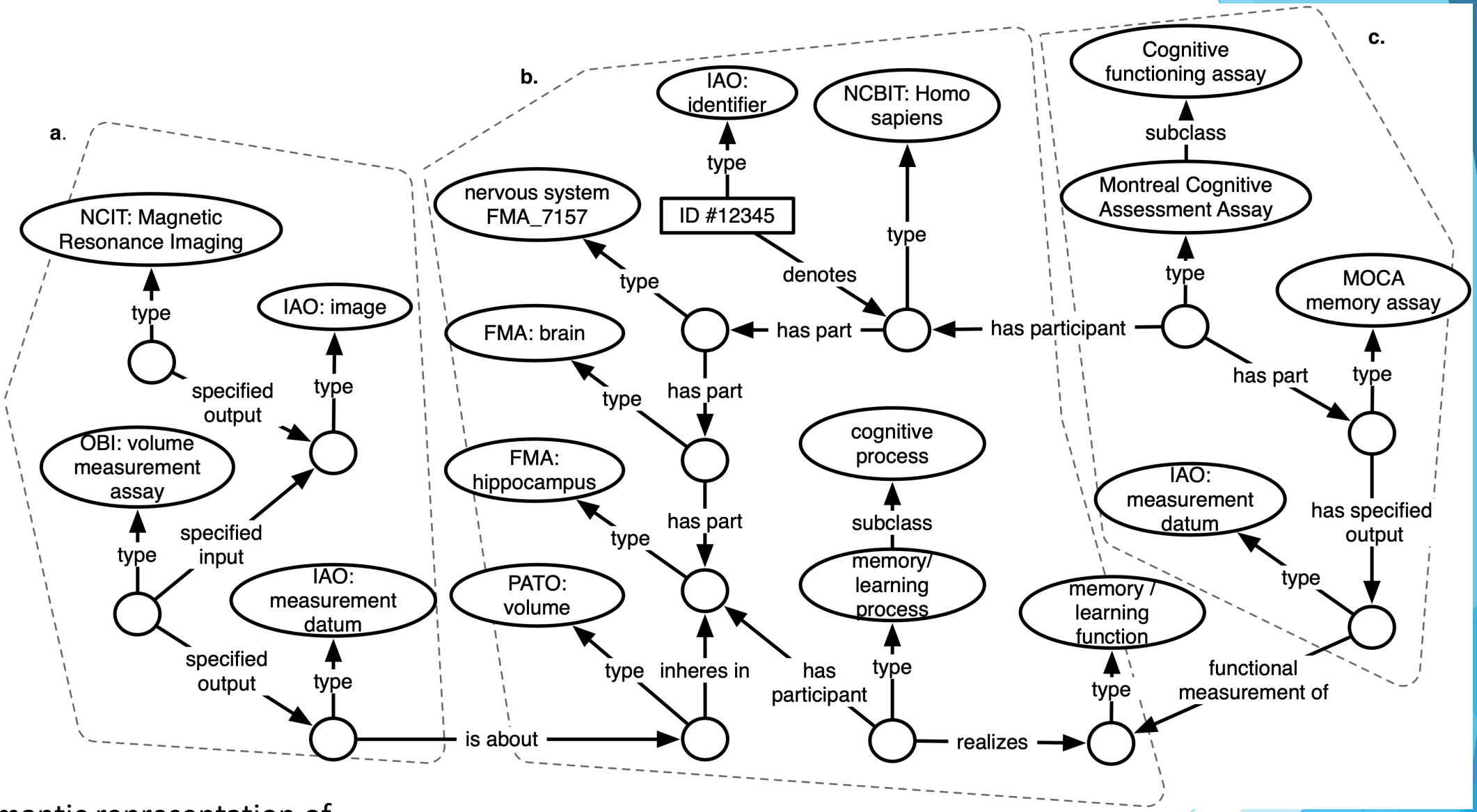




		Cohort 1	Cohort 2	Cohort 3
Imaging & derived features	MRI	■	■	■
	MRI-derived imaging features	■	■	■
	Gait video	■	■	■
	Gait-derived features	■	■	■
	Resting Electroencephalography (EEG)	■	■	■
Motor assessments	Gait-assessment floor mat	■	■	■
	Accelerometry from wearable body sensors	■	■	■
	digitized gloves	■	■	■
	handwriting/drawing assessments on a digitizing tablet	■	■	■
Cognitive assessments	Montreal Cognitive Assessment	■	■	■
	St Louis University Mental Status	■	■	■
	Repeatable Battery for Assessment of Neuropsychological Status	■	■	■
	SCales for Outcomes in PArkinson's disease - COGnition	■	■	■
	Neuropsychological Assessment Battery	■	■	■
	Frontal Assessment Battery	■	■	■
Clinical rating scales	etc	■	■	■
Demographics		■	■	■
Clnical		■	■	■


MONTREAL COGNITIVE ASSESSMENT (MOCA)
Version 7.1 Original Version
VISUOSPATIAL / EXECUTIVE
NAMING
3D View 2
Overlay list
Location
Dx
group
R_Hipp
L_Hipp
ICV
Integrated Semantic Repository


ARIES
Arkansas Image Enterprise System
8
Query
Inference
Exploration
Arkansas Image Enterprise System (ARIES)
7
ontotext
Semantic database
mongoDB.
6
ARIES data repository
1
Imaging
Research data
Clinical data
Posda
De-identification
DICOM verification
PRISM
Platform for Imaging in Precision Medicine
5
2
Arkansas Clinical Data Repository
(AR-CDR) - research data warehouse
Identifier management service
4704756
3024870
8054019
5974908
4192455
0000001
0000002
0000003
0000004
0000005
4
3



Semantic representation of

a) an image capture and image-derived volume measure

b) the subject

c) a cognitive assessment

Ongoing work

- Ongoing representation & integration of TCIA collections data
- PRISM semantic use cases
- Semantic query interface development
- Testing and use of PRISM platform within ARIES project

Acknowledgements

This work has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health under Contract No. HHSN261200800001E, subcontract 16X011 and grants: U01CA187013, U24CA215109, 3U24CA215109-02S1.

The PRISM Team

UAMS
University of Arkansas for Medical Sciences

- **Fred Prior, PhD**
- Jonathan Bona, PhD
- Kirk Smith
- Lawrence Tarbox, PhD
- Mathias Brochhausen, PhD
- Roosevelt Dobbins
- Tracy Nolan
- William Bennett

EMORY UNIVERSITY

- **Ashish Sharma, PhD**
- Annie Gu
- Mohanapriya Narapareddy
- Monjoy Saha, PhD
- Pradeeban Kathiravelu, PhD

Stony Brook University

- **Joel Saltz, MD, PhD**
- Erich Bremer
- Rajrisi Gupta MD
- Tahsin Kurc, PhD
- Tammy DiPrima

QARC
QUALITY ASSURANCE REVIEW CENTER

- TJ Fitzgerald, MD
- Fran Laurie