Standardization of (Imaging) Data Formats

Lessons Learned

Practical Big Data Workshop

David Clunie (dclunie@dclunie.com)

Pixelmed Publishing, LLC.

Conflict of Interest

- Grants/Research Support: NCI (Essex, BWH)
- Consulting Fees: MDDX, Carestream, GE, Curemetrix, NEMA
- Editor of DICOM Standard (NEMA/MITA Contractor)
- Other: Owner of PixelMed Publishing

- None directly relevant to topic of this presentation

DICOM and Big Data

- DICOM data elements
- DICOM coded concepts and values
- Actually used count
- Single Attribute vs. structured context
- Identification, acquisition (incl. workflow), derivation (incl. quantitative parametric maps, ROIs, measurements, categorical)
- non-image DICOM: SEG, PS, SR, RTSS
- DICOM RT-specific – 1st and 2nd generation

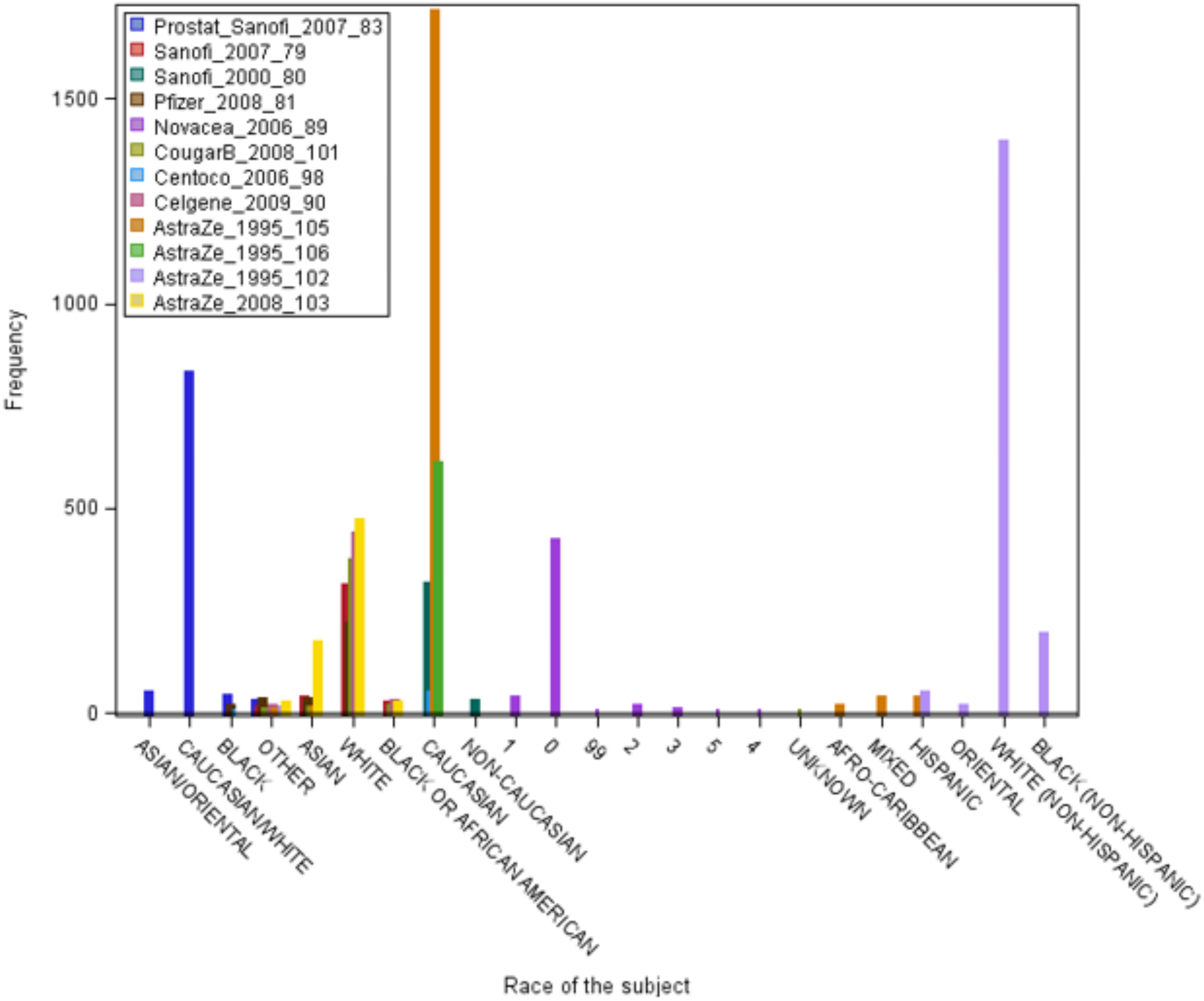
DICOM Elements Actually Used

- Defined versus used
  - what is defined in various image and non-image IODs
  - including “enhanced” family images (much more detail)
  - what is actually encountered in clinical practice
- 2006 review of large oncology clinical trial archive
  - standard had 2527 data elements
  - 618 data elements seen in archive
  - in more than 25% of images, 125 data elements
  - in more than 90% of images, 54 data elements
  - admittedly a biased sample CT >> MR >> NM, PT, CR, DX

Standard Values for Attributes

- “Common Data Elements” are not enough for big data
- Need “Common Value Sets” for those CDEs too
- Legacy objects – few enumerated values and defined terms
- Enhanced family – many more, but less often used
- Codes
  - from external vocabulary, e.g., SNOMED
  - defined by DICOM (PS3.16 Annex D)
- Codes used for
  - anatomy, etc. in newer images
  - DICOM SR
  - worklists, acquisition context and protocols

Figure 5: Race Group Names Bar Graph



Gene Lightfoot. Project DataSphere – Reviewing Data and Quality. SAS. 2017.

Codes, Controlled Terminology

- General need, and in an RT-context
- Anatomy – SNOMED, FMA – could use for Organs at Risk
- Regions for specific purposes, e.g., GTV
  - code or string?
  - poor DICOM RTSS (implementation) precedent – not even a code for GTV in DICOM !@#$
  - could easily add SNOMED to DICOM context group
- Recent CPs to improve RTSS and align with Segmentation codes – CP 1287, 1314, 1586

Codes for Irradiated Volumes

- E.g., SNOMED Irradiated Volume concepts
  - (R-429E0, SRT, “Gross tumor volume”)
  - (R-429EB, SRT, “Clinical target volume”)
  - (R-429EC, SRT, “Planning target volume”)
- Being added in Sup 147 “Prescription and Segment Annotation”
  - in CID SUP147070 “Radiotherapy Targets”
  - 2nd generation, status is frozen draft for trial implementation
  - defines yet another RT-specific annotation IOD that doesn’t re-use non-RT objects (such as DICOM SR)
  - not back-ported to define for use in RTSS

Efforts to Standardize RT Names

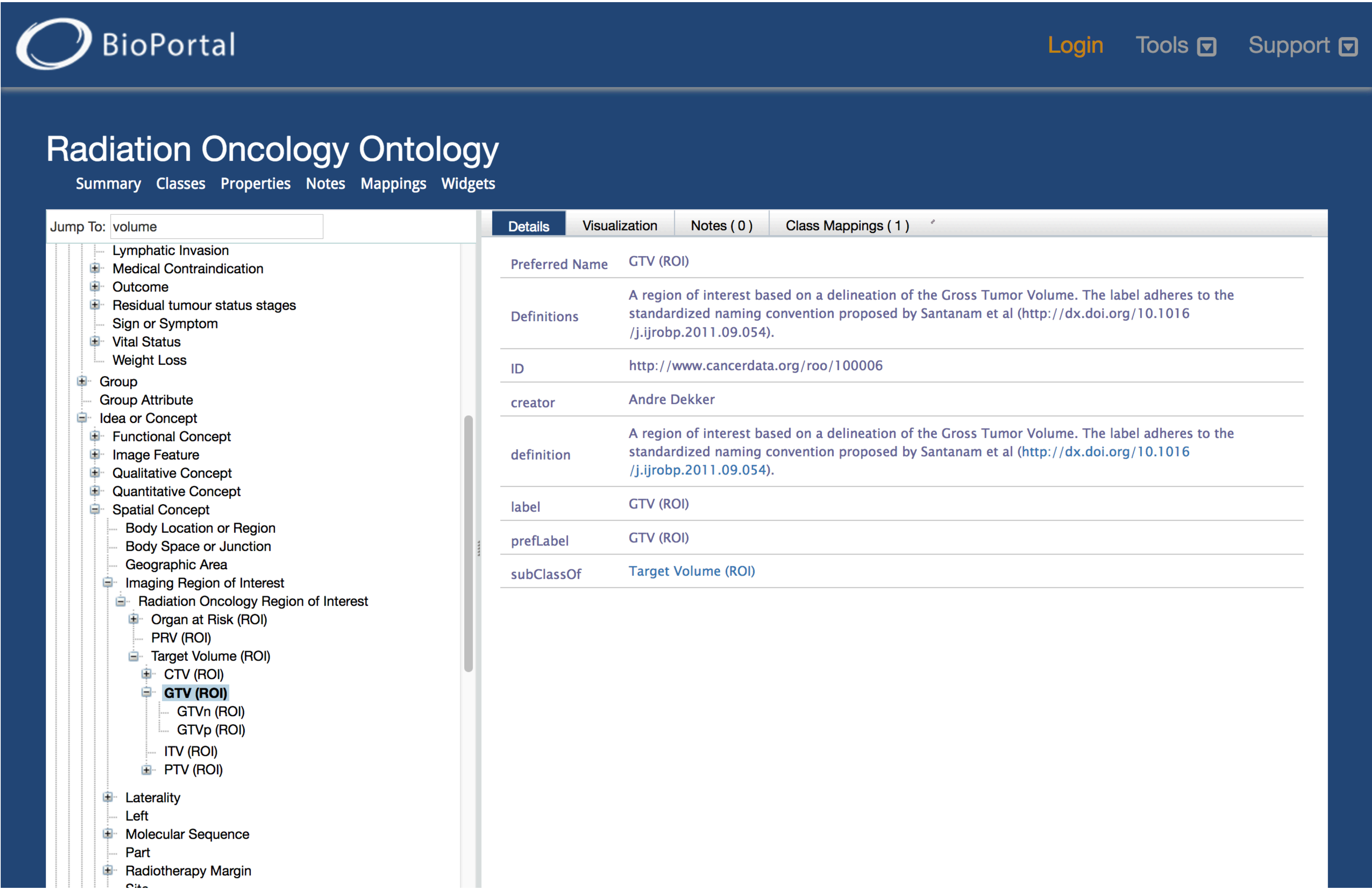
- Santanam et al. Standardizing Naming Conventions in Radiation Oncology. 2012. doi:10.1016/j.ijrobp.2011.09.054
- Miller. A Rational Informatics-enabled approach to Standardised Nomenclature of Contours and Volumes in Radiation Oncology Planning. 2014. http://ojs.jroi.org/index.php/jroi/article/view/22
- Denton et al. Guidelines for treatment naming in radiation oncology. 2016. doi:10.1120/jacmp.v17i2.5953
- AAPM TG 263 – Standardizing Nomenclature for Radiation Therapy
- NRG Structure Name Library
- Danger of constructing string names with embedded syntax versus true codes and ontologies

Radiation Oncology Ontology

- “aims to cover the radiation oncology domain with a strong focus on re-using existing ontologies”

- https://www.cancerdata.org/roo-information
- http://bioportal.bioontology.org/ontologies/ROO
- https://github.com/RadiationOncologyOntology/ROO

- ? add as new Coding Scheme to DICOM
- ? use codes from wherever re-used concepts came from
- not using SNOMED since not free (for non-DICOM folks)
- Open Source – Apache License
- Distributed as an OWL file



Radiation Oncology Ontology

Jump To: volume

- Lymphatic Invasion
- Medical Contraindication
- Outcome
- Residual tumour status stages
- Sign or Symptom
- Vital Status
- Weight Loss

- Group
- Group Attribute
- Idea or Concept
  - Functional Concept
  - Image Feature
  - Qualitative Concept
  - Quantitative Concept
  - Spatial Concept
    - Body Location or Region
    - Body Space or Junction
    - Geographic Area
    - Imaging Region of Interest
      - Radiation Oncology Region of Interest
        - Organ at Risk (ROI)
        - PRV (ROI)
        - Target Volume (ROI)
          - CTV (ROI)
          - **GTV (ROI)**
            - GTVn (ROI)
            - GTVp (ROI)
          - ITV (ROI)
          - PTV (ROI)
    - Laterality
    - Left
    - Molecular Sequence
    - Part
    - Radiotherapy Margin

Preferred Name	GTV (ROI)
Definitions	A region of interest based on a delineation of the Gross Tumor Volume. The label adheres to the standardized naming convention proposed by Santanam et al (http://dx.doi.org/10.1016/j.ijrobp.2011.09.054).
ID	http://www.cancerdata.org/roo/100006
creator	Andre Dekker
definition	A region of interest based on a delineation of the Gross Tumor Volume. The label adheres to the standardized naming convention proposed by Santanam et al (http://dx.doi.org/10.1016/j.ijrobp.2011.09.054).
label	GTV (ROI)
prefLabel	GTV (ROI)
subClassOf	Target Volume (ROI)

Structural Context

- The values of a data element extracted from its “context” may be meaningless
- Multiple different “volumes” in same “row” of extracted table if insufficient “context”
- E.g., “volume” = “12.34” “mm3”
- Volume of what?
- Measured how?
- Modifiers: mean, max, peak (e.g., SUV)
- Pre-coordinated vs. post-coordinated

N.3.4 Left Ventricle Volumes and Ejection Fraction

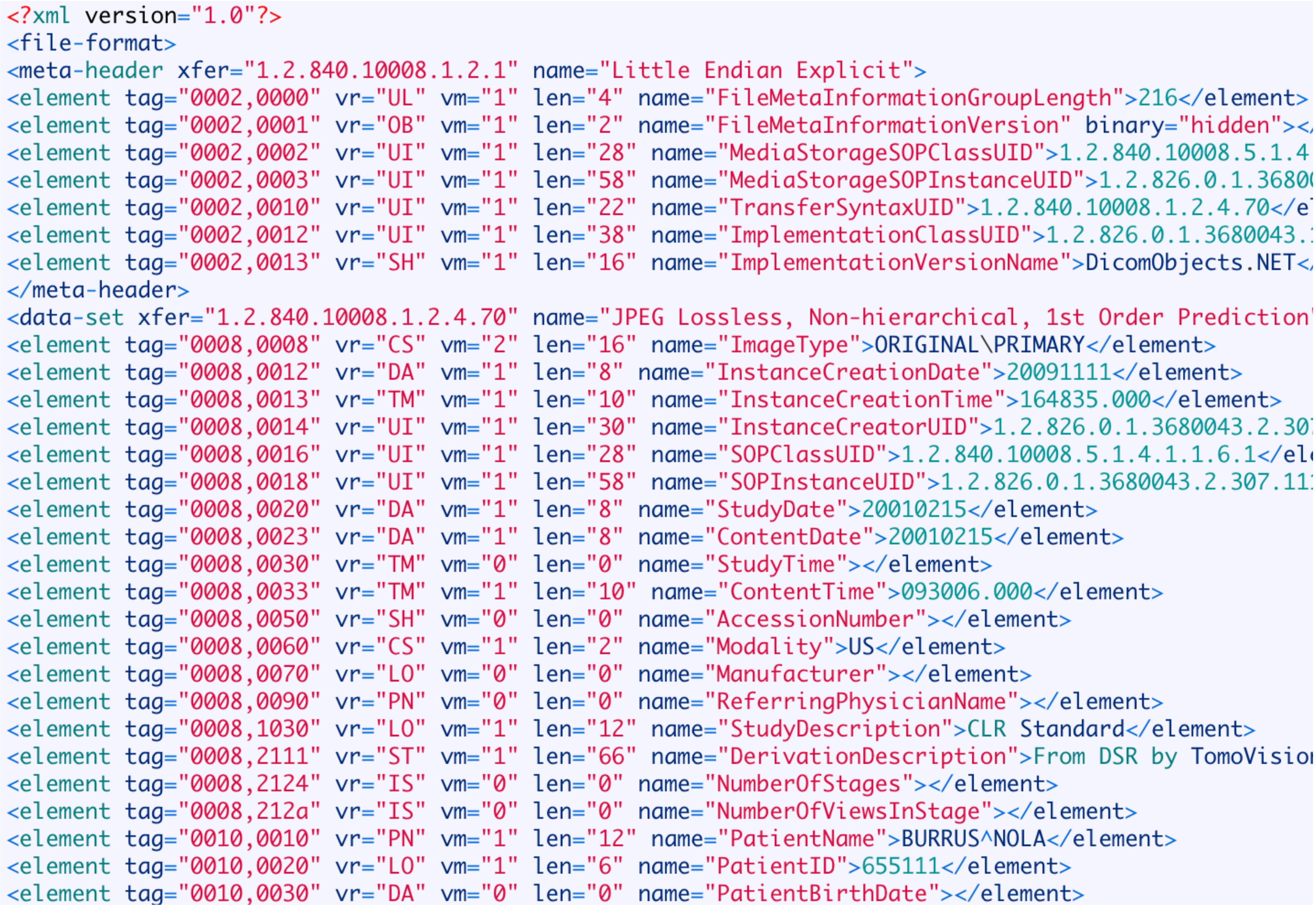
Name of ASE Concept	Base Measurement Concept Name	Concept or Acquisition Context Modifiers
Left Ventricular End Diastolic Volume	(18026-5, LN, "Left Ventricular End Diastolic Volume")	
Left Ventricular End Diastolic Volume by Teichholz Method	(18026-5, LN, "Left Ventricular End Diastolic Volume")	(G-C036, SRT, "Measurement Method") = (125209, DCM, "Teichholz")
Left Ventricular End Diastolic Volume by 2-D Single Plane by Method of Disks (4-Chamber)	(18026-5, LN, "Left Ventricular End Diastolic Volume")	(111031, DCM, "Image View") = (G-A19C, SRT, "Apical Four Chamber") (G-C036, SRT, "Measurement Method") = (125208, DCM, "Method of Disks, Single Plane")
Left Ventricular End Diastolic Volume by 2-D Biplane by Method of Disks	(18026-5, LN, "Left Ventricular End Diastolic Volume")	(G-C036, SRT, "Measurement Method") = (125207, DCM, "Method of Disks, Biplane")
Left Ventricular End Systolic Volume	(18148-7, LN, "Left Ventricular End Systolic Volume")	
Left Ventricular End Systolic Volume by Teichholz Method	(18148-7, LN, "Left Ventricular End Systolic Volume")	(G-C036, SRT, "Measurement Method") = (125209, DCM, "Teichholz")
Left Ventricular End Systolic Volume by 2D Single Plane by Method of Disks (4-Chamber)	(18148-7, LN, "Left Ventricular End Systolic Volume")	(111031, DCM, "Image View") = (G-A19C, SRT, "Apical Four Chamber") (G-C036, SRT, "Measurement Method") = (125208, DCM, "Method of Disks, Single Plane")
Left Ventricular End Systolic Volume by 2-D Biplane by Method of Disks	(18148-7, LN, "Left Ventricular End Systolic Volume")	(G-C036, SRT, "Measurement Method") = (125207, DCM, "Method of Disks, Biplane")
Left Ventricular EF	(18043-0, LN, "Left Ventricular Ejection Fraction")	
Left Ventricular EF by Teichholz Method	(18043-0, LN, "Left Ventricular Ejection Fraction")	(G-C036, SRT, "Measurement Method") = (125209, DCM, "Teichholz")

Push or Pull

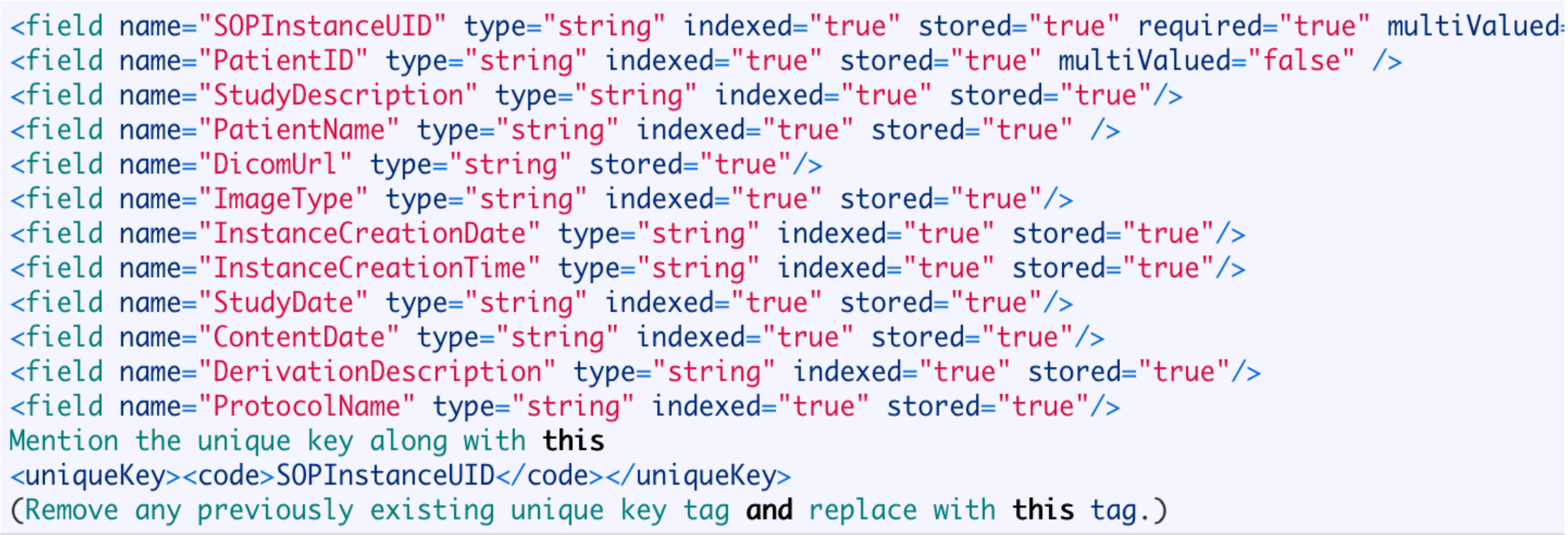
- Pull
  - known inputs into known fields in “template” or “schema”
- Push
  - recognized input into known fields
  - any other input into unknown fields
- Predefined “schema” vs. adaptive data modeling
- Name-value pairs, RDF tuples, mixture
- Automated ETL rather than hand-mapped
- How do (input) standards help?
  - what to expect
  - what it actually “means” (versus “lexical semantics”)

DICOM Big Data Example

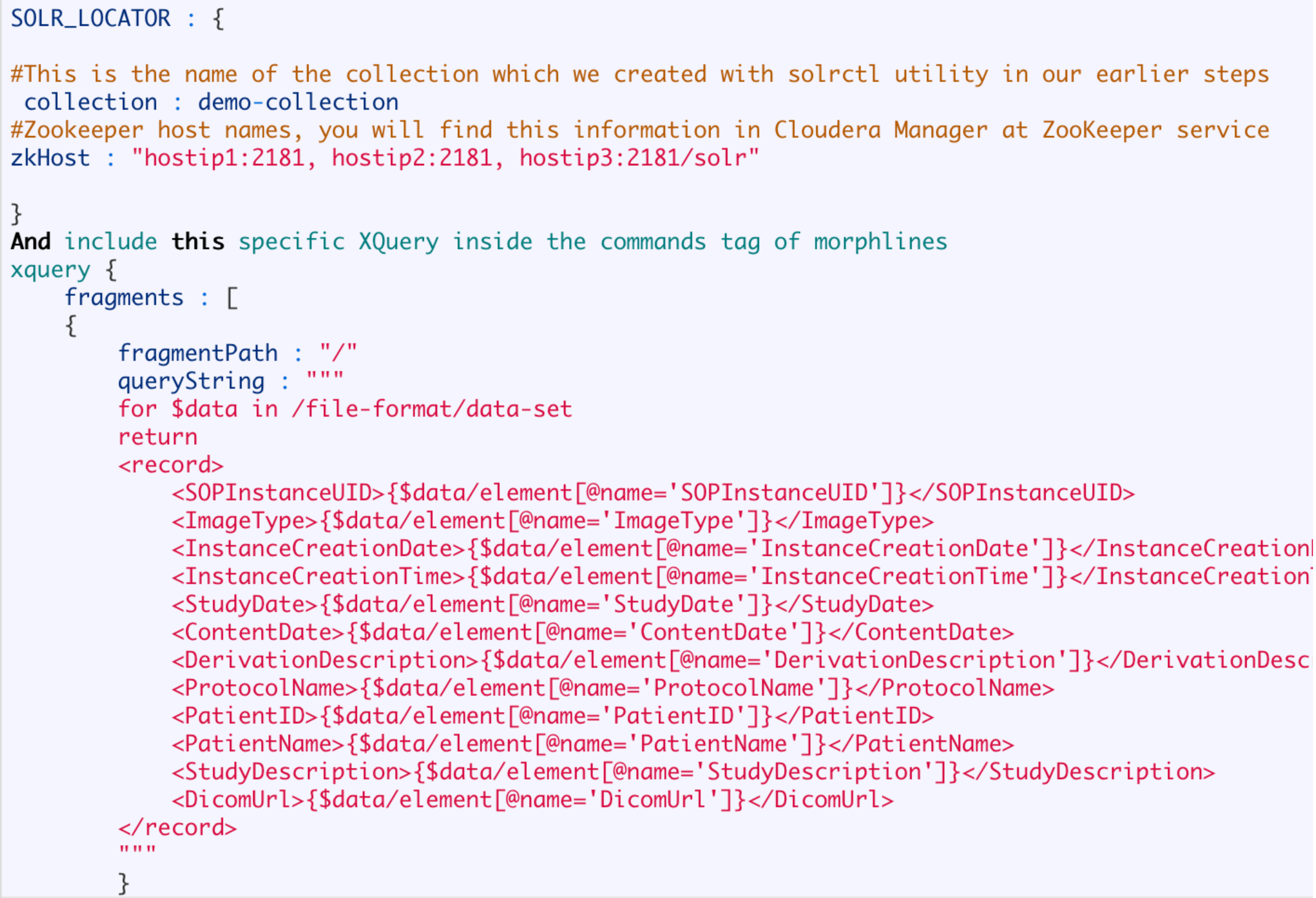
- https://blog.cloudera.com/blog/2016/05/how-to-process-and-index-medical-images-with-apache-hadoop-and-apache-solr/
- dcmtk dcm2xml
- Apache Solr schema.xml file
- Morphlines configuration file
- MapReduceIndexerTool
- Hue for view/search



```
<?xml version="1.0"?>
<file-format>
<meta-header xfer="1.2.840.10008.1.2.1" name="Little Endian Explicit">
<element tag="0002,0000" vr="UL" vm="1" len="4" name="FileMetaInformationGroupLength">216</element>
<element tag="0002,0001" vr="OB" vm="1" len="2" name="FileMetaInformationVersion" binary="hidden"></
<element tag="0002,0002" vr="UI" vm="1" len="28" name="MediaStorageSOPClassUID">1.2.840.10008.5.1.4
<element tag="0002,0003" vr="UI" vm="1" len="58" name="MediaStorageSOPInstanceUID">1.2.826.0.1.3680
<element tag="0002,0010" vr="UI" vm="1" len="22" name="TransferSyntaxUID">1.2.840.10008.1.2.4.70</e
<element tag="0002,0012" vr="UI" vm="1" len="38" name="ImplementationClassUID">1.2.826.0.1.3680043.
<element tag="0002,0013" vr="SH" vm="1" len="16" name="ImplementationVersionName">DicomObjects.NET</
</meta-header>
<data-set xfer="1.2.840.10008.1.2.4.70" name="JPEG Lossless, Non-hierarchical, 1st Order Prediction"
<element tag="0008,0008" vr="CS" vm="2" len="16" name="ImageType">ORIGINAL\PRIMARY</element>
<element tag="0008,0012" vr="DA" vm="1" len="8" name="InstanceCreationDate">20091111</element>
<element tag="0008,0013" vr="TM" vm="1" len="10" name="InstanceCreationTime">164835.000</element>
<element tag="0008,0014" vr="UI" vm="1" len="30" name="InstanceCreatorUID">1.2.826.0.1.3680043.2.30
<element tag="0008,0016" vr="UI" vm="1" len="28" name="SOPClassUID">1.2.840.10008.5.1.4.1.1.6.1</el
<element tag="0008,0018" vr="UI" vm="1" len="58" name="SOPInstanceUID">1.2.826.0.1.3680043.2.307.11
<element tag="0008,0020" vr="DA" vm="1" len="8" name="StudyDate">20010215</element>
<element tag="0008,0023" vr="DA" vm="1" len="8" name="ContentDate">20010215</element>
<element tag="0008,0030" vr="TM" vm="0" len="0" name="StudyTime"></element>
<element tag="0008,0033" vr="TM" vm="1" len="10" name="ContentTime">093006.000</element>
<element tag="0008,0050" vr="SH" vm="0" len="0" name="AccessionNumber"></element>
<element tag="0008,0060" vr="CS" vm="1" len="2" name="Modality">US</element>
<element tag="0008,0070" vr="LO" vm="0" len="0" name="Manufacturer"></element>
<element tag="0008,0090" vr="PN" vm="0" len="0" name="ReferringPhysicianName"></element>
<element tag="0008,1030" vr="LO" vm="1" len="12" name="StudyDescription">CLR Standard</element>
<element tag="0008,2111" vr="ST" vm="1" len="66" name="DerivationDescription">From DSR by TomoVisio
<element tag="0008,2124" vr="IS" vm="0" len="0" name="NumberOfStages"></element>
<element tag="0008,212a" vr="IS" vm="0" len="0" name="NumberOfViewsInStage"></element>
<element tag="0010,0010" vr="PN" vm="1" len="12" name="PatientName">BURRUS^NOLA</element>
<element tag="0010,0020" vr="LO" vm="1" len="6" name="PatientID">655111</element>
<element tag="0010,0030" vr="DA" vm="0" len="0" name="PatientBirthDate"></element>
```



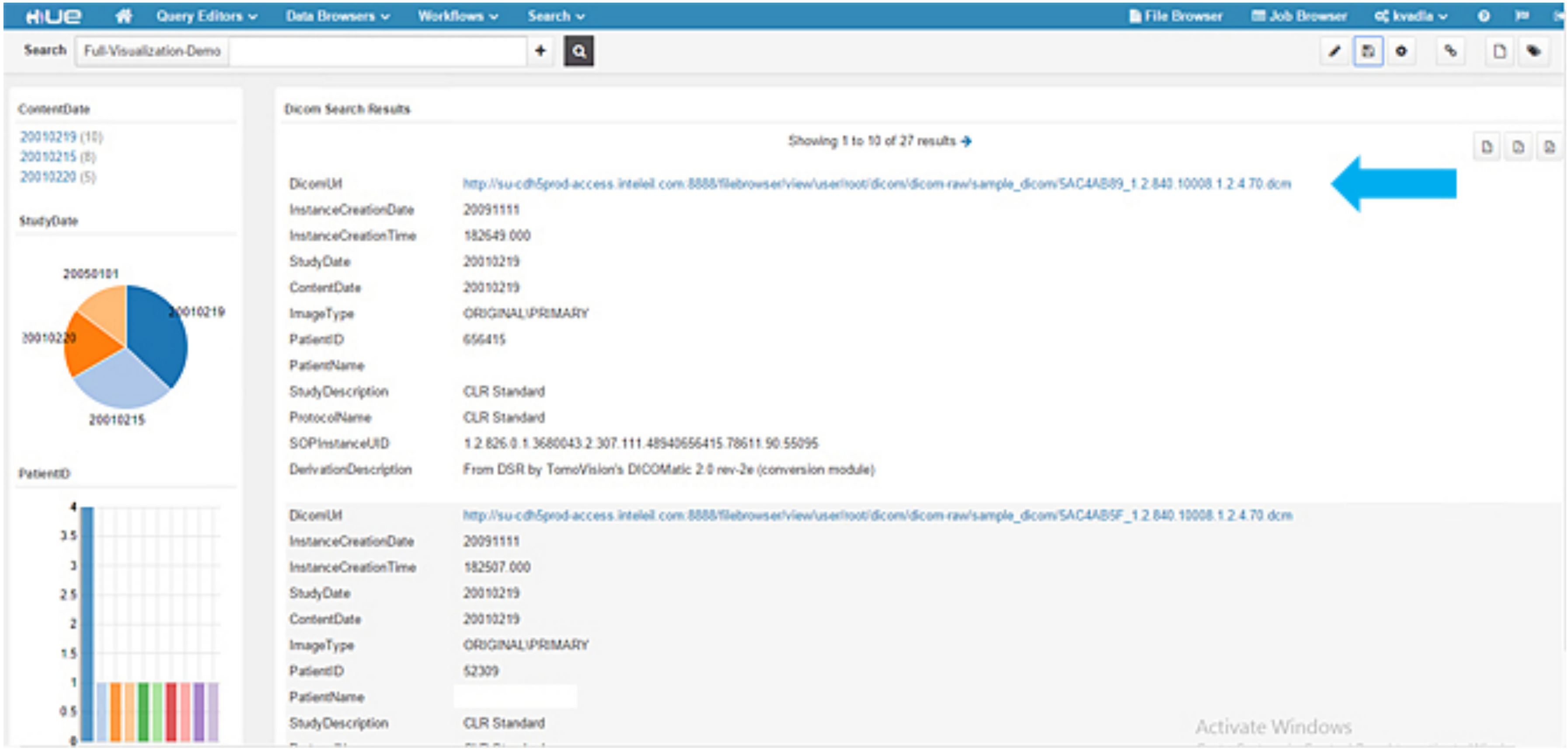
```
<field name="SOPInstanceUID" type="string" indexed="true" stored="true" required="true" multiValued=
<field name="PatientID" type="string" indexed="true" stored="true" multiValued="false" />
<field name="StudyDescription" type="string" indexed="true" stored="true"/>
<field name="PatientName" type="string" indexed="true" stored="true" />
<field name="DicomUrl" type="string" stored="true"/>
<field name="ImageType" type="string" indexed="true" stored="true"/>
<field name="InstanceCreationDate" type="string" indexed="true" stored="true"/>
<field name="InstanceCreationTime" type="string" indexed="true" stored="true"/>
<field name="StudyDate" type="string" indexed="true" stored="true"/>
<field name="ContentDate" type="string" indexed="true" stored="true"/>
<field name="DerivationDescription" type="string" indexed="true" stored="true"/>
<field name="ProtocolName" type="string" indexed="true" stored="true"/>
Mention the unique key along with this
<uniqueKey><code>SOPInstanceUID</code></uniqueKey>
(Remove any previously existing unique key tag and replace with this tag.)
```



```
SOLR_LOCATOR : {

#This is the name of the collection which we created with solrctl utility in our earlier steps
 collection : demo-collection
#Zookeeper host names, you will find this information in Cloudera Manager at ZooKeeper service
zkHost : "hostip1:2181, hostip2:2181, hostip3:2181/solr"

}
And include this specific XQuery inside the commands tag of morphlines
xquery {
    fragments : [
    {
        fragmentPath : "/"
        queryString : """
        for $data in /file-format/data-set
        return
        <record>
            <SOPInstanceUID>{$data/element[@name='SOPInstanceUID']}</SOPInstanceUID>
            <ImageType>{$data/element[@name='ImageType']}</ImageType>
            <InstanceCreationDate>{$data/element[@name='InstanceCreationDate']}</InstanceCreation
            <InstanceCreationTime>{$data/element[@name='InstanceCreationTime']}</InstanceCreation
            <StudyDate>{$data/element[@name='StudyDate']}</StudyDate>
            <ContentDate>{$data/element[@name='ContentDate']}</ContentDate>
            <DerivationDescription>{$data/element[@name='DerivationDescription']}</DerivationDesc
            <ProtocolName>{$data/element[@name='ProtocolName']}</ProtocolName>
            <PatientID>{$data/element[@name='PatientID']}</PatientID>
            <PatientName>{$data/element[@name='PatientName']}</PatientName>
            <StudyDescription>{$data/element[@name='StudyDescription']}</StudyDescription>
            <DicomUrl>{$data/element[@name='DicomUrl']}</DicomUrl>
        </record>
        """
        }
```



ContentDate

20010219 (10)
20010215 (8)
20010220 (5)

Dicom Search Results

Showing 1 to 10 of 27 results

DicomUrl	http://su-cdh5prod-access.inteleil.com:8888/filebrowser/view/user/root/dicom/dicom-raw/sample_dicom/5AC4AB89_1.2.840.10008.1.2.4.70.dcm
InstanceCreationDate	20091111
InstanceCreationTime	182549.000
StudyDate	20010219
ContentDate	20010219
ImageType	ORIGINAL\PRIMARY
PatientID	656415
PatientName	
StudyDescription	CLR Standard
ProtocolName	CLR Standard
SOPInstanceUID	1.2.826.0.1.3680043.2.307.111.48940656415.78611.90.55095
DerivationDescription	From DSR by TomoVision's DICOMatic 2.0 rev-2e (conversion module)

DicomUrl	http://su-cdh5prod-access.inteleil.com:8888/filebrowser/view/user/root/dicom/dicom-raw/sample_dicom/5AC4AB5F_1.2.840.10008.1.2.4.70.dcm
InstanceCreationDate	20091111
InstanceCreationTime	182507.000
StudyDate	20010219
ContentDate	20010219
ImageType	ORIGINAL\PRIMARY
PatientID	52309
PatientName	
StudyDescription	CLR Standard

Other DICOM Big Data Examples

- Hadoop – Hbase – http://coders-log.blogspot.com/2008/10/hadoop.html
- Hadoop – Mazurek et al. *Medical data preservation at scale*. 2015. https://tnc15.terena.org/core/presentation/108
- Hadoop – Hbase – bulk data – Bao et al. *Strategies for Improving Latency and Throughput of the Apache Hadoop Ecosystem for Medical Imaging Data*. 2016. http://www.dre.vanderbilt.edu/~gokhale/WWW/papers/Middleware16_HBaseOpt.pdf
- Hadoop – image feature extraction from bulk data – Schaer R. Using MapReduce for Large-scale Medical Image Analysis. 2012. https://www.slideshare.net/IIG_HES/20120927-hisb-usingmapreduce
- Hadoop – PACS basis – Ganapathy et al. *Circumventing Picture Archiving and Communication Systems Server with Hadoop Framework in Health Care Services*. 2010. http://thescipub.com/abstract/10.3844/jssp.2010.310.314
- RDF – SPARQL – Jena – Tello et al. *RDF-ization of DICOM Medical Images towards Linked Health Data Cloud*. 2014. https://link.springer.com/chapter/10.1007/978-3-319-13117-7_193
- Gfarm – Hiroyasu et al. *Distributed PACS using distributed file system with hierarchical meta data servers*. 2012. http://www.is.doshisha.ac.jp/academic/papers/pdf/12/201209_minamitaniembc.pdf
- MIRTH – PostgreSQL – Langer S. *A Flexible Database Architecture for Mining DICOM Objects: the DICOM Data Warehouse*. 2012. http://www.springerlink.com/content/77448527x3k40221/fulltext.html

RT Data Mining Examples

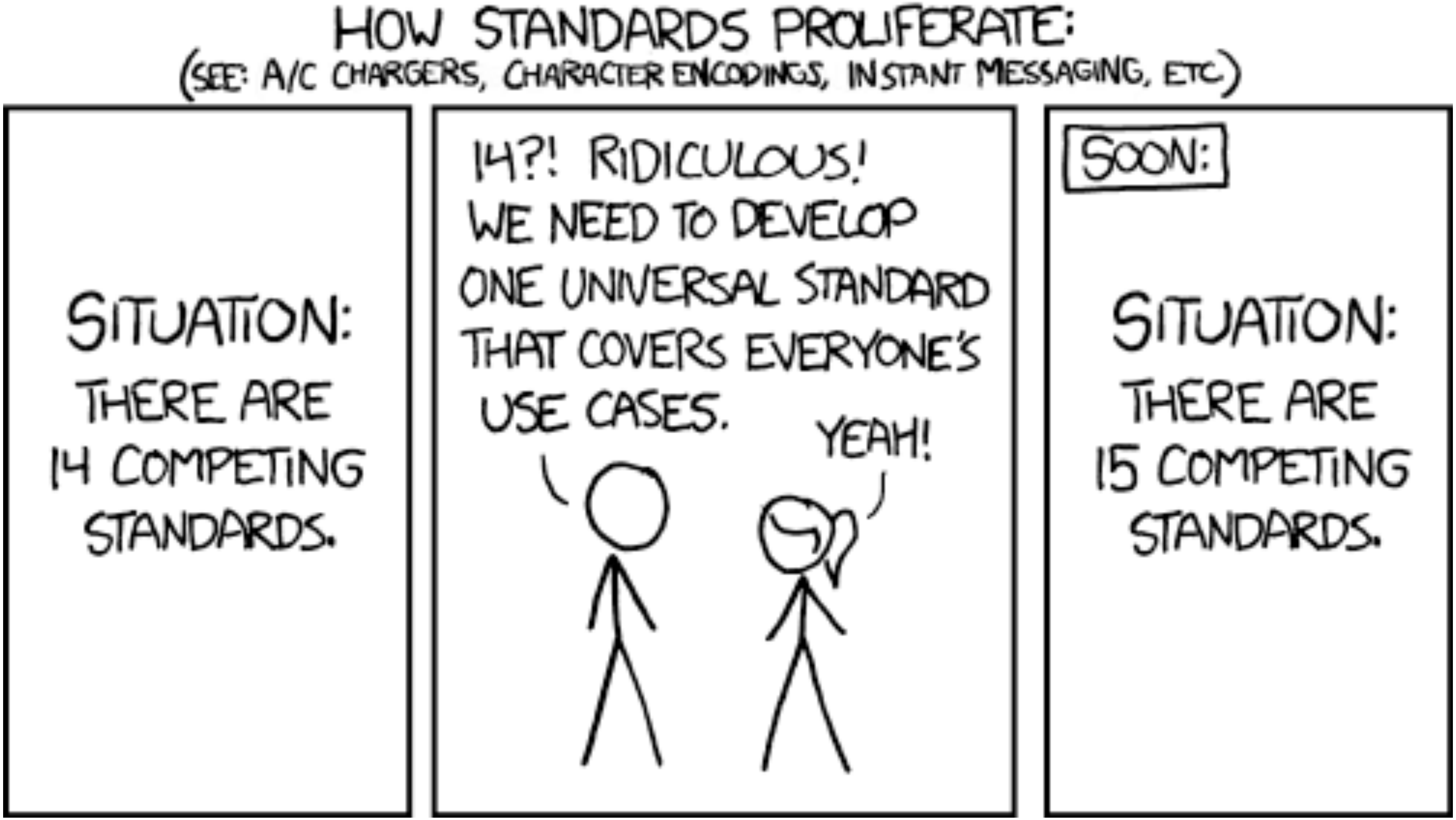
- Roelofs et al. *International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining*. 2014. doi:10.1016/j.radonc. 2013.11.001
- DICOM for RT, SNOMED for clinical data
- Italian language translation

Beyond Imaging – Integrative Queries

- Diagnostic radiology (imaging) – routine or "radiomic" (e.g., feature extraction)
- Anatomical pathology – reports, images (WSI), automated analysis results
- Genomic and proteomic
- Clinical data – demographics, disease, anatomy, pathology (biopsy), staging (incl. TNM), outcome (death, recurrence, survival), treatment (medical, surgical, radiation)
- Radiation therapy

Other Initiatives

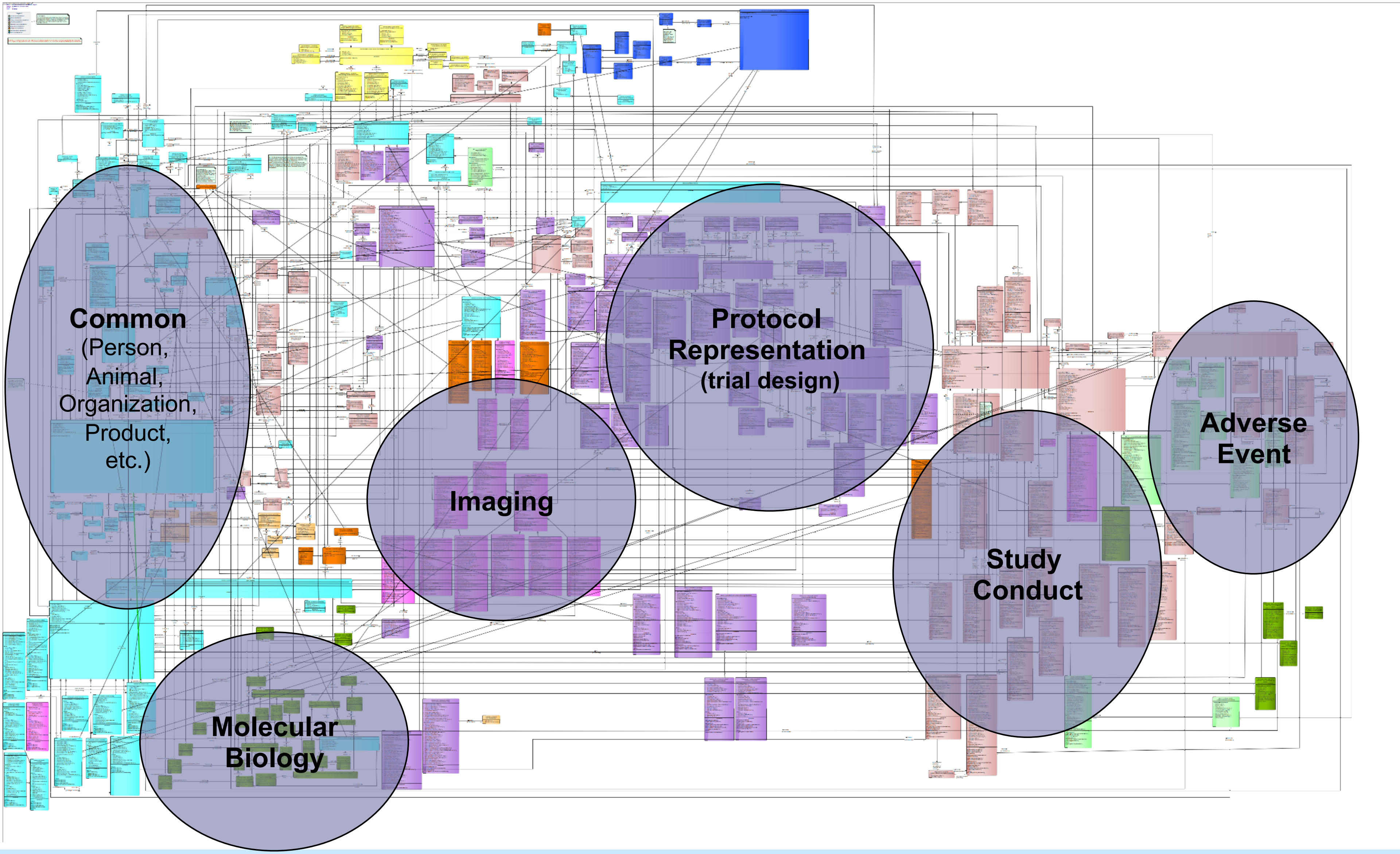
- Some of which may have mapping to DICOM
- BRIDG
- CDISC SDTM esp. Oncology Domain
- Genomic Data Commons (GDC) – cross-study and study-specific
- HL7 V2
- HL7 V3 RIM
- HL7 Clinical Document Architecture (CDA)
- HL7 FHIR
- Registries - SEER


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

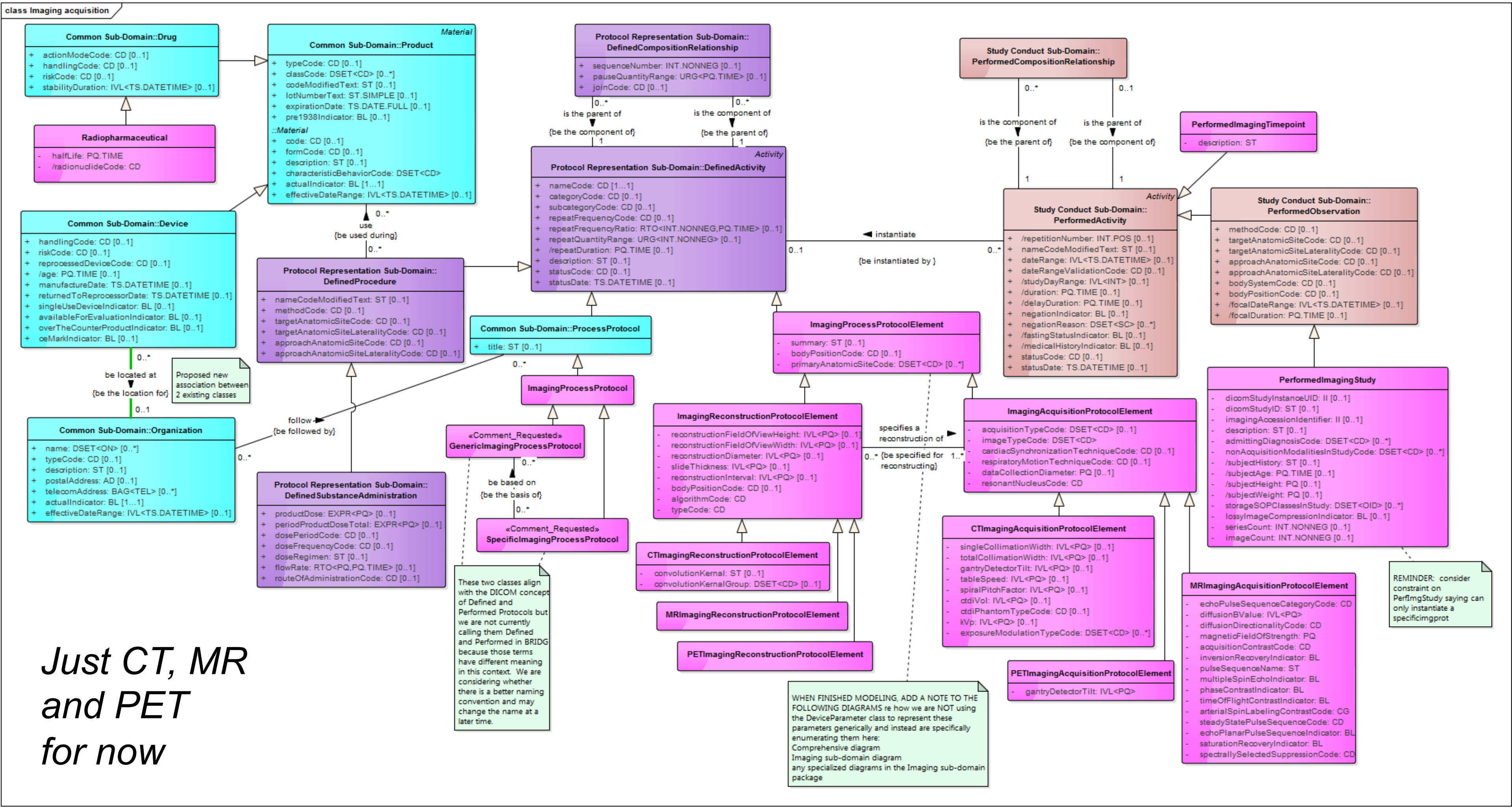
BRIDG Model Overview

- BRIDG – Biomedical Research Integrated Domain Group Model
- Protocol-driven research and translational sciences research
- Collaborative standard developed by CDISC, FDA, HL7, ISO and NCI
  - ISO 14199 Standard 2015
- BRIDG is a Domain Information Model for Translational research
  - a UML model and class diagram (in Enterprise Architect)
  - combined semantics from CDISC, HL7 and ISO to enable semantic interoperability
- Scope changed in 2014 to include translational sciences
  - includes in vivo imaging, pathology and clinical genomics
- BRIDG contains CDISC data standards harmonized over last 8 years
  - CDISC SDTM required for FDA Division of Oncology submissions

BRIDG Model


Common
(Person,
Animal,
Organization,
Product,
etc.)
Protocol
Representation
(trial design)
Imaging
Adverse
Event
Study
Conduct
Molecular
Biology

DICOM Imaging added to BRIDG



class Imaging acquisition

Common Sub-Domain::Drug
- + actionModeCode: CD [0..1]
- + handlingCode: CD [0..1]
- + riskCode: CD [0..1]
- + stabilityDuration: IVL<TS.DATETIME> [0..1]

Radiopharmaceutical
- - halfLife: PQ.TIME
- - /radionuclideCode: CD

Material
Common Sub-Domain::Product
- + typeCode: CD [0..1]
- + classCode: DSET<CD> [0..*]
- + codeModifiedText: ST [0..1]
- + lotNumberText: ST.SIMPLE [0..1]
- + expirationDate: TS.DATE.FULL [0..1]
- + pre1938Indicator: BL [0..1]

::Material
- + code: CD [0..1]
- + formCode: CD [0..1]
- + description: ST [0..1]
- + characteristicBehaviorCode: DSET<CD>
- + actualIndicator: BL [1...1]
- + effectiveDateRange: IVL<TS.DATETIME> [0..1]

Common Sub-Domain::Device
- + handlingCode: CD [0..1]
- + riskCode: CD [0..1]
- + reprocessedDeviceCode: CD [0..1]
- + /age: PQ.TIME [0..1]
- + manufactureDate: TS.DATETIME [0..1]
- + returnedToReprocessorDate: TS.DATETIME [0..1]
- + singleUseDeviceIndicator: BL [0..1]
- + availableForEvaluationIndicator: BL [0..1]
- + overTheCounterProductIndicator: BL [0..1]
- + ceMarkIndicator: BL [0..1]

Common Sub-Domain::Organization
- + name: DSET<ON> [0..*]
- + typeCode: CD [0..1]
- + description: ST [0..1]
- + postalAddress: AD [0..1]
- + telecomAddress: BAG<TEL> [0..*]
- + actualIndicator: BL [1...1]
- + effectiveDateRange: IVL<TS.DATETIME> [0..1]

Proposed new association between 2 existing classes

Protocol Representation Sub-Domain:: DefinedProcedure
- + nameCodeModifiedText: ST [0..1]
- + methodCode: CD [0..1]
- + targetAnatomicSiteCode: CD [0..1]
- + targetAnatomicSiteLateralityCode: CD [0..1]
- + approachAnatomicSiteCode: CD [0..1]
- + approachAnatomicSiteLateralityCode: CD [0..1]

Protocol Representation Sub-Domain:: DefinedSubstanceAdministration
- + productDose: EXPR<PQ> [0..1]
- + periodProductDoseTotal: EXPR<PQ> [0..1]
- + dosePeriodCode: CD [0..1]
- + doseFrequencyCode: CD [0..1]
- + doseRegimen: ST [0..1]
- + flowRate: RTO<PQ,PQ.TIME> [0..1]
- + routeOfAdministrationCode: CD [0..1]

Protocol Representation Sub-Domain:: DefinedCompositionRelationship
- + sequenceNumber: INT.NONNEG [0..1]
- + pauseQuantityRange: URG<PQ.TIME> [0..1]
- + joinCode: CD [0..1]

Activity
Protocol Representation Sub-Domain::DefinedActivity
- + nameCode: CD [1...1]
- + categoryCode: CD [0..1]
- + subcategoryCode: CD [0..1]
- + repeatFrequencyCode: CD [0..1]
- + repeatFrequencyRatio: RTO<INT.NONNEG,PQ.TIME> [0..1]
- + repeatQuantityRange: URG<INT.NONNEG> [0..1]
- + /repeatDuration: PQ.TIME [0..1]
- + description: ST [0..1]
- + statusCode: CD [0..1]
- + statusDate: TS.DATETIME [0..1]

Common Sub-Domain::ProcessProtocol
- + title: ST [0..1]

ImagingProcessProtocol

«Comment_Requested»
GenericImagingProcessProtocol

«Comment_Requested»
SpecificImagingProcessProtocol

These two classes align with the DICOM concept of Defined and Performed Protocols but we are not currently calling them Defined and Performed in BRIDG because those terms have different meaning in this context. We are considering whether there is a better naming convention and may change the name at a later time.

ImagingProcessProtocolElement
- - summary: ST [0..1]
- - bodyPositionCode: CD [0..1]
- - primaryAnatomicSiteCode: DSET<CD> [0..*]

ImagingReconstructionProtocolElement
- - reconstructionFieldOfViewHeight: IVL<PQ> [0..1]
- - reconstructionFieldOfViewWidth: IVL<PQ> [0..1]
- - reconstructionDiameter: IVL<PQ> [0..1]
- - slideThickness: IVL<PQ> [0..1]
- - reconstructionInterval: IVL<PQ> [0..1]
- - bodyPositionCode: CD [0..1]
- - algorithmCode: CD
- - typeCode: CD

CTImagingReconstructionProtocolElement
- - convolutionKernal: ST [0..1]
- - convolutionKernalGroup: DSET<CD> [0..1]

MRImagingReconstructionProtocolElement

PETImagingReconstructionProtocolElement

WHEN FINISHED MODELING, ADD A NOTE TO THE FOLLOWING DIAGRAMS re how we are NOT using the DeviceParameter class to represent these parameters generically and instead are specifically enumerating them here:
Comprehensive diagram
Imaging sub-domain diagram
any specialized diagrams in the Imaging sub-domain package

Study Conduct Sub-Domain:: PerformedCompositionRelationship

PerformedImagingTimepoint
- - description: ST

Activity
Study Conduct Sub-Domain:: PerformedActivity
- + /repetitionNumber: INT.POS [0..1]
- + nameCodeModifiedText: ST [0..1]
- + dateRange: IVL<TS.DATETIME> [0..1]
- + dateRangeValidationCode: CD [0..1]
- + /studyDayRange: IVL<INT> [0..1]
- + /duration: PQ.TIME [0..1]
- + /delayDuration: PQ.TIME [0..1]
- + negationIndicator: BL [0..1]
- + negationReason: DSET<SC> [0..*]
- + /fastingStatusIndicator: BL [0..1]
- + /medicalHistoryIndicator: BL [0..1]
- + statusCode: CD [0..1]
- + statusDate: TS.DATETIME [0..1]

Study Conduct Sub-Domain:: PerformedObservation
- + methodCode: CD [0..1]
- + targetAnatomicSiteCode: CD [0..1]
- + targetAnatomicSiteLateralityCode: CD [0..1]
- + approachAnatomicSiteCode: CD [0..1]
- + approachAnatomicSiteLateralityCode: CD [0..1]
- + bodySystemCode: CD [0..1]
- + bodyPositionCode: CD [0..1]
- + /focalDateRange: IVL<TS.DATETIME> [0..1]
- + /focalDuration: PQ.TIME [0..1]

PerformedImagingStudy
- - dicomStudyInstanceUID: II [0..1]
- - dicomStudyID: ST [0..1]
- - imagingAccessionIdentifier: II [0..1]
- - description: ST [0..1]
- - admittingDiagnosisCode: DSET<CD> [0..*]
- - nonAcquisitionModalitiesInStudyCode: DSET<CD> [0..*]
- - /subjectHistory: ST [0..1]
- - /subjectAge: PQ.TIME [0..1]
- - /subjectHeight: PQ [0..1]
- - /subjectWeight: PQ [0..1]
- - storageSOPClassesInStudy: DSET<OID> [0..*]
- - lossyImageCompressionIndicator: BL [0..1]
- - seriesCount: INT.NONNEG [0..1]
- - imageCount: INT.NONNEG [0..1]

REMINDER: consider constraint on PerfImgStudy saying can only instantiate a specificimgprot

ImagingAcquisitionProtocolElement
- - acquisitionTypeCode: DSET<CD> [0..1]
- - imageTypeCode: DSET<CD>
- - cardiacSynchronizationTechniqueCode: CD [0..1]
- - respiratoryMotionTechniqueCode: CD [0..1]
- - dataCollectionDiameter: PQ [0..1]
- - resonantNucleusCode: CD

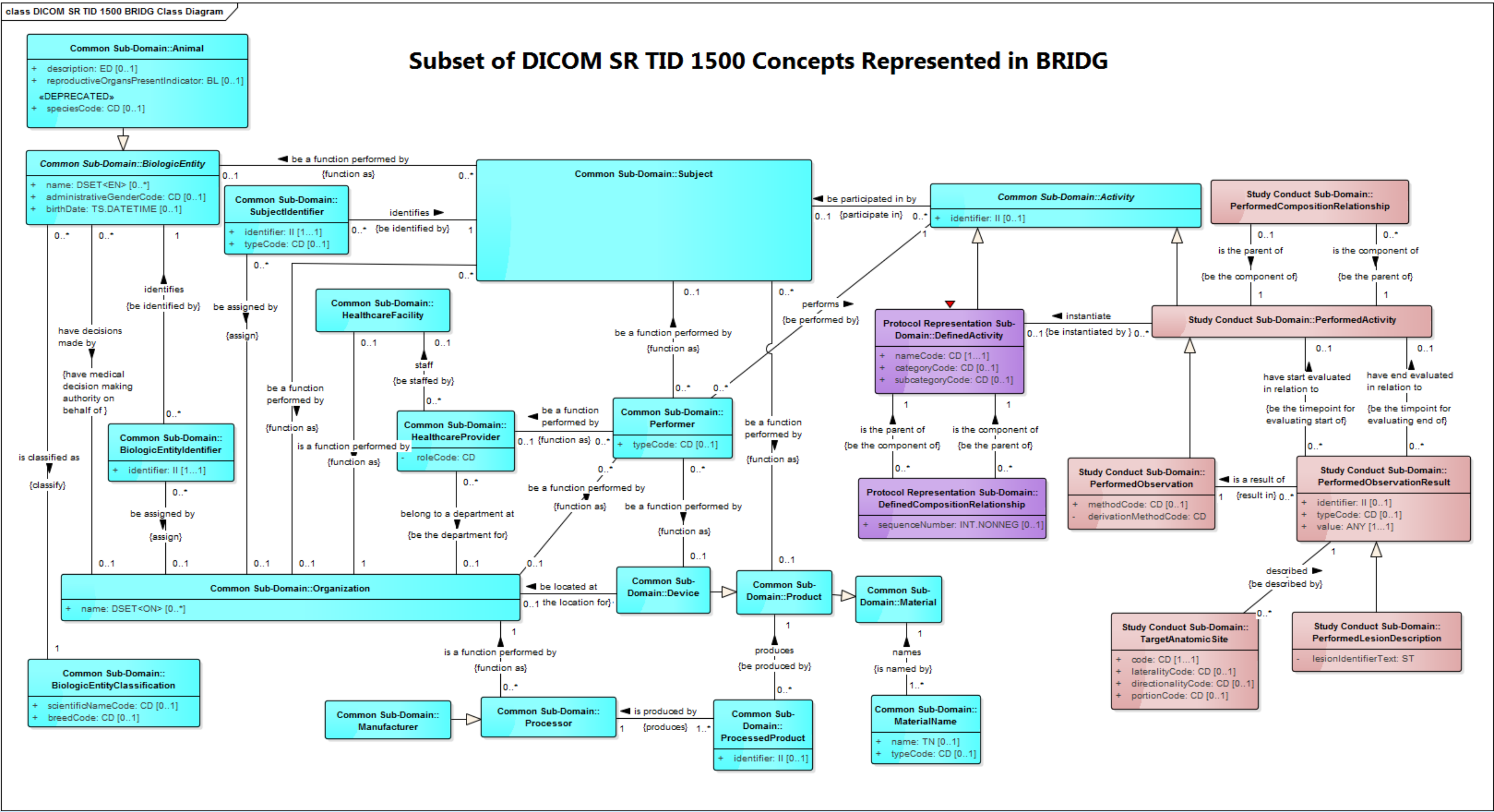
CTImagingAcquisitionProtocolElement
- - singleCollimationWidth: IVL<PQ> [0..1]
- - totalCollimationWidth: IVL<PQ> [0..1]
- - gantryDetectorTilt: IVL<PQ> [0..1]
- - tableSpeed: IVL<PQ> [0..1]
- - spiralPitchFactor: IVL<PQ> [0..1]
- - ctdiVol: IVL<PQ> [0..1]
- - ctdiPhantomTypeCode: CD [0..1]
- - kVp: IVL<PQ> [0..1]
- - exposureModulationTypeCode: DSET<CD> [0..*]

PETImagingAcquisitionProtocolElement
- - gantryDetectorTilt: IVL<PQ>

MRImagingAcquisitionProtocolElement
- - echoPulseSequenceCategoryCode: CD
- - diffusionBValue: IVL<PQ>
- - diffusionDirectionalityCode: CD
- - magneticFieldOfStrength: PQ
- - acquisitionContrastCode: CD
- - inversionRecoveryIndicator: BL
- - pulseSequenceName: ST
- - multipleSpinEchoIndicator: BL
- - phaseContrastIndicator: BL
- - timeOfFlightContrastIndicator: BL
- - arterialSpinLabelingContrastCode: CG
- - steadyStatePulseSequenceCode: CD
- - echoPlanarPulseSequenceIndicator: BL
- - saturationRecoveryIndicator: BL
- - spectrallySelectedSuppressionCode: CD

Just CT, MR and PET for now

BRIDG – DICOM SR TID 1500



class DICOM SR TID 1500 BRIDG Class Diagram

Subset of DICOM SR TID 1500 Concepts Represented in BRIDG

Common Sub-Domain::Animal
- + description: ED [0..1]
- + reproductiveOrgansPresentIndicator: BL [0..1]
- «DEPRECATED»
- + speciesCode: CD [0..1]

Common Sub-Domain::BiologicEntity
- + name: DSET<EN> [0..*]
- + administrativeGenderCode: CD [0..1]
- + birthDate: TS.DATETIME [0..1]

Common Sub-Domain:: SubjectIdentifier
- + identifier: II [1...1]
- + typeCode: CD [0..1]

Common Sub-Domain::Subject

Common Sub-Domain::Activity
- + identifier: II [0..1]

Study Conduct Sub-Domain:: PerformedCompositionRelationship

Common Sub-Domain:: HealthcareFacility

Common Sub-Domain:: BiologicEntityIdentifier
- + identifier: II [1...1]

Common Sub-Domain:: HealthcareProvider
- - roleCode: CD

Common Sub-Domain:: Performer
- + typeCode: CD [0..1]

Protocol Representation Sub-Domain::DefinedActivity
- + nameCode: CD [1...1]
- + categoryCode: CD [0..1]
- + subcategoryCode: CD [0..1]

Study Conduct Sub-Domain::PerformedActivity

Protocol Representation Sub-Domain:: DefinedCompositionRelationship
- + sequenceNumber: INT.NONNEG [0..1]

Study Conduct Sub-Domain:: PerformedObservation
- + methodCode: CD [0..1]
- - derivationMethodCode: CD

Study Conduct Sub-Domain:: PerformedObservationResult
- + identifier: II [0..1]
- + typeCode: CD [0..1]
- + value: ANY [1...1]

Common Sub-Domain::Organization
- + name: DSET<ON> [0..*]

Common Sub-Domain::Device

Common Sub-Domain::Product

Common Sub-Domain::Material

Study Conduct Sub-Domain:: TargetAnatomicSite
- + code: CD [1...1]
- + lateralityCode: CD [0..1]
- + directionalityCode: CD [0..1]
- + portionCode: CD [0..1]

Study Conduct Sub-Domain:: PerformedLesionDescription
- - lesionIdentifierText: ST

Common Sub-Domain:: BiologicEntityClassification
- + scientificNameCode: CD [0..1]
- + breedCode: CD [0..1]

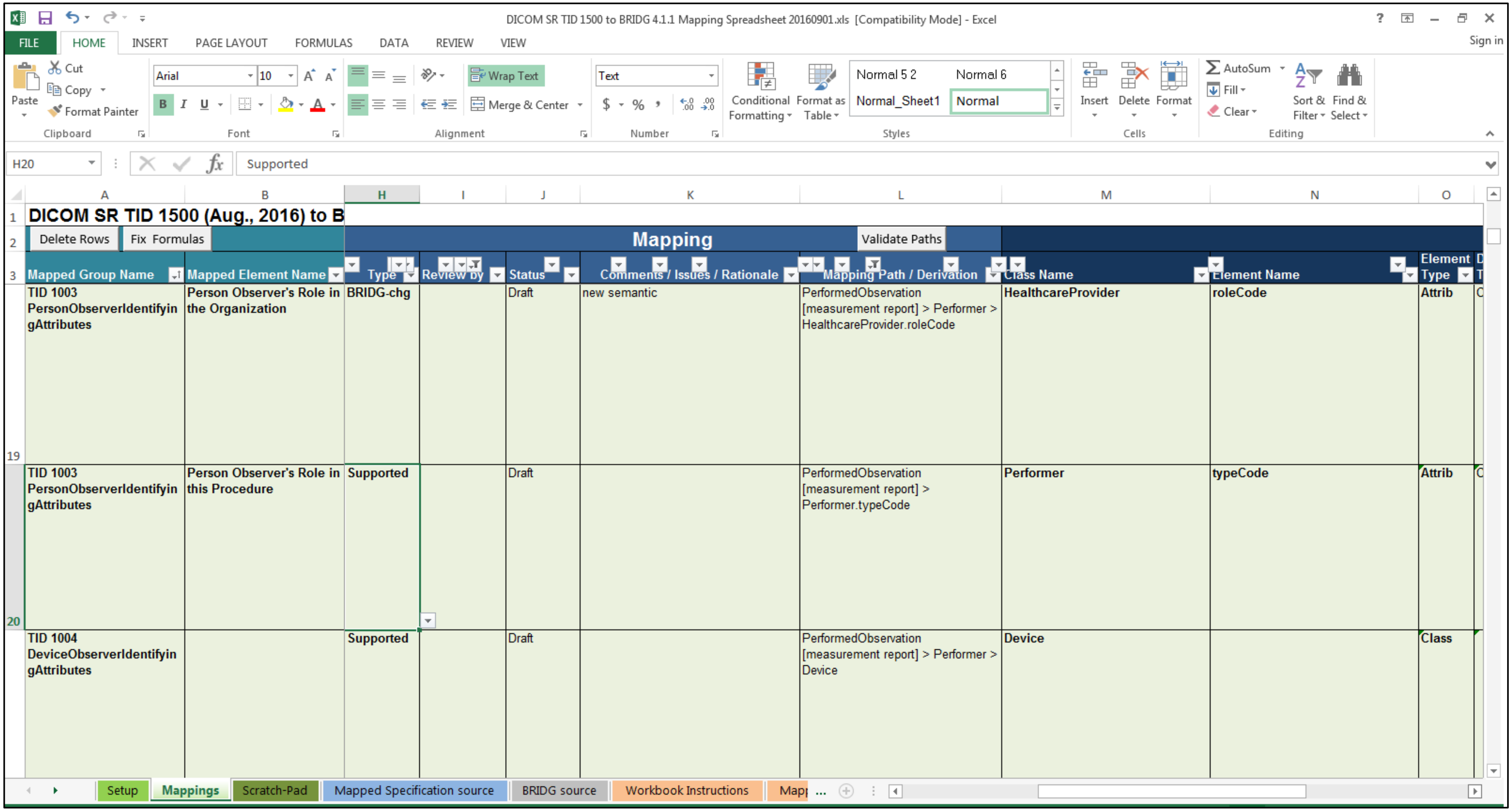
Common Sub-Domain:: Manufacturer

Common Sub-Domain:: Processor

Common Sub-Domain:: ProcessedProduct
- + identifier: II [0..1]

Common Sub-Domain:: MaterialName
- + name: TN [0..1]
- + typeCode: CD [0..1]

BRIDG – DICOM SR TID 1500



DICOM SR TID 1500 to BRIDG 4.1.1 Mapping Spreadsheet 20160901.xls [Compatibility Mode] - Excel

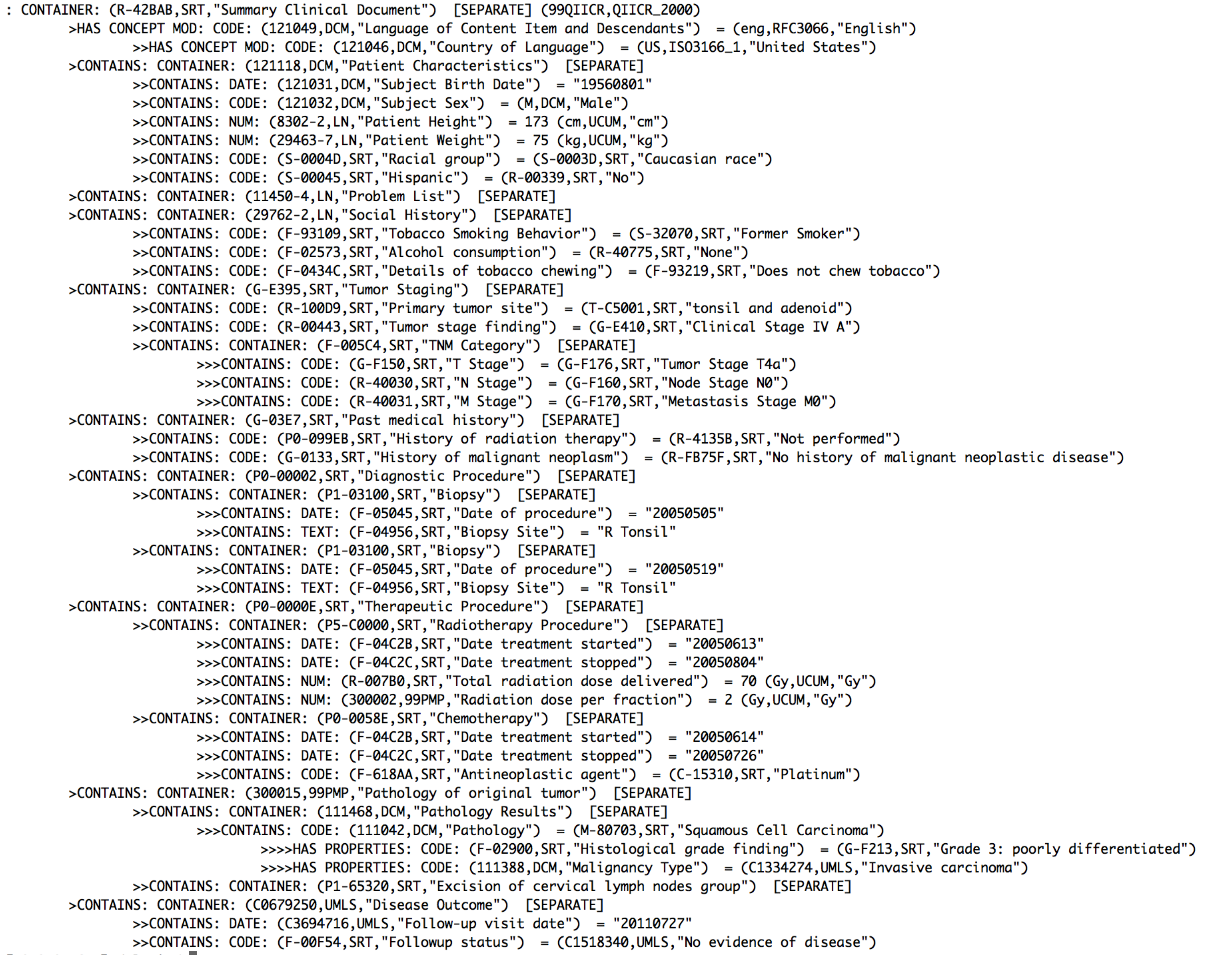
H20 Supported

DICOM SR TID 1500 (Aug., 2016) to B

	A	B	H	I	J	K	L	M	N	O
3	Mapped Group Name	Mapped Element Name	Type	Review by	Status	Comments / Issues / Rationale	Mapping Path / Derivation	Class Name	Element Name	Element Type
19	TID 1003 PersonObserverIdentifyingAttributes	Person Observer's Role in the Organization	BRIDG-chg		Draft	new semantic	PerformedObservation [measurement report] > Performer > HealthcareProvider.roleCode	HealthcareProvider	roleCode	Attrib
20	TID 1003 PersonObserverIdentifyingAttributes	Person Observer's Role in this Procedure	Supported		Draft		PerformedObservation [measurement report] > Performer.typeCode	Performer	typeCode	Attrib
	TID 1004 DeviceObserverIdentifyingAttributes		Supported		Draft		PerformedObservation [measurement report] > Performer > Device	Device		Class

Clinical Data

- Disease
- Anatomical pathology and staging
- Treatment
- Outcome – recurrence, survival
- Not ideal, but can use DICOM SR
- Leverage “relevant clinical information” templates (intended pre-imaging)
- QIICR – 10.7717/peerj.2057
- NCI CIP DI-cubed project



```
: CONTAINER: (R-42BAB,SRT,"Summary Clinical Document")  [SEPARATE] (99QIICR,QIICR_2000)
        >HAS CONCEPT MOD: CODE: (121049,DCM,"Language of Content Item and Descendants")  = (eng,RFC3066,"English")
                >>HAS CONCEPT MOD: CODE: (121046,DCM,"Country of Language")  = (US,ISO3166_1,"United States")
        >CONTAINS: CONTAINER: (121118,DCM,"Patient Characteristics")  [SEPARATE]
                >>CONTAINS: DATE: (121031,DCM,"Subject Birth Date")  = "19560801"
                >>CONTAINS: CODE: (121032,DCM,"Subject Sex")  = (M,DCM,"Male")
                >>CONTAINS: NUM: (8302-2,LN,"Patient Height")  = 173 (cm,UCUM,"cm")
                >>CONTAINS: NUM: (29463-7,LN,"Patient Weight")  = 75 (kg,UCUM,"kg")
                >>CONTAINS: CODE: (S-0004D,SRT,"Racial group")  = (S-0003D,SRT,"Caucasian race")
                >>CONTAINS: CODE: (S-00045,SRT,"Hispanic")  = (R-00339,SRT,"No")
        >CONTAINS: CONTAINER: (11450-4,LN,"Problem List")  [SEPARATE]
        >CONTAINS: CONTAINER: (29762-2,LN,"Social History")  [SEPARATE]
                >>CONTAINS: CODE: (F-93109,SRT,"Tobacco Smoking Behavior")  = (S-32070,SRT,"Former Smoker")
                >>CONTAINS: CODE: (F-02573,SRT,"Alcohol consumption")  = (R-40775,SRT,"None")
                >>CONTAINS: CODE: (F-0434C,SRT,"Details of tobacco chewing")  = (F-93219,SRT,"Does not chew tobacco")
        >CONTAINS: CONTAINER: (G-E395,SRT,"Tumor Staging")  [SEPARATE]
                >>CONTAINS: CODE: (R-100D9,SRT,"Primary tumor site")  = (T-C5001,SRT,"tonsil and adenoid")
                >>CONTAINS: CODE: (R-00443,SRT,"Tumor stage finding")  = (G-E410,SRT,"Clinical Stage IV A")
                >>CONTAINS: CONTAINER: (F-005C4,SRT,"TNM Category")  [SEPARATE]
                        >>>CONTAINS: CODE: (G-F150,SRT,"T Stage")  = (G-F176,SRT,"Tumor Stage T4a")
                        >>>CONTAINS: CODE: (R-40030,SRT,"N Stage")  = (G-F160,SRT,"Node Stage N0")
                        >>>CONTAINS: CODE: (R-40031,SRT,"M Stage")  = (G-F170,SRT,"Metastasis Stage M0")
        >CONTAINS: CONTAINER: (G-03E7,SRT,"Past medical history")  [SEPARATE]
                >>CONTAINS: CODE: (P0-099EB,SRT,"History of radiation therapy")  = (R-4135B,SRT,"Not performed")
                >>CONTAINS: CODE: (G-0133,SRT,"History of malignant neoplasm")  = (R-FB75F,SRT,"No history of malignant neoplastic disease")
        >CONTAINS: CONTAINER: (P0-00002,SRT,"Diagnostic Procedure")  [SEPARATE]
                >>CONTAINS: CONTAINER: (P1-03100,SRT,"Biopsy")  [SEPARATE]
                        >>>CONTAINS: DATE: (F-05045,SRT,"Date of procedure")  = "20050505"
                        >>>CONTAINS: TEXT: (F-04956,SRT,"Biopsy Site")  = "R Tonsil"
                >>CONTAINS: CONTAINER: (P1-03100,SRT,"Biopsy")  [SEPARATE]
                        >>>CONTAINS: DATE: (F-05045,SRT,"Date of procedure")  = "20050519"
                        >>>CONTAINS: TEXT: (F-04956,SRT,"Biopsy Site")  = "R Tonsil"
        >CONTAINS: CONTAINER: (P0-0000E,SRT,"Therapeutic Procedure")  [SEPARATE]
                >>CONTAINS: CONTAINER: (P5-C0000,SRT,"Radiotherapy Procedure")  [SEPARATE]
                        >>>CONTAINS: DATE: (F-04C2B,SRT,"Date treatment started")  = "20050613"
                        >>>CONTAINS: DATE: (F-04C2C,SRT,"Date treatment stopped")  = "20050804"
                        >>>CONTAINS: NUM: (R-007B0,SRT,"Total radiation dose delivered")  = 70 (Gy,UCUM,"Gy")
                        >>>CONTAINS: NUM: (300002,99PMP,"Radiation dose per fraction")  = 2 (Gy,UCUM,"Gy")
                >>CONTAINS: CONTAINER: (P0-0058E,SRT,"Chemotherapy")  [SEPARATE]
                        >>>CONTAINS: DATE: (F-04C2B,SRT,"Date treatment started")  = "20050614"
                        >>>CONTAINS: DATE: (F-04C2C,SRT,"Date treatment stopped")  = "20050726"
                        >>>CONTAINS: CODE: (F-618AA,SRT,"Antineoplastic agent")  = (C-15310,SRT,"Platinum")
        >CONTAINS: CONTAINER: (300015,99PMP,"Pathology of original tumor")  [SEPARATE]
                >>CONTAINS: CONTAINER: (111468,DCM,"Pathology Results")  [SEPARATE]
                        >>>CONTAINS: CODE: (111042,DCM,"Pathology")  = (M-80703,SRT,"Squamous Cell Carcinoma")
                                >>>>HAS PROPERTIES: CODE: (F-02900,SRT,"Histological grade finding")  = (G-F213,SRT,"Grade 3: poorly differentiated")
                                >>>>HAS PROPERTIES: CODE: (111388,DCM,"Malignancy Type")  = (C1334274,UMLS,"Invasive carcinoma")
                >>CONTAINS: CONTAINER: (P1-65320,SRT,"Excision of cervical lymph nodes group")  [SEPARATE]
        >CONTAINS: CONTAINER: (C0679250,UMLS,"Disease Outcome")  [SEPARATE]
                >>CONTAINS: DATE: (C3694716,UMLS,"Follow-up visit date")  = "20110727"
                >>CONTAINS: CODE: (F-00F54,SRT,"Followup status")  = (C1518340,UMLS,"No evidence of disease")
```

Summary

- Extracting information from DICOM images and non-image objects
- Many tools to extract the DICOM context (for both individual elements and structured content, e.g., SR) to feed the ETL process
- Which attributes – many to choose from – sparseness
- Consistency of attribute values is challenging (esp. free text values)
- Use of standard codes (DCM, SNOMED)
- Specific RT attribute/value standardization efforts
- “Context” of each use (place in a tree flattened to a row)
- Role of standard mapping from DICOM to broader based models (e.g., BRIDG)