DICOM for Digital Pathology – Current Landscape

David Clunie, PixelMed

Disclosures

Editor of the DICOM Standard (NEMA contract) NCI FNL Leidos Essex sub-contractor (SME DICOM, de-identification) NCI Imaging Data Commons (IDC) sub-contractor Consult with various equipment manufacturers re-DICOM

Overview

Introduction to DICOM for WSI Connectathon overview (past and current) DICOM, Imaging Data Commons (IDC) and Slim Viewer DICOM conformance tools (validators) and their utility to implementers Updates on new activities



Meyer-Ebrecht D. [Electronic Archival System for X-Rays Images - Work proposal for a research project in the years 1974 and 1975] Elektronisches Archivierungssystem für Röntgenbilder – Arbeitsvortenag jur ein Forschungsprojekt in den Jahren 1974 und 1975. Hamburg, Germany: Philips Research Labs; 1973 Oct.

DICOM – Interoperability Boundaries – Commoditization



DICOM – Interoperability for Pathology



DICOM – Analysis Systems – Computational Pathology



DICOM – Enterprise Imaging



FDA "Entire Pixel Pathway" – Non-interoperable Black Box







Interoperability boundaries for pathology imaging

Access to input data images \checkmark DICOM WSMI (tiled pyramids) annotations \checkmark DICOM SEG, SR, ANN protocols ~ DICOMweb query, metadata, frames Sharing of output data images \checkmark DICOM parametric maps (tiled pyramids) annotations </ DICOM SEG, SR, ANN protocols ~ DICOMweb STOW Management of workflow application selection X application orchestration ~ *DICOMweb UPS*





Herrmann 2018



Information Source

Microscope

Laboratory Information System

Herrmann 2018

Electronic Medical Record

VL Whole Slide Microscopy Image							
Patient	General Study	General Series	Specimen		General Image	Image Pixel	
Patient's Name Patient's Sex Patient's Birth Date Patient ID 	Study Instance UID Study Date Study Time Accession Number 	Series Instance UID Series Date Series Time Series Number 	Container Identifier Container Description Specimen Identifier Specimen Description Seq.		Instance Number Acquisition Date Acquisition Time Acquisition Number 	Samples Per Pixel Rows Columns Pixel Data 	





Herrmann 2018



Herrmann 2018

So where are we at?

Most scanner vendors have implemented DICOM WSI Demonstrated at successive Connectations since 2017 Few have FDA clearance (at all, and to operate with DICOM) Few sites are actually using the DICOM capability if present Some very large sites are insisting on DICOM Some converting proprietary formats to DICOM themselves DICOM can have mathematically identical pixels Two popular open-source libraries now support DICOM WSI Testing annotations in recent Hackathons/Connectathons

DICOM WG 26 WSI Connectathons – Participation

	PV'17	Pl'18	ECDP'18	PV'18	ECDP'19	PV'19	PV'20
AidPath	View		Archive, View				
Corista		Analyze, View				View	
GE							Archive
Gestalt				Archive, View			
Infinitt						Archive	
J4Care							Archive, View
MGH							View
Neagen				Archive, View	Archive, View	Archive, View	Archive, View
PathCore	Archive, View	Archive, View		Archive, View		Archive, View	Archive, View
Sectra		View	View	View		View	
3DHistech					Scan	Scan	Scan
Hamamatsu		Scan	Scan		Scan	Scan	Scan
Huron						Scan	Scan
Leica	Scan	Scan		Scan			Scan
Motic				Scan		Scan	Scan
Nference							Scan
Philips	Scan		Scan	Scan	Scan	Scan	Scan
Roche Ventana	Scan	Scan	Scan	Scan	Scan	Scan	Scan, View

DICOM, Imaging Data Commons and Slim Viewer

IDC is the imaging node of the NCI CBIIT Cancer Research Data Commons Goal is to make images available in the Cloud for computing Distinct from other archives that are not Cloud-based Shares content from The Cancer Image Archive (TCIA) and other sources Radiology and Pathology – both in DICOM format ONLY WSI converted to DICOM (has not yet received any native WSI DICOM) IDC consists of DICOM files in Google and AWS buckets, Google DICOM Store and DICOM metadata automatically extracted to BigQuery Application-appropriate DICOMweb viewers – radiology: OHIF, WSI: Slim Images, annotations, other DICOM objects (radiotherapy, etc.)



NATIONAL CANCER INSTITUTE Imaging Data Commons	Explore Images	Collections	Getting Started [2] User Forum [2]				
			Sound of the source of the sou	News About Help 🗹			→] Sign
	U	Col	lection Name	∧ Total # of 0	Cases	# of Cases(this cohort)	
▶ □ HTAN	0		RMS-Mutation-Prediction ()	403		95	
MIDI	0						
▶ □ NLM	0						
▶ □ QIBA	0	Selected	Cases				
Analysis Results	\$ Q	Showing 1	to 10 of 95 entries Show 10 v entries	s Page Go Previou	is 1 2 3 4 5 10 Next	Find by Case ID:	Go
Sort by: Count Alpha		00	Collection Name	Case ID 🚯	A Total # of Studies	Total # of Series	
□ Hide analysis results with 0 cases			RMS-Mutation-Prediction	RMS2150	1	2	
show 11 more Check All / U	Uncheck All		RMS-Mutation-Prediction	RMS2152	1	2	
 RMS-Mutation-Prediction-Expert- Annotations I 	95		RMS-Mutation-Prediction	RMS2153 💕	1	2	
None 🕄	308		RMS-Mutation-Prediction	RMS2154	1	2	
TotalSegmentator-CT-Segmentations BAME-AIMI-Apportations	0		DMC Mutation Dradiation		1	-	
nnU-Net-BPR-annotations	0		RMS-Mutation-Prediction	RM32133 L		2	
			RMS-Mutation-Prediction	RMS2202	1	2	
earch Configuration			RMS-Mutation-Prediction	RMS2203	1	2	
Hide attribute values with 0 cases			RMS-Mutation-Prediction	RMS2204	1	2	
The autoute values with o cases			RMS-Mutation-Prediction	RMS2206	1	2	
ORIGINAL DERIVED RE	ELATED		RMS-Mutation-Prediction	RMS2207	1	2	
Primary Site Location	\$ Q						
Sort by: Count Alpha							
show 40 more Check All / L	Uncheck All	Selected	Studies				
U various	95	Chaude - 4	to 1 of 1 option Chan 10	Remain Co.			<u> </u>
□ Breast		Snowing 1	to For Fentries Snow 10 V entries	Page G0 Previous	INEXC.	Find by Study Instance UID:	Go
Lung	0	00	Case ID 🚯 🔹 ^ Study Instance	ce UID Study Date	Study Description	# of Series Vie	w
Prostate	0		RMS2152 2.25.155948	338876 🌓 10-31-2018	Histopathology	2	
Cancer Type	# Q						



🗁 🔶 🕻 😂 https://viewer.imaging.datacommons.cancer.gov/slim/studies/2.25.155017484756498730492136597238994838876/series/1.3.6.1.4.1.5962.99.1.3434988325.342625608.16870622011... 😫 🛧



NIH NATIONAL CANCER INSTITUTE Imaging Data Commons

Patient

ID: RMS2152 Name: RMS2152 Gender: Male Birthdate:

Study

Accession #: RMS2152 ID: RMS2152 Date: 2018-10-31 Time: 16:31:25

Slides

PAPJAD-0BGUTB



^



🔏 😘 🔍 🔕 🖸 🦓

NIH NATIONAL CANCER INSTITUTE Imaging Data Commons

^

^

Patient

ID: RMS2152 Name: RMS2152 Gender: Male Birthdate:

Study

Accession #: RMS2152 ID: RMS2152 Date: 2018-10-31 Time: 16:31:25

Slides

PAPJAD-0BGUTB





DICOM Conformance Tools (Validators)

The DICOM Standard describes metadata in meticulous detail Structural, identifying and descriptive information for multiple entities Patient, Study, Series, Equipment, Specimen, Preparation, Acquisition, Some is required, some is optional, some specific values are required Easy for implementors to make mistakes, omit critical information Can assist them with mechanical tools Hand-written or generated from the formal representation of the standard Examples: *dciodvfy* and *dcentvfy* from <u>dicom3tools</u> package *dciodvfy*: detects errors within a single file (instance) *dcentvfy*: detects errors across a set of files (consistency between files) Not a panacea: valid files may not "work", invalid files may "work" (e.g., display) Used routinely in Integrating the Healthcare Enterprise (IHE) & WG26 Connectations

IE	Module	Reference	Usage
Patient	Patient	<u>C.7.1.1</u>	M
	Clinical Trial Subject	<u>C.7.1.3</u>	U
Study	General Study	<u>C.7.2.1</u>	M
	Patient Study	<u>C.7.2.2</u>	U
	Clinical Trial Study	<u>C.7.2.3</u>	U
Series	General Series	<u>C.7.3.1</u>	Μ
	Whole Slide Microscopy Series	<u>C.8.12.3</u>	M
	Clinical Trial Series	<u>C.7.3.2</u>	U
Frame of Reference	Frame of Reference	<u>C.7.4.1</u>	M
Equipment	General Equipment	<u>C.7.5.1</u>	M
	Enhanced General Equipment	<u>C.7.5.2</u>	M
Acquisition	General Acquisition	<u>C.7.10.1</u>	M
Multi-Resolution Pyramid	Multi-Resolution Pyramid	<u>C.7.11.1</u>	U - Shall be present only if Image Type Value 3 is VOLUME or THUMBNAIL.
Image	General Image	<u>C.7.6.1</u>	M
	General Reference	<u>C.12.4</u>	U
	Microscope Slide Layer Tile Organization	<u>C.8.12.14</u>	M
	Image Pixel	<u>C.7.6.3</u>	M
	Acquisition Context	<u>C.7.6.14</u>	M
	Multi-frame Functional Groups	<u>C.7.6.16</u>	M
	Multi-frame Dimension	<u>C.7.6.17</u>	M
	Specimen	<u>C.7.6.22</u>	M
	Whole Slide Microscopy Image	<u>C.8.12.4</u>	M
	Optical Path	<u>C.8.12.5</u>	M
	Slide Label	<u>C.8.12.8</u>	C - Required if Image Type (0008,0008) Value 3 is LABEL; may be present otherwise
	SOP Common	<u>C.12.1</u>	Μ
	Common Instance Reference	<u>C.12.2</u>	Μ
	Frame Extraction	<u>C.12.3</u>	C - Required if the SOP Instance was created in response to a Frame-Level retrieve request

Attribute Name	Тад	Туре	Attribute Description			
Total Pixel Matrix Columns	(0048,0006)	1	Total number of columns in pixel matrix; i.e., width of total imaged volume in pixels. See Section C.8.12.14.1.1			
Total Pixel Matrix Rows	(0048,0007)	1	Total number of rows in pixel matrix; i.e., height of total imaged volume in pixels. See Section C.8.12.14.1.1			
Total Pixel Matrix Focal Planes	(0048,0303)	1C	Total number of focal planes (Z locations) in the pixel matrix; i.e., depth of total imaged volume in pixels. See <u>Section C.8.12.14.1.1</u> Required if Dimension Organization Type (0020,9311) is present with a value of TILED_FULL. May be present otherwise. Note Total Pixel Matrix Focal Planes (0048,0303) describes the number of focal planes separately encoded, and is distinct from Number of Focal Planes (0048,0013) which describes in what manner different focal planes were combined into a single encoded plane (focus stacking)			
Total Pixel Matrix Origin Sequence	(0048,0008)	1	Location of pixel 1\1 of the total pixel matrix in the Slide Coordinate System Frame of Reference. Only a single Item shall be included in this Sequence. See <u>Section C.8.12.14.1.2</u> and <u>Section C.8.12.2.1.1</u> for further explanation.			
>X Offset in Slide Coordinate System	(0040,072A)	1	The X offset in millimeters from the Origin of the Slide Coordinate System.			
>Y Offset in Slide Coordinate System	(0040,073A)	1	The Y offset in millimeters from the Origin of the Slide Coordinate System.			
>Z Offset in Slide Coordinate System	(0040,074A)	1C	The Z offset in μm from the image substrate reference plane (i.e., utilized surface of a glass slide). Required if the Z offset is not zero. May be present otherwise. Note The conditional requirement is used because, historically, this Attribute was not present.			
Image Orientation (Slide)	(0048,0102)	1C	The direction cosines of the first row and the first column of the total pixel matrix with respect to the Slide Coordinate System Frame of Reference. See <u>Section C.8.12.14.1.2</u> . Required if Plane Position (Slide) Sequence (0048,021A) is present within a Functional Group Sequence or Dimension Organization Type (0020,9311) is present with a value of TILED_FULL. May be present otherwise. Note This condition will always be satisfied when this Module is included in the Whole Slide Microscopy Image IOD.			

```
% dciodvfv compressed instance 3 6.dcm
(0x0048,0x0105) SQ Optical Path Sequence - Error - Bad Value Length - not a multiple of 2 - VL is 0xc9 should be 0xca
(0x0028,0x2000) OB ICC Profile - Error - Bad Value Length - not a multiple of 2 - VL is 0x3 should be 0x4
(0x1002,0x1001) ? - Warning - Unrecognized tag - assuming explicit value representation OK
Error - Dicom dataset read failed
Error - Illegal root for UID - "870896693490043133834121745960504521625957166000" in (0x0040,0x0554) Specimen UID
Warning - Missing attribute or value that would be needed to build DICOMDIR - Study ID
Warning - Missing attribute or value that would be needed to build DICOMDIR - Instance Number
Error - Value invalid for this VR - (0x0002,0x0003) UI Media Storage SOP Instance UID UI [1] = <1.2.416.0.0010.3.1.4.537.1.17438.870896693490043133834121745960504521625957166000.3.1.1.1.1.1.1.) - Length invalid for this VR = 95, expected <= 64
Error - Value invalid for this VR - (0x0002,0x0003) UI Media Storage SOP Instance UID UI [1] = <1.2.416.0.0010.3.1.4.537.1.17438.870896693490043133834121745960504521625957166000.3.1.1.1.1.1.1.> - Leading zeroes in embedded numeric component(s)
Error - Value invalid for this VR - (0x0008,0x0018) UI SOP Instance UID UI [1] = <1.2.416.0.0010.3.1.4.537.1.17438.870896693490043133834121745960504521625957166000.3.1.1.1.1.1.1.- - Length invalid for this VR = 95, expected <= 64
Error - Value invalid for this VR - (0x0008,0x0018) UI SOP Instance UID UI [1] = <1.2.416.0.0010.3.1.4.537.1.17438.870896693490043133834121745960504521625957166000.3.1.1.1.1.1.1.> - Leading zeroes in embedded numeric component(s)
Warning - Value dubious for this VR - (0x0008.0x0090) PN Referring Physician's Name PN [1] = <SOME-PHYSICIAN> - Retired Person Name form
Warning - Value dubious for this VR - (0x0010,0x0010) PN Patient's Name PN [1] = <DCM 4_7c-12c> - Retired Person Name form
Error - Value invalid for this VR - (0x0020.0x000d) UI Study Instance UID UI [1] = <1.2.416.0.0010.3.1.2.537.1.17433.870896693490043133834121745960504521625957166000> - Length invalid for this VR = 81. expected <= 64
Error - Value invalid for this VR - (0x0020,0x000d) UI Study Instance UID UI [1] = <1.2.416.0.0010.3.1.2.537.1.17433.870896693490043133834121745960504521625957166000> - Leading zeroes in embedded numeric component(s)
Error - Value invalid for this VR - (0x0020,0x000e) UI Series Instance UID UI [1] = <1.2.416.0.0010.3.1.3.537.1.17433.870896693490043133834121745960504521625957166000.3> - Length invalid for this VR = 83, expected <= 64
Error - Value invalid for this VR - (0x0020,0x000e) UI Series Instance UID UI [1] = <1.2.416.0.0010.3.1.3.537.1.17433.870896693490043133834121745960504521625957166000.3> - Leading zeroes in embedded numeric component(s)
Error - Value invalid for this VR - (0x0028,0x2002) CS Color Space CS [1] = <sRGB> - Character invalid for this VR = 's' (0x73)
Error - Dicom dataset contains invalid data values for Value Representations
VLWholeSlideMicroscopvImage
Error - Missing attribute Type 2 Required Element=<PatientBirthDate> Module=<Patient>
Error - Missing attribute Type 2 Required Element=<StudyID> Module=<GeneralStudy>
Error - Empty attribute (no value) Type 1 Required Element=<Manufacturer> Module=<EnhancedGeneralEquipment>
Error - Missing attribute Type 1 Required Element=<ManufacturerModelName> Module=<EnhancedGeneralEquipment>
Error - Missing attribute Type 1 Required Element=<DeviceSerialNumber> Module=<EnhancedGeneralEquipment>
Error - Missing attribute Type 1 Required Element=<SoftwareVersions> Module=<EnhancedGeneralEquipment>
Error - Missing attribute Type 2 Required Element=<InstanceNumber> Module=<GeneralImage>
Error - Attribute present when condition unsatisfied (which may not be present otherwise) Type 1C Conditional Element=<ImageOrientationSlide> Module=<MicroscopeSlideLaverTileOrganization>
Error - Missing attribute Type 1 Required Element=<InstanceNumber> Module=<MultiFrameFunctionalGroupsCommon>
Error - Missing attribute Type 1 Required Element=<PlanePositionSlideSequence> Module=<PlanePositionSlideMacro>
Error - Missing attribute Type 1 Required Element=<wholeSlideMicroscopyImageFrameTypeSequence> Module=<wholeSlideMicroscopyImageFrameTypeMacro>
Error - Required unless DimensionOrganizationType is TILED FULL - attribute <PerFrameFunctionalGroupsSequence>
Warning - CodingSchemeDesignator is deprecated - attribute <CodingSchemeDesignator> = <SRT>
Warning - Attribute is not present in standard DICOM IOD - (0x0008,0x9007) CS Frame Type
Warning - Attribute is not present in standard DICOM IOD - (0x0040,0x0710) SQ Whole Slide Microscopy Image Frame Type Sequence
Error - Attribute with an even group number is not a recognized standard attribute - (0x1002.0x1001) ?
Warning - Attribute is not present in standard DICOM IOD - (0x1002,0x1001) ?
Warning - Dicom dataset contains attributes not present in standard DICOM IOD - this is a Standard Extended SOP Class
```

% dcentvfy *.dcm

Error - String attribute has different value - Element=</anufacturer> IE=<Equipment> for file <f0ec15e1-56e3-45c4-ad53-91be8df1d90a.dcm> versus <fcc7af34-713d-41a6-869e-0ed1c215090d.dcm> Value 1 <Hamamatsu> versus <Leica Biosystems>

Error - String attribute has different value - Element=«HanufacturerModeUName> IE=<Equipment> for file <f0ec15e1-56e3-45c4-ad53-91be8df1d90a.dcm> versus <fcc7af34-713d-41a6-869e-0ed1c215909d.dcm> Value 1 https://doi.org/10.1111/jac234031-30150-351122351412:-7163

Error - String attribute has different value - Element=Hanufacturer> IE=<Equipment> for file <db50117c-dcdc-4987-8f54-dc1e83b991de.dcm> versus <fcc7af34-713d-41a6-869e-eedic215090d.dcm> Value 1 <Hamamatsu> versus <<.ica Biosystems>

Error - String attribute has different value - Element-ManufacturerModelNames IE-Equipment> for file doSbil7c-dcdc-4987-4954-dcl8330931dc.dcm versus <fc7af34-713d-41a6-869e-dedL235098dc.dcm Value 174an2700z200000- ros.pixelend for file doSbil7c-dcdc-4987-4954-dcl8330931dc.dcm versus <fc7af34-713d-41a6-869e-dedL235098dc.dcm Value 174an2700z200000- ros.pixelend for file doSbil7c-dcdc-4987-4954-dcl8330931dc.dcm versus <fc7af34-713d-41a6-869e-dedL2350980dc.dcm Versus <fc7af34-713d-41a6-869e-dedL2350980dc.

Updates on new pathology-related activities in DICOM

Annotations

Color management – already present but needs more testing New compression schemes – JPEG-XL, Deflate for bitmaps More web-based API – e.g., spatial queries (esp. annotations)



Wen et al. A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image. JPI. 2017.

DICOM WSI Annotations

Use cases: human generated (few) vs. machine generated (millions) Standard supports vector or rasterized (bit mask, soon label map) Supports overlapping and non-overlapping segmentations Coded classification for semantic interoperability (SNOMED CT) Anatomy & property (category/type) (e.g., brain, abnormal structure/necrosis) Hackathons, Connectathons for testing are underway (ECDP 2024) Some use cases, representations & optimizations yet to be explored Considering performance issues with different levels of zoom Spatial query API is probably needed (e.g., annotations that intersect a tile) [000-00-0001] Koehler^August^W^Mr^Dr

M 20000101

20171001 M 20000101 []

000-00-0001 [112-0001] [000-00-0001] Koehler/August/W/Mr/Dr

000-00-0001 [112-0001] 20171001 Series #382183762

1280x1280x1232 +8 YBR FULL 422 [PEG 1:1

023117.000000 023117.000000 074813.000000 [] With ICC Profile Applied VOLUME

023117.000000 023117.000000 074813.000000 [] **No ICC Profile Applied** VOLUME

x1232 + 8 YBR FULL 422 19FG 1

DICOM WSI Color Management

Goal is color consistency throughout pixel pathway Color "improvement" (normalization) outside DICOM's scope Depends on International Color Consortium (ICC) profiles Consistent with approach adopted by scanner vendors All DICOM WSI require presence of an ICC profile IFF downstream systems are calibrated, consistency possible DICOMweb – client or server may apply profile Web-browser-based clients have distinct limits in this respect

DICOM WSI Compression

State of the art is modest level of lossy compression of WSI Mostly using baseline JPEG, some JPEG 2000 Possibility of doing better with new JPEG-XL More flexible (effective, complex) than baseline JPEG Can re-encode existing JPEG images without further loss Potentially save > 30% in size

For segmentation bit masks, need compression, since sparse Considering use of Deflate (zip), in future maybe JBIG2

DICOM and FDA

Recognized consensus standard – what does that "mean"? Understanding when pixel data is mathematically the same ... as the manufacturer's proprietary format already cleared ... as what will be displayed to the user (vs. already cleared) Compression (JPEG) is independent of wrapper (DICOM) What role does color management (ICC in DICOM) play? Are there any display-specific standard issues (like GSDF)? Analogy of plain X-Ray vs. mammography in radiology (relative perceived risk and classification) – eventually relaxed to "5MP", no processing, requirements, mix-and-match but not DICOM-related