

Minutes

MEETING NAME 09-WG33: WG-33 Data Archive and Management

MEETING PLACE/DIAL IN

DATE & TIME Wednesday, July 22, 2020 | 11:00 am – 12:30 pm US ET

PRESIDING OFFICERS Matthew Bishop, UnityPoint Health
 Keith Eklund, Healthcare Tech Solutions

VOTING MEMBERS PRESENT

Ambra Health	Ostrow, Daniel
Canon Medical Research USA Inc.	O'Donnell, Kevin
Canon/Vital Images, Inc.	Whitby, Jonathan
DesAcc EMEA Ltd	King, Graham
GE Healthcare	Nichols, Steven
Grafimedia	Georgiadis, Pantelis
Healthcare Tech Solutions	Eklund, Keith
Laitek, Inc.	Costea-Barluti, Razvan
Laitek, Inc.	Solomon, Harry
Mayo Clinic	Persons, Kenneth
Pixelmed Publishing	Clunie, David
Society for Imaging Informatics In Medicine	Bishop, Matthew
Society for Imaging Informatics in Medicine	Carey, Cheryl

OTHERS

CitiusTech Healthcare Technology Ltd.	Mahalle, Prashant
DeJarnette Research Systems	Wineke, Steve
Hyland Software Inc.	Ullrich, Mike
Ochsner Health System	Hayes, Matt

TFT

Bagheri, Majid

VOTING

AAPM

Bevins, Nicholas

MEMBERS

AAPM

Knazik, Shayna

ABSENT

Argentix Informatics

Silver, Elliot

Canon/Vital Images, Inc.

Dawson, Tim

Change Healthcare

Ho, Kinson

European Society of
Radiology

Mildenberger, Peter

GE Healthcare

Numan, Jouke

Laitek, Inc.

Sluis, Douglas

Mach7 Technologies

Ulanov, Alexey

Society for Imaging Informatics
in Medicine

Henson, Kyle

Varian Medical Systems

Schwere, Thomas

DICOM

Anna Zawacki, SIIM

SECRETARIAT

**1 CALL TO ORDER AND REVIEW OF ANTI-TRUST RULES AND DICOM
PATENT POLICY (Co-Chairs, Secretariat)**

The meeting was called to order. Guidelines for Conducting NEMA Meetings were read and attendance was recorded.

2 REVIEW AND APPROVE AGENDA (Co-Chairs)

The agenda was reviewed and approved.

3 REVIEW MINUTES (Co-Chairs)

The minutes of the previous meeting were reviewed and approved.

4 TOPIC ITEMS TO BE DISCUSSED (All)

- Resume reviewing (at #5/6) open issues to be addressed in the Supplement

ftp://d9-workgrps:Private15@medical.nema.org/MEDICAL/Private/Dicom/WORKGRPS/WG33/2020/2020-07-22/Consensus_Positions_and_Open_Issues-20200722.pptx

5a Use cases and Profiles

- Migration [Complete]
- Quality control [Advanced]
 - Provenance of updates
 - UID cross-references
 - Equipment
- EMR synchronization [Study-Only]
 - Includes complete study and patient info, and perhaps Modalities in Study, but no Series or Instance level info
- Minimal [?]

What are our use cases and what content do they need?

Migration - Robust set of metadata to facilitate export of data to the new PACS.

For QC processes – there are additional things that need to be retained.

For EMR synchro – maybe we only need study level info and not have to get down to series and instance level.

Thoughts on how rich the metadata in the inventory needs to be?

So, are we saying you've got access to instances or the only other way you are getting it is through this?

D. Clunie was talking about some use case that all you need is UIDs.

If the metadata does not need cohesion cause it's already correct in the reference files, then you really just need a list of stuff and arguably don't even need the UIDs. Goal just take a bunch of stuff and move it, then metadata is relatively minimal.

Little bit of data is relatively harmless, but the manifest size grows.

KOS – the need to provide hierarchical structure when it's unnecessary.

Lot's to be learned from DICOMDIR and KYOS.

Can agree that we need different levels of inventory content for different use cases

Don't need general capability for arbitrary content

Feels like it's a spectrum – at one end of it is something like David is describing, just the UID, and on the other end – how far does the spectrum go in the other direction, i.e. is it complete metadata for every single instance.

Are there particular points on the spectrum that we want to define as useful.

Is there a basic set of metadata.

If the receiver gets too much metadata, they can just ignore it.

There are probably 3-4 points on the continuum that are reasonable places to say – yes that's a set of metadata that would be appropriate for this sets of use cases.

Essentially profiling. Here is a list of attributes that would be expected to be present.

One that has 15, one that have 3, and 1.

Variability – when we think about a complete set:

- Complete as in all attributes for all files, or
- Complete as in all attributes that are different from what's on disc for all files

Some profiles will be list-based, and some will be rule-based.

Have a baseline and mandatory extensibility mechanism. Provider is adaptive to the needs of the consumer. Able to adapt to any metadata that the consumer is demanding. May be a bit too burdensome for the producer but may be worth considering. Ability for provider to reject the request.

Or consumer to interrogate the provider about what he can provide and then make a decision based on that.

Does the server need to confirm to the desires of the consumer or is it the other way around?

There is some space in between that we may be trying to find.

Main use case is migration, so we want minimal burden on the producer.

Conceptually is more or less a C-FIND but it's different in many ways.

How much we are trying to pay attention to optimization.

Very hard to read the DICOM header on every file, very time consuming for the producer, so if there is reliable metadata that is sufficient to produce your recipient index without having to consult the DICOM file, then that's a sweet spot in terms of optimization that we should enable, if the producer can satisfy it.

Don't exclude possibility of optimization.

Leave this open and we'll come back to it when we get to an actual draft.

5b

Open Issue – updated metadata provenance

- Use case – migration should preserve traceability of provenance of metadata updates
 - Minimally datetime of change, and source of changed data
- DICOM capability in Original Attributes Sequence ([Part 3 Section C.12.1.1.9](#))
- **? Proposal: inventory includes record of provenance of metadata updates**
 - In advanced profile only?

Need to define recording the provenance of metadata updates.

Original attributes sequence:

- Source of previous values
- Modifying system
- Reason for the modification

To the extent that you might be able to have this it would be nice to be able to record the provenance of changes

Make it an optional capability, for research applications it is faced a lot, collections of images do change a lot

6

Open Issue – Inventory data structure

- Options:
 - Hierarchical – similar to hierarchical Query response, DICOMDIR, FHIR ImagingStudy
 - Flat/denormalized – similar to relational Query response; no consolidated study or series entity
- **Proposal: inventory uses a hierarchical structure**

How to structure data

Harry – of the opinion it should be hierarchical structure to the data, reduces repetitions, duplications

Put your series level stuff at the series level and it should be different with different instances, but it happens

We have a normalized data model, but if the source PACS doesn't instantiate that model, do we need to accommodate deviations from the standard data model?

Do we want to migrate non-compliant datasets?

Study with two different dates and times, the producer doesn't know which one is correct

Strong argument for being able to send unnormalized data

Hybrid – send twice

They all have to come up with some answer to answer C-FIND queries

Whatever it is they have decided to be in the C-FIND request this will be the same

Hierarchical makes sense but need to flag that there may be some weird exceptions. This is a good candidate for considering a trial implementation.

7

Open Issue – distributed inventory structure

- Enterprise inventories may be enormous
- Inventory instance may apply to a described scope of subset (see #4)
- Use case: inventory instances might be created and persisted for partitions of the archive
 - Allows incremental or parallel production of inventory
 - Allows tractable object sizes
- **Proposal: inventory supports inclusion by reference of other inventory instances for subsets of archive**
- See note in Part 3 Section F.1: *The directory information has been defined so that a future version may be extended to support the distribution of the directory information among a logical tree of several files (with the DICOMDIR file at its root). However in this version, the entire directory information is specified to be stored in a single File with a DICOMDIR File ID.*

In what manner will the manifest be delivered?

Static file, set of files, and API that includes HTTP features?

What is produced by the inventory is a DICOM instance that will be accessed either through file system access or DICOMweb access or DICOM DIMSE access.

HTTP range is a physical offset-based

Is random access to the manifest a feature that we require?

Should we be able to include sub-directories/sub-manifests by reference?

Considered that back when we did DICOMDIR.

If we are thinking of this as instances. Once you have composed an instance and nothing has changed, you don't have to re-generate it.

Could need a mechanism to request a set of manifests with respect to a certain query.

If you have these manifests already created for some subsets of the archive, and there were no patient updates, those are good forever.

What's the instance of metadata updates? About 10%. But it happens all over the map when it happens.

If you can do it, it would simplify production and access.

If I ask for a result and you give it to me, if I open it and see there are 2 pointers I need to get, I need to reach back out. If I just stack it up somewhere and open 4 months later and see it's not accurate.

What is the longevity of the manifest you previously downloaded?

Manifests from the consumer perspective would be transient – get them, use and discard.

Last thing I want to see is pointers from one manifest to another manifest. Each manifest should stand alone. Keep it simple. A manifest is just a manifest.

Use the manifest as your source of truth instead of going back to the database.

For the response – am I sending you back a stream of records or I am handing you a pointer saying your answer is in this file, go get it.

For larger requests – it has to be the latter.

Have an escape path – I will respond with a file location.

Can I give you a list of 3 file locations and it's a combination of those.

Put on the producer to pre-compose that data – problem – it could be billions of records.

Don't like an idea of having an index of an index.

Keep it on the table, but they always need to contain records and not the pointers.

Table this issue for future experimentation.

8

Open Issue - inventory creation invocation

- Use case – apps external to PACS/VNA (e.g., migration controller) may need production of inventory of subsets of archive
 - Subset parameters may be determined dynamically
 - Asynchronous mechanism required – possible extended period to create inventory
 - SCP can set security permissions on which external applications can invoke service (authorizations not standardized in DICOM)
- **Requirement: creation of inventory may be invoked by standard service with asynchronous completion**
 - Mechanism specifies parameters of subset, profile level
 - Separate conformance claim from Inventory Storage SOP Class – could create inventory from PACS local admin UI without supporting this service

Asynchronous – I don't know now and ask me later. It's up to you to periodically come back and ask again.

8a

Open Issue – Mechanism of subset specification

- C-FIND Query uses simple and well-established key parameter search specification
 - But range structure in VR allowed only in Query – need to record in Inventory SOP Instance
- Hanging Protocol uses a more robustly specified, but somewhat complex, Filter Operation Sequence (0072,0400)
- **Proposal: use C-FIND style key parameter matching**

Possibly need to do a tweak to Part 5.

Need to define types of queries we need and pick mechanisms that can satisfy those.

8b

Open Issue – service message elements

- Requirement: asynchronous invocation / response, with progress indication
- Minimal (like Storage Commitment)
 - Initiate, Cancel: N-ACTION from SCU
 - Progress indication, Completion: N-EVENT-REPORT push from SCP
- Possible re-use/profile of Unified Procedure Step ([Part 4 Annex CC](#))
UPS Push + UPS Event SOP Classes
 - Initiate: N-CREATE from SCU
 - Cancel: N-ACTION from SCU
 - Progress indication: N-GET pull from SCU, N-EVENT-REPORT push from SCP
 - Completion: N-EVENT-REPORT push from SCP
 - Many attributes inappropriate for a non-patient related task
- Simplified version of UPS using N-CREATE, N-ACTION, N-GET, N-EVENT-REPORT

Have a WATCHED folder – content of which is always an active work

Need indication of when the producer is done

It's not the DICOM way.

Do it the simple way, not the DICOM complex way.

It should be on the list.

Need to figure out how it fits in the structure of DICOM standard.

Add the WATCH folder to the list.

PUB SUB type stuff? Any work on that in WG-27? Notification stuff?

Kind of done, but checking how it fits with other notifications.

How advanced things are with long-term subscriptions.

Easiest way to get a notification is via email.

All notifications can be replicated by polling.

add SMTP notification

12

Open Issue – inventory within DICOM IODs

- How do we characterize the inventory object amongst the DICOM standardized IODs
- Option 1 – “special” IOD, similar to Basic Directory IOD ([Part 3 Annex F](#))
 - Special attribute tags (0004,xxxx)
 - Special place in services architecture (limited to Media Storage Service)
- Option 2 – In category of Non-Patient Objects ([Part 3 Section 7.13](#))
 - Color Palette, Hanging Protocol, Defined Protocol, etc.
 - Normal composite (but single IE level) IODs
- **Proposal: Inventory instances are in class of DICOM Non-Patient Objects**

Presumption – it should be a DICOM object, but not a complicated one.

9

Open Issue – inventory anonymization

- Use case – research may require an anonymized inventory
- Many varieties of anonymization (See [Part 15 Annex E](#))
 - Allows placing original attribute values in Encrypted Attributes Sequence
- Anonymization is an optional parameter of the DICOMweb-URI Retrieve DICOM Instance Transaction ([Part 18 Section 9.4](#)), but not DICOMweb-RS or DIMSE protocols
 - IHE Teaching File and Clinical Trial Export Profile specifies a separate Export Manager actor to process anonymization
- **Proposal: use invocation parameter to control anonymization of inventory**
- **Alternative Proposal : anonymization is out of scope**

Use de-identification rather than anonymization

Separate work item

Doing it is out of scope, but anticipation of doing it is in scope

5 OLD BUSINESS

6 NEW BUSINESS

7 DATE AND TIME OF NEXT MEETINGS (Secretariat)

- Continue T-con meetings bi-weekly for the time being (at least until the New Work Item Proposal is submitted to DSC)
- Next call is August 5, 2020 between 11:00 am and 12:30 pm ET

<u>NEMALINK CODE</u>	09-WG33
<u>SUBMITTED BY</u>	Hull, Carolyn
<u>SUBMITTED ON</u>	8/12/2020
<u>LEGAL REVIEW</u>	Peter Tolsdorf, 8/14/2020
<u>UPLOAD LOCATION</u>	Enter upload location.