JSON Representation of DICOM Structured Reports

DICOM WG 23
David Clunie
Presentation of First Read Draft to WG 6
2019/09/09
AI CHANGES THE GAME

http://medium.com/adhive/disruptive-ai-controlled-advertising-cd90a07452cb
Annotation interoperability matters now

• Previously:
  • little incentive to annotate
  • few tools to create or view annotations
  • annotation interoperability was a low priority for product managers
  • presentation rather than semantics were the priority for annotation tools

• Now:
  • semantic annotations have (real monetary) value beyond primary use case
  • recognition of existence of unanticipated re-use cases
  • annotations are expensive to create/recreate retrospectively
  • more expensive to process if proprietary rather than OTS standard
  • AI-generated annotations need to be interoperable for display
  • “interactive” AI requires interoperable annotation exchange
  • AI vendors unlikely to be the same as scanner/PACS vendors – mix and match
DICOM SR and AI

• DICOM SR is a generic solution for:
  • fundamental encoding of measurements, categorical results, using codes and referencing images, waveforms as well as spatial and temporal coordinates
  • reusable sub-templates for specific scenarios that are common to different use cases and applications
  • generic root level templates for non-specific measurements (e.g., TID 1500)
  • linking other objects related to results and measurements (such as SEG, Parametric Map and RWVM)

• Specific templates for:
  • traditional CAD applications that are relevant to AI
  • traditional human operator measurements that may now be made by AI
DICOM SR and the developer

• Traditional DICOM SR encoding requires use of a toolkit and an API with a non-trivial learning curve (binary encoding intractable by hand)

• AI algorithm developer may not need to know about the “composite context” (patient/study/series +/- workflow metadata) of the encounter

• Impedance mismatch between
  • PACS-orientated “DICOM image in, DICOM SEG + SR out”
  • Algorithm-developer orientated “PNG in, PNG + JSON out”

• Even XML is deemed excessive/too complicated by AI developer community

• DICOMweb JSON encoding is also intractable for SR, since it is hexadecimal tag, individual data element orientated (no SR content item abstraction)
Goals for Simplified DICOM SR in JSON

• Full-fidelity round trip with actual DICOM SR for all constructs (any template)
• Simple (enough to hand write or copy from examples)
• Compact (even terse)
• Understandable (relatively)
• Unambiguous (easily parsable)
• Leverage any existing actual or de facto JSON or evolving AI standards
• Platform independent
• Capable of encoding extracts separated from composite context (such as without “header” rather than content tree, image library, etc., which could be added by separate tool/pass)
Pipeline to add missing stuff to JSON

AI Algorithm

Lesion Manager

DICOM Image Aware System

Patient-Study Aware System

PACS
Design Decisions – Business Names

• No hexadecimal numbers for “header” attributes – leverage DICOMweb JSON encoding but with PS3.6 keywords rather than numeric tags
• Abstract the content items (i.e., as if they were attributes)
• No obscure alphanumeric codes in content tree – use “business names” concept from Green CDA (not dissimilar to JSON-LD)
• Codes are defined in separate “business names” JSON file that acts as a dictionary – do not need to be standardized (but may be in future, like keywords)
Design Decisions – JSON Structure

• Use business name as name of JSON object’s name-value pair
• Use JSON arrays to preserve order
• Use JSON arrays to allow sibling objects with same name
• Use a JSON array to encode children of an object
• Omit explicit value type and relationship type if they can be deduced from context, or defined in the separate business names file
• Add annotations (specific object names starting with “@” symbol) to resolve ambiguities, and to provide target and source for by-reference relationships
Example 1 – hexdump of the original (partial)

```hex
0000390  20 00 11 00 49 53 04 00 34 35 37 38 20 00 13 00 | ...IS..4578 ...
00003c0  49 53 02 00 31 20 40 00 40 a0 43 53 0a 00 43 4f | IS..1 @ @.CS..COI
00003c0  4e 54 41 49 4e 54 20 40 00 43 a0 53 51 00 00 | INTAINER @.C.SQ..I
00003c0  ff ff ff ff fe ff 00 e0 ff ff ff ff 08 00 00 01 | ....................................
00003e0  53 48 06 00 31 32 36 30 30 30 08 00 02 01 53 48 | ISH..126000...SHI
00003e0  04 00 44 43 4d 20 08 00 04 01 4c 4f 00 49 6d | DCM ...LO..ImI
00003e0  61 67 69 6e 67 20 20 4d 65 20 73 65 20 6d 65 | Iaging MeasurementIt Report...
0000400  74 20 52 65 70 6f 72 74 ff 0d e0 00 00 00 00 | @..P..CS...
0000410  fe ff dd e0 00 00 00 00 40 00 50 a0 43 53 08 00 | ......................@..C.
0000420  53 45 50 41 32 41 54 45 44 45 50 48 07 a0 53 51 00 | SEPARATE@.x.SQ..I
0000430  ff ff ff ff fe ff 00 e0 ff ff ff ff 08 00 80 00 | ......................
0000440  4c 4f 00 00 08 00 82 00 53 51 00 00 ff ff ff | LO..SQ......
0000450  fe ff dd e0 00 00 00 00 40 00 01 11 53 51 00 00 | @...SQ...
0000460  ff ff ff fe ff dd e0 00 00 00 00 40 00 84 a0 | ......................
0000470  43 53 04 00 50 53 3e 4e 20 40 00 23 a1 50 4e 14 | CS..PSN @.@.PN..I
0000480  61 63 63 6f 6d 70 6f 72 74 6c 6f 77 6c 6f 77 6f | accomplished_peal
0000490  6f 00 00 00 00 00 00 00 00 00 00 00 00 00 00 | fowl...........
00004a0  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 | ...
00004b0  ff ff dd e0 00 00 00 00 40 00 75 a3 53 51 00 00 | ...@.R..SQ...
00004c0  ff ff ff fe ff 00 e0 ff ff ff ff 08 00 15 11 | ......................
00004d0  53 51 00 00 00 00 00 00 00 00 00 00 00 00 00 | SQ.............
00004e0  08 00 99 11 53 51 00 00 ff ff ff ff fe ff 0e 00 | ........P..UI..1..2..I
00004f0  ff ff ff ff 08 00 50 11 55 49 1a 00 31 2e 32 2e | 840..1008.5.1.4.1
0000500  38 34 30 2e 31 30 30 30 38 2e 35 2e 32 3e 34 2e | 1.1.2...UI@.I
0000510  31 2e 31 2e 32 2e 32 00 80 00 55 11 55 49 40 00 | 3.6.1..4..1.4519..I
0000520  32 2e 36 2e 31 2e 34 2e 31 2e 31 34 35 31 39 2e | 5.2.1.9203.4004..I
0000530  35 2e 32 2e 31 2e 39 32 30 33 3e 32 34 30 30 2e | 1268018422888185I
0000540  32 36 38 30 31 38 34 32 32 32 38 38 38 31 38 35 | 73226516023762..I
0000550  37 33 32 32 36 35 31 36 30 32 33 37 36 32 3e ff | 173226516023762..I
```
Example 1 – dcsrcmp of the original

: CONTAINER: (126000,DCM,"Imaging Measurement Report")  [SEPARATE] (DCMR,1500)
  >CONTAINS: CONTAINER: (126010,DCM,"Imaging Measurements")  [SEPARATE]
    >>CONTAINS: CONTAINER: (125007,DCM,"Measurement Group")  [SEPARATE]
      >>>HAS OBS CONTEXT: TEXT: (112039,DCM,"Tracking Identifier")  = "5b6eb4301d3175942d29985a3d0fbb00"
      >>>HAS OBS CONTEXT: UIDREF: (112040,DCM,"Tracking Unique Identifier")  = "1.3.6.1.4.1.5962.1.1.0.0.0.1535644357.22655.1"
      >>>CONTAINS: NUM: (G-D7PE,SRT,"Length")  = 66.43856134 (mm,UCUM,"mm")
        >>>INFERRED FROM: SCOORD: = POLYLINE (172.835357666016,270.064086914062,133.798889160156,343.045318603516)
          >>>SELECTED FROM: IMAGE:  = (1.2.840.10008.5.1.4.1.1.2,1.3.6.1.4.1.14519.5.2.1.9203.4004.26801842288810573226516023762)
Example 1 – JSON of the content tree (only)
{
  "TrackingUniqueIdentifier": [
    "1.3.6.1.4.1.5962.1.1.0.0.0.1535644357.22655.1"
  ],
  {
    "Length": [
      "66.43856134",
      "mm",
      [
        {
          "": [
            "POLYLINE",
            [
              172.83535766601562,
              270.0640869140625,
              133.7988916015625,
              343.0453186035156
            ]
          ],
          [
            {
              "": [
                "1.2.840.10008.5.1.4.1.1.2",
                "1.3.6.1.4.1.14519.5.2.1.9203.4004.268018422288818573226516023762"
              ]
            }
          ]
        }
      ]
    ]
  }
}
Example 1 – JSON of result only (no id, coord)

```
"ImagingMeasurementReport": [
  {
    "ImagingMeasurements": [
      {
        "MeasurementGroup": [
          {
            "Length": [
              "66.43856134",
              "mm"
            ]
          }
        ]
      }
    ]
  }
]
```
Example 1 – Business Names file (partial)
HAS ACQ CONTEXT

{
  "FindingSite": {
    "@cv": "G-C0E3",
    "@csd": "SRT",
    "@cm": "Finding Site",
    "@vt": [
      "CODE"
    ],
    "@rel": [
      "HAS CONCEPT MOD"
    ]
  }
}

,
Current Discussion and Open Issues

• Some folks are distressed by the terseness and nesting of objects and arrays
• Proposals for simplification – e.g., omit array of values when only a single value
• Proposals for clarity at the expense of compactness – e.g., include annotation to indicate type rather than relying on positional parameters, such as for graphic type, coordinates, etc.
• Concern about “anonymous” concept names (no name for name-value pair) allowed by both SR infrastructure and many templates, esp. for IMAGE and SCOORD content items – harmless but lead to lots of nesting
• Adding higher level abstractions (than content item level) might address some of these concerns, but also make for a more complicated parser
Out of Scope (for this development cycle)

- A DICOMweb API to transform JSON SR to/from the standard binary DICOM SR persistent form (in a WADO-RS or STOW-RS “application/dicom+json” like manner, e.g., “application/x.dicom+json”)
- A DICOMweb API to access, create or modify the DICOM SR content tree abstraction (cf. the existing RetrieveMetadata individual DICOM attribute level access)
- A DICOMweb API to create and manage individual (or sets of) annotations separately from the storage/retrieval of entire DICOM SR object
- A DICOMweb API to perform/manage the various steps of the authoring pipeline that adds lesion management, image references and descriptions, and patient/study/series/workflow composite context