

Strawman statement of requirements: Compound Documents

Rev. 0.2, September 11, 2001

Motivation and Definition:

Use Cases:

Requirements for Compound Documents:

Open Issues:

Motivation and Definition:

The futuristic vision is that an electronic health record (EHR) will be constructed from multiple CDA documents created in diverse practice settings. A document management system will manage all those CDA documents, and index them according to the various markups they include. The document management system will be able to get queries and dynamically retrieve a subset of the documents based on some fields e.g. get all documents of John Doe during a specific period. However, sometimes there is a need to aggregate several documents that cannot be aggregating by posing queries on the document management system. Sometimes, there is a need for a document that statically aggregates several other CDA documents.

A **compound document** is a document that statically aggregates several CDA documents by including references to those CDAs. In many cases, the compound document includes also its own content about the aggregated group of documents. The compound document semantically (not in-line) includes the referenced documents and is not considered valid without the existence of those referenced documents. However, applications may choose to exchange just the compound document as long as they have the means to get the referenced documents per demand.

A compound document differs from <observation_media> because each referenced document is a regular CDA document with its own authentication. A compound document differs from <link_html> because it is strongly connected to the referenced documents. So perhaps, the links in a compound document will require some of the functionality of the <onservation_media> such as the need to be able to retrieve the referenced documents, the need to express the instance identifier and checksum of the referenced documents, etc.

Use Cases:

- DICOM, which is a popular standard in the radiology domain, has the notion of a patient that includes several studies. The study includes several series; each series is taken from a different modality and may contain several images. Also, a radiologist report is generally attached to the study or series. We think that a series will be represented in a CDA document. Then, a study compound document will include references to all its series documents, a reference to the study report, and its own content such as modalities_in_study (a DICOM tag).

- In the bone marrow transplantation (BMT) domain, there are several documents such as disease document, donor document, conditioning document, etc. A BMT compound document includes references to all those BMT documents that relate to the same transplantation episode as well as its own content such as `transplantaion_episode_number`. Each referenced document has its own authentication, and in fact the disease document is created in another department of the hospital. The compound document is not valid and not accepted in the external BMT registry without the existence of all the referenced documents.

Requirements for compound documents:

1. It should be simple for a document management system to identify a document as a compound document.
2. A compound document is a CDA document, so it has its own header including authentication and version.
3. A compound document reference other CDA documents and semantically (not in-line) includes them i.e. the compound document is not valid without the existence of the referenced documents.
4. A compound document may include its own content in addition to the references to other CDAs.
5. A compound document can be either level 1, 2, or 3, and it can reference CDA documents of any level. The level of the compound document is decided according to the granularity of the markup of the document own content, and does not depend on the level of the referenced documents. For example, the compound document can be of level 1 and reference two documents – one in level 2 and one in level 3.
6. A compound document can reference documents with different version numbers (`<version_nbr>` element).
7. The authentication of the compound document applies to the content in the compound document and not to the content in the referenced documents. Each referenced document has potentially its own authentication.

Open Issues:

1. How the document management system knows that a document is compound? It seems that the document management system can better manage the documents if it knows easily when a document is a compound one. Do we need a specific `document_type_cd` for compound documents?
2. How a compound document has references to other CDAs? Do we use the regular `<link>` and `<link_html>`? Do we use `<observation_media>`? Do we need to develop new means for referencing? Should we have `<link_cda>` under `<link>`?
3. Are there any restrictions on confidentiality of the compound and referenced documents? Should we restrict that the compound and referenced documents should have the same confidentiality?