Metadata Requirements for Portals

Tim BENSONa,1
a Abies Ltd, R-Outcomes Ltd and UCL CHIME

Abstract. Consensus around the requirements for metadata in patient and clinical portals would provide a sound basis for the adoption of standards. We propose a set of requirements for metadata in a way that is generic and platform independent. These requirements cover both Clinical Documents and Clinical Statements, addressing the what, who, when and where of each item.

Keywords. Electronic Health Records, Information Storage and Retrieval, Metadata, Patient Portals, Clinical Portals

Introduction

Metadata is information about any information item that is used in search to find it later. The Dublin Core [1] is an example of a successful metadata standard that is widely used across the Internet for indexing public documents. The UK e-Government Metadata Standard (e-GMS) [2] set out to standardise metadata across government, building on the Dublin Core foundations.

The ebXML registry standard [3] provides a broad specification for document portals, used in business and provides the foundation on which IHE (Integrating the Health Enterprise XDS (Cross-enterprise Data Sharing) has been built [4]. IHE XDS is the basis of many patient portals. HL7 CDA (Clinical Document Architecture) has been widely adopted in health care and contains a common document header [5], which can be thought of as metadata. However, IHE XDS and CDA were specified by different groups and are not fully aligned. Furthermore the high level of optionality in both standards allows implementers wide scope for local modifications. All elements are optional and allow an unlimited number of repeats. Implementers need to specify the local Affinity Domain (XDS) and/or Templates (CDA).

Portal architecture usually includes a central registry and one or more repositories, which may be physically and logically separate. Additional functionality is required to manage identity and access control (not discussed here). A book library provides a good analogy. The library has shelves of books and a central index. An index card points to the shelf address of each book. Similarly, the portal registry contains metadata describing the content of each item in repositories and a pointer to its location.

Common metadata is required for retrieving and sharing data that has originated from multiple heterogeneous sources, as typically found in health and care services. This supports information sharing horizontally across each patient’s wider web of care, including health, social and voluntary sector services and family, friends and carers.

1 tim.benson@abies.co.uk
Metadata may also support vertical interoperability for secondary uses in management, commissioning, analysis and research.

Health portals typically contain items at two distinct levels of granularity, which we refer to as Clinical Documents and Clinical Statements. These have different uses and granularity, but need similar metadata to enable retrieval. Both share common properties of persistence, coherence, wholeness, human readability, stewardship and potential for authentication.

A Clinical Document is a discrete electronic composition about an identified patient, intended to be read or used by a human, like a paper document, such as a report or letter. Examples of Clinical Documents include pdf, web pages, medical images, audio and video.

A Clinical Statement is the smallest meaningful category of stand-alone medical information. Think of it as a single line in notes with an independent existence, which can be selected and combined with others in different ways to generate new ways of looking at data. Examples include a diagnosis, medication entry, procedure or test result, such as a blood pressure reading. They may be selected and reordered on the fly to compile lists or charts from multiple sources, sorted by type, date, author or source.

It is usual to start with Clinical Documents and extend later to cover Clinical Statements. However this extension is only possible if the architecture is designed from the outset to handle both.

1. Methods

This paper builds on work undertaken as part of the Sintero Project, funded by the Wellcome Trust at Cardiff University 2009-11 [6] and health interoperability standards, including HL7, IHE XDS and SNOMED CT [7].

The requirements for metadata include:

(1) Information about what each item’s content, who created it about whom, when and where.

(2) Metadata must be complete, because if one item contains a value for a metadata element but a similar item containing similar item does not, then search will find the former but not the latter. This can be a real clinical safety risk. It is safer to specify a small set of mandatory elements, than a larger optional set.

(3) Metadata can only be the union of information coming from multiple source systems. It needs to be standardised. Suppliers and their customers will wish to support the minimum number of options, preferably just one.

(4) All metadata elements should be required (not optional) and multiple instances not supported. Each item’s metadata relates to exactly one patient, one creator and one creation time. If more than one, then add additional records.

(5) Metadata is computer-processed, requiring unambiguous codes, identifiers and dates. Free text is often ambiguous.

(6) Metadata is derived from multiple source systems, and should avoid the limitations imposed by any specific implementation platform. It needs to be platform-independent.
2. Results

2.1 What

Metadata includes what the item is. There are ten relevant items:

**Class** (required code) describes the general kind of item using a course-grained classification. It is always important to find and review all applicable records and not miss relevant items that have been given a slightly different name or fine-level category. These considerations point to using a small finite set of well-understood and mutually exclusive codes, so we will not miss important information. Examples of Class codes for Clinical Documents are: clinical notes, correspondence, order, report and image. Examples of Class codes for Clinical Statements include: problem (including diagnosis and allergy), medication, procedure, finding, history, and actual care event, proposed care event, family history and demographics.

**Type** (required code) describes the item at a fine level of granularity. This describes the type of item in some detail, but should exclude information about the clinical specialty or mode of care (which are recorded as Specialty and CareType respectively). A candidate list of document types has been developed by ACIG (Academy of Royal Colleges Information Group). This list has not been published but the UK Terminology Centre has provided each item with local SNOMED CT codes. The list of possible Type codes for both Clinical Statements and Clinical Documents is likely to be quite long (hundreds or thousands).

For Clinical Documents the Type is usually subsumed by a single Class (for example discharge summaries are a type of clinical correspondence). However, this is often not the case for Clinical Statements. In Clinical Statements Class may usefully be used to distinguish between events that are planned and those that have actually taken place, between those about the patient and about others (family history), or to indicate negative findings (in negative findings, subsumption works upside down – knowing someone does not have liver cancer does not imply that they have neither cancer nor a liver problem). This is one reason why we need to have separate Class and Type codes.

**Specialty** (required code) represents the clinical specialty or service responsible and should be specified as precisely as possible.

**CareType** (required code) represents the type of event that gave rise to this item, such as inpatient, outpatient or out-of-hours service.

**Title** (required) represents the title of the item in a human-readable text form suitable for display in a browser, for example “St Mary’s Hospital Urology inpatient discharge summary”. This is the only metadata item that is free text.

**Confidentiality** (required code) represents the level of sensitivity of the item. The default is normal. This can be useful to support access control.

**Language** (required code) records the original language (ISO 639 format).

**Format** (required code) specifies the precise format of the item. It will specify whether the item is a Clinical Document or a Clinical Statement, and will also provide finer-level detail to assist computer processing, such as MIME type.

**UUID** (required) is a universally unique identity for each Clinical Document or Clinical Statement. It may be a URL.

**OriginalID** (required) is an identifier of the Clinical Document or Clinical Statement instance assigned by the creator.
Table 1. Examples of metadata for a clinical document and a clinical statement

<table>
<thead>
<tr>
<th>Example</th>
<th>Urology inpatient discharge summary</th>
<th>Patient weight recorded at home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Correspondence (Clinical Document)</td>
<td>Finding (Clinical Statement)</td>
</tr>
<tr>
<td>Type</td>
<td>Final discharge summary</td>
<td>Weight</td>
</tr>
<tr>
<td>Specialty</td>
<td>Urology</td>
<td>Patient-initiated</td>
</tr>
<tr>
<td>CareType</td>
<td>Inpatient</td>
<td>Patient (alone)</td>
</tr>
<tr>
<td>Title</td>
<td>Urology discharge summary</td>
<td>Weight record</td>
</tr>
<tr>
<td>Confidentiality</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>

Two examples, for a discharge summary (clinical document) and a weight measurement (clinical statement) are shown in Table 1. These omit all IDs, Language Format and date/times.

2.2 When

*CreationTime* (required) is the time that the item was created. This should be recorded as a date/time, accurate to the nearest second, so that items can be listed in strict chronological, or reverse chronological order.

*EventTime* (required) represents the date of the event that this item relates to. This is usually either earlier or the same as the CreationTime (an exception is for future appointments). For events that have duration, the creator needs to select the most appropriate date. EventTme should not be recorded with more precision than is known.

2.3 Who

Each Clinical Document or Clinical Statement refers to exactly one identified patient and has one creator.

*PatientID* (required) is the patient’s Id provided by the originating item. Items without a patient Id are not supported. The patientId should be a common identifier such as the NHS number, but other identification number schemes, may also be used, so long as each is associated with a recognised and unique id scheme identifier. This extra flexibility allows extension to all countries, across the wider health and social care divide, or where legacy source systems do not contain standardised IDs.

Patient name, sex, date of birth and address are important for checking, but are not metadata. Name is not stable and people can change this at will. Similarly, some people alter their apparent sex and age. Some people have no address and it can change frequently.

*CreatorId* (required) is a unique identifier of the author or organisation responsible for creating this item at the time of creation. The searcher is usually looking to know who was responsible for the item in the first place. It may be more important to know that an item comes from a specific clinical team than the actual author. The author should be stated in the item itself. The Creator may be the patient.

There is no clear requirement to know the intended destination of the item, although this can often be found within the item itself.
2.4 Where

Location (required) is the identifier for the organisation location responsible for creating the item.

2.5 Extensions

While optionality is deprecated, there are some occasions when it is justifiable to provide an extension mechanism. Optional tags provide this.

Tag (optional code) provides a way to tag the item in an additional precise way, which will be use case specific. SNOMED CT codes are recommended.

3. Discussion

Metadata is used in search outside the originating organisation, not to support internal business transactions locally. In consequence, metadata does not include transaction-specific information, even though transaction-specific items are routinely found within clinical records.

Metadata is inevitably limited by what can be provided by source systems. We should not assume that anyone will ever actually retrieve any specific item, or that items found in a portal are complete and up-to-date (although each item should be accurate at the time it is created).

The historical failure to address and build consensus around stringent requirements for common metadata has delayed the progress of clinical and patient portals and increased the costs of deployment. This paper sets out a set of stringent requirements and a candidate set of items and constraints that meet these requirements.

This is low-hanging fruit, which can enable health interoperability in a way that is particularly useful for patients and clinicians.

References