Consolidating CDA Documents from Multiple Data Sources

Session 7, February 20, 2017
Brian E. Dixon and Masoud Hosseini

DISCLAIMER: The views and opinions expressed in this presentation are those of the author and do not necessarily represent official policy or position of HIMSS.
Speaker Introduction

Brian E. Dixon, MPA, PhD, FHIMSS
Research Scientist
Regenstrief Institute, Inc.

Masoud Hosseini, MS, PhD
Doctoral Graduate
Indiana University School of Informatics and Computing
Brian E. Dixon, MPA, PhD, FHIMSS
Dr. Dixon has a 1/8th appointment at the U.S. Department of Veterans Affairs, making him a part-time federal employee.

Masoud Hosseini, MS, PhD
Has no real or apparent conflicts of interest to report.
Agenda

• Introduction to Interoperability and the Clinical Document Architecture
• The Challenge of Document-Centric Health Information Exchange
• DeDupIT: An Approach to Reducing Information Overload
• Evaluation of DeDupIT in the Indiana HIE
• Other Potential Solutions to the Challenge
• Future Directions
Learning Objectives

• Identify the challenges inherent when exchanging multiple CDA documents for a patient across multiple facilities within a health system or community.

• Describe the results of an evaluation of a system designed to consolidate data within multiple Clinical Document Architecture (CDA) Documents for a single patient.

• Propose solutions for addressing outstanding issues in resolving conflicting information in CDA documents received from multiple facilities within a health system or community.
The “E”s: Efficient, Exchange, Easy to Use

Reduce Information Overload
By consolidating data in C-CDA documents
What is Interoperability?

• The ability for one health information system to receive data or information from another system without prior negotiation, and have data queries and business rules work reliably against the received data or information (Dolin, 2011)

• The ability of different information systems and software applications to communicate, exchange data and use the info exchanged (HIMSS, 2013)
Interoperability Building Blocks

- **Semantic**
  - Vocabulary
    - The specific words that need to be used in the letter

- **Syntactic**
  - Content structure
    - The structure and specific type of information in the letter or package – how to write a business letter

- **Technical**
  - Transport
    - The method used to move the letter or package (e.g., truck, plane, boat) from point A to point B
  - Security
    - Sealing the envelope or the package
  - Services
    - Delivering to intended recipient and being able to find their address
HL7 Clinical Document Architecture (CDA)

- CDA is a standard which provides a common architecture, coding, semantic framework, and markup language for the creation of electronic clinical documents.
- It is XML based, so it can be rendered as HTML or PDF (human readable).
- Exchange of medical information without loss of meaning.
- Include pertinent clinical, demographic, and administrative data for a specific patient.
CDA Document

Header
- <id>
- <code>
- <title>
- <recordTarget>
- <patient>

Body
- <structuredBody>
  - <section>
  - <code>
  - <title>Vital Signs</title>
  - <text>Temp is 98.6</text>
  - <entry>
    - <observation>
      - <code: value, system, name>
      - <classCode>
      - <moodCode>
      - <statusCode>
      - <effectiveTime>
      - <value>

Source: http://www.healthit.gov/~Consolidated CDA Overview
Continuity of Care Document (CCD)

- CCD is an electronic document exchange standard for sharing patient summary information. (NOT complete but critical)
- CCD is one of CDA templates.

**CCD Template Sections**

1. Problems
2. Procedures
3. Family history
4. Social history
5. Payers
6. Advance directives
7. Alerts
8. Medications
9. Immunizations
10. Medical equipment
11. Vital signs
12. Functional status
13. Results
14. Encounters
15. Plan of care
What the CCD Is / Is Not

• CCD is
  – A summary of an episode of care
  – Useful for transfer of care scenarios

• CCD is not
  – A complete longitudinal health record
  – A replacement for more specific document types
    • History and Physical
    • Operative Note
    • Discharge Summary (specifically excluded from CCR/CCD)
Drivers for Using CCDs / C-CDA

- HITSP Adopted it for its C32 specification
  - Required in Stage 1 MU
  - Required for use in the eHealth Exchange

- Many EHR vendors and providers latched on to the CCD concept
  - While originally CCD used, many platforms now use C-CDA
  - C-CDA Use in Stage 2 and Stage 3 MU
Problem Statement

• Because CCDs (and now C-CDAs) are NOT longitudinal medical records, they often contain duplicate medical information

• Many HIE interfaces ask physicians to look at lists of CCDs / C-CDAs when viewing the patient’s past medical history
  – Information overload
  – Poorly perceived ease of use
CCDs from Multiple Sources

- HIEs receive CCDs from multiple sources that exchange information with each other
**Example from the VA**

Click the ‘Summarization of Episode’ you would like to view

<table>
<thead>
<tr>
<th>Continuity of Care Document</th>
<th>Multicare Health System:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MedVirginia Summarization of Episode Note</td>
<td>MedVirginia - HIE:</td>
</tr>
<tr>
<td>Summarization Of Episode</td>
<td>Inland Northwest Health Services (INHS) - HIE:</td>
</tr>
<tr>
<td>Summarization Of Episode</td>
<td>San Diego Beacon UCSD - MIRTH:</td>
</tr>
<tr>
<td>Summarization of episode note</td>
<td>Pensacola Health Information Exchange (HIE):</td>
</tr>
<tr>
<td>SUMMARIZATION OF EPISODE NOTE</td>
<td>Idaho Health Information Network (HIE):</td>
</tr>
</tbody>
</table>
VistAWeb View of the Allergy Domain

*Note: VA and Non-VA Allergies are combined in this view

<table>
<thead>
<tr>
<th>Reactant</th>
<th>Allergy Type</th>
<th>Verification Date</th>
<th>Observation/Historical</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>propensity to adverse reactions to drug</td>
<td>09/23/2009 00:00</td>
<td></td>
<td>Kaiser Permanente#</td>
</tr>
<tr>
<td>Aspirin</td>
<td>propensity to adverse reactions to drug</td>
<td>08/01/2013 00:00</td>
<td></td>
<td>Kaiser Permanente#</td>
</tr>
<tr>
<td>Beer</td>
<td>propensity to adverse reactions to food</td>
<td>04/14/2012 00:00</td>
<td></td>
<td>Kaiser Permanente#</td>
</tr>
<tr>
<td>Cats</td>
<td>Propensity to adverse reactions</td>
<td>03/04/2013 00:00</td>
<td></td>
<td>Utah HIN#</td>
</tr>
<tr>
<td>Cheddar Cheese</td>
<td>FOOD</td>
<td>08/26/2010 11:22</td>
<td>HISTORICAL</td>
<td>EAST ORANGE-VA NEW JERsey HCS#</td>
</tr>
<tr>
<td>Chocolate Concentrate</td>
<td>Drug allergy</td>
<td>02/05/2013 00:00</td>
<td></td>
<td>Idaho Health Data Exchange#</td>
</tr>
<tr>
<td>Dog Dander</td>
<td>propensity to adverse reactions to substance</td>
<td>02/12/2014 00:00</td>
<td></td>
<td>Kaiser Permanente#</td>
</tr>
<tr>
<td>Egg/White Poultry</td>
<td>Allergy</td>
<td></td>
<td></td>
<td>WILLIAM BEAUMONT AMC-FT. BLISS+</td>
</tr>
<tr>
<td>Eggs</td>
<td>DRUG,FOOD</td>
<td>04/14/2009 14:54</td>
<td>HISTORICAL</td>
<td>EL PASO VA HCS#</td>
</tr>
<tr>
<td>EggS</td>
<td>DRUG,FOOD</td>
<td>06/14/2010 09:14</td>
<td>HISTORICAL</td>
<td>HAMPTON (VAMC)#</td>
</tr>
<tr>
<td>EggS (Edible)</td>
<td>propensity to adverse reactions to drug</td>
<td>08/01/2013 00:00</td>
<td></td>
<td>Kaiser Permanente#</td>
</tr>
</tbody>
</table>
The Indiana HIE

- Over 58 sites within the IHIE network regularly began sending C-CDA documents in 2014

- To date, over 1 million C-CDA documents have been exchanged
  - 176,000+ from one health system
  - More than 50% had 2+ documents
  - Around 10% patients had 10+ documents
Enter DeDupIT

• Hoosier Healthcare Innovation Challenge
  – Local competition for innovative software to solve complex health care problems

• One Use Case focused on CCD Duplication
  – Sponsored by Indiana HIE
  – Winner would have eternal fame and glory
  – Solution had to be open source
What Does the Software Do?

• Consolidates multiple C-CDA documents into one

• Removes coded elements that are the same
  – Semantic identifiers need to be equal for the same date of observation and result

• Output is a “master C-CDA document”
System Architecture

Application API (REST, SOAP, Java, etc.)
- List of CCDs
- Merged CCDs

CDA Consolidation Engine
- Audit Record

Audit

MongoDB
- MapReduce Processor

Configuration User Interface

List of CCDs
- Merged CCDs

Pre-Format Rules
Primary Rule
Duplication Rules
Post-Format Rules

Evaluation of DeDupIT

- Preliminary Evaluation (In vitro)
  - 150 synthetic CCDs

- Real-World Evaluation (In vivo)
  - 176,169 C-CDA documents in the wild

- Impact on Clinical Workflow and Decision-Making
  - Tested software with 9 providers using human factors methods
Preliminary Results (Aim 1)

- 100% consolidation accuracy compared to manual de-duplication
- Time for 150 CCDs is 169 seconds (1.13 sec per CCD)
- CCDs are valid and compliant with NIST validator post hoc

<table>
<thead>
<tr>
<th></th>
<th>A: Number of Entries in Input CCDs (N=150)</th>
<th>B: Number of Entries in Consolidated CCDs (N=50)</th>
<th>Percentage Decrease from A to B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems Section</td>
<td>7,511</td>
<td>3830</td>
<td>49%</td>
</tr>
<tr>
<td>Medication Section</td>
<td>4,822</td>
<td>1900</td>
<td>60.6%</td>
</tr>
<tr>
<td>Allergy Section</td>
<td>1,309</td>
<td>275</td>
<td>79%</td>
</tr>
<tr>
<td>File size on disc (KB)</td>
<td>92,205</td>
<td>39,199</td>
<td>57.5%</td>
</tr>
<tr>
<td>Number of lines in CCDs</td>
<td>1,636,385</td>
<td>687,192</td>
<td>58%</td>
</tr>
</tbody>
</table>
Real-World C-CDA Docs from IHIE

• 176,169 C-CDA documents from 24 health facilities (hospitals)

• Randomly sampled 50 patients with at least 2 C-CDA documents exchanged during a 16-month period
  – N=30,370; Range: 2 to 36 C-CDAs per patient

<table>
<thead>
<tr>
<th>Number of Patients</th>
<th>Number of CCDs</th>
<th>Problem entries</th>
<th>Allergy entries</th>
<th>Medication entries</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>522</td>
<td>6,531</td>
<td>1,845</td>
<td>3,255</td>
<td>11,631</td>
</tr>
</tbody>
</table>
Results of Real-World Test

- The 522 C-CDA documents were manually de-duplicated AND ran through the system independently
  - 0.38 sec per CCD (compare to 150 manual hours)

<table>
<thead>
<tr>
<th></th>
<th>Problems</th>
<th>Allergies</th>
<th>Medications</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of entries in 522 CCDs</td>
<td>6,531</td>
<td>1,845</td>
<td>3,255</td>
<td>11,631</td>
</tr>
<tr>
<td>Decrease in number of entries through manual deduplication (gold standard)</td>
<td>89.9%</td>
<td>91.9%</td>
<td>57.3%</td>
<td>81.1%</td>
</tr>
<tr>
<td>Decrease in number of entries through auto-system deduplication</td>
<td>89.1%</td>
<td>79.9%</td>
<td>52.5%</td>
<td>77.4%</td>
</tr>
</tbody>
</table>
Reasons for Inaccuracy of System

- Reference to free-text rather than coded entry
  - Problems, Allergies, and Medications

- Conflicting codes used for the same medication
  - RxNorm encapsulates dosage along with ingredient
  - No translation codes provided in C-CDA
Example: Text Reference

- No code for the problem and only reference to the text section of CCD (GI Bleed, Uterine CA, Vision Impairment)

```xml
<observation classCode="OBS" moodCode="EVN">
  <templateId root="2.16.840.1.113883.10.20.22.4.4" />
  <id root="" />
  <code code="55607006" displayName="Problem" codeSystem="2.16.840.1.113883.6.96" />
  <text>
    <reference value="#PROBLEM" />
  </text>
  <statusCode code="completed" />
  <effectiveTime>...</effectiveTime>
  <value xsi:type="CD" nullFlavor="UNK">
    <originalText>
      <reference value="#PROBLEM" />
    </originalText>
    <value>
      <author>...</author>
      <entryRelationship typeCode="REFR">...</entryRelationship>
    </value>
  </value>
</observation>
```

--- Merged Information by CCD DeDuplication Engine ---

<content ID="PROBLEM">Possible Postmenopausal osteoporosis(<content ID="CONFirmed">Confirmed</content>)</content>
Human Factors

Patient 1
- 3 CCDs
- 8 problems
- 5 allergies
- 16 meds

Patient 2
- 1 CCD
- 8 problems
- 5 allergies
- 16 meds

Patient 3
- 5 CCDs
- 5 problems
- 4 allergies
- 17 meds

Patient 4
- 1 CCD
- 5 problems
- 4 allergies
- 17 meds

Patient 5
- 7 CCDs
- 6 problems
- 4 allergies
- 22 meds

Patient 6
- 1 CCD
- 6 problems
- 4 allergies
- 22 meds

Time recorded:
- Referral
- Med rec
- Problem rec

Before after before after before after
NASA TLX

Task Questionnaire - Part 1

Click on each scale at the point that best indicates your experience of the task.

Mental Demand
Low
High
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand
Low
High
How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand
Low
High
How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance
Poor
Good
How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort
Low
High
How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration
Low
High
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
### Results – Perceived Workload

<table>
<thead>
<tr>
<th></th>
<th>Referral scenario</th>
<th>Med rec scenario</th>
<th>Problems rec scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R1 Mean (SD)</td>
<td>R2 Mean (SD)</td>
<td>p-Value</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>131.2 (89.6)</td>
<td>54.1 (36.4)</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>MD</strong></td>
<td>39.4 (25.1)</td>
<td>24.4 (13.6)</td>
<td>0.041</td>
</tr>
<tr>
<td><strong>PD</strong></td>
<td>16.7 (10.6)</td>
<td>12.2 (7.5)</td>
<td>0.026</td>
</tr>
<tr>
<td><strong>TD</strong></td>
<td>31.1 (27.0)</td>
<td>16.1 (10.8)</td>
<td>0.020</td>
</tr>
<tr>
<td><strong>PE</strong></td>
<td>27.2 (16.8)</td>
<td>20.6 (15.5)</td>
<td><strong>0.052</strong></td>
</tr>
<tr>
<td><strong>EF</strong></td>
<td>28.9 (22.7)</td>
<td>22.8 (15.2)</td>
<td><strong>0.128</strong></td>
</tr>
<tr>
<td><strong>FR</strong></td>
<td>22.2 (26.5)</td>
<td>12.8 (12.3)</td>
<td><strong>0.179</strong></td>
</tr>
<tr>
<td><strong>OW</strong></td>
<td>32.4 (21.7)</td>
<td>20.3 (10.2)</td>
<td>0.032</td>
</tr>
</tbody>
</table>

MD=Mental Demands, PD=Physical demands, TD=Temporal Demands, EF=Effort, PE=Performance, FR=Frustration, OW=Overall Workload
# Results – Efficiency

## Percentage reduction after consolidation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Referral scenario</th>
<th>Med rec scenario</th>
<th>Problem rec scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>58.8%</td>
<td>38.1%</td>
<td>65.1%</td>
</tr>
<tr>
<td>MD</td>
<td>38.0%</td>
<td>19.8%</td>
<td>44.9%</td>
</tr>
<tr>
<td>PD</td>
<td>26.7%</td>
<td>26.8%</td>
<td>29.4%</td>
</tr>
<tr>
<td>TD</td>
<td>48.2%</td>
<td>36.1%</td>
<td>29.7%</td>
</tr>
<tr>
<td>PE</td>
<td>24.5%</td>
<td>10.6%</td>
<td>26.3%</td>
</tr>
<tr>
<td>EF</td>
<td>21.2%</td>
<td>16.8%</td>
<td>24.6%</td>
</tr>
<tr>
<td>FR</td>
<td>42.5%</td>
<td>19.7%</td>
<td>22.0%</td>
</tr>
<tr>
<td>OW</td>
<td>37.2%</td>
<td>18.4%</td>
<td>31.5%</td>
</tr>
</tbody>
</table>

MD=Mental Demands, PD=Physical demands, TD=Temporal Demands, EF=Effort, PE=Performance, FR=Frustration, OW=Overall Workload
Interview questions (open-ended)

• What are the challenges you face in finding information from patient’s medical records in general?
  – Too much information (I need relevant information).
  – Uncertainty (missing, inaccurate or not complete data).
  – Reconciliation is a huge time commitment.
  – Inconsistencies (conflicts in different documents).
  – Finding information (Where to look).
  – Complex UI (I am not sure where the information is).
Open-ended questions

• If you had a tool that could bring together the information from medical documents into a single view, would this be helpful to you? If so, how?
  – Helps finding different types of information easily.
  – Less time
  – Less searching
  – Less clicking
Summary of DeDupIT Evaluation

• Multiple documents for the same patient exist, are growing, and can lead to information overload if unchecked
  – >30,000 patients with 2 to 36 documents a year

• The DeDupIT approach quickly and accurately reduced information volume on demand for C-CDA documents*

• Consolidating information was well received by providers and resulted in meaningful reductions in work and cognitive loads
Other Potential Solutions

• The Transport Record Summary (TRS) Constructor
  – Another doctoral thesis project from Towson University

• Electronic Mapping, Reporting, and Coding (eMaRC) Plus
  – A cancer registry tool developed by the U.S. Centers for Disease Control and Prevention (CDC)

• Integration Platforms (in general)
TRS Constructor

- Similar in concept to DeDupIT
- Component of a larger TRS Architecture concept that queries HIE network in real-time for the individual CDA documents
- Includes an MPI to resolve identities
- Tested as an input to CDSS to suggest differential diagnosis based on symptoms

Electronic Mapping, Reporting, and Coding (eMaRC) Plus

• Software consumes CDA documents, parses contents, and stores data as individual reports from hospitals (AbsRefID Abstract ID)

• Multiple Record Management algorithms de-duplicate data after they have been entered as records into a Central Cancer Registry
  – Based on unique Patient/Tumor/Reporting Entity ID

• Created by the National Program of Cancer Registries (NPCR)

https://www.cdc.gov/cancer/npcr/tools/registryplus/mp.htm
eMaRC Plus

-Generically this approach could work in a commercial EHR system or within an HIE network that involves a centralized data repository.

-While reasoning over tumor record entries is challenging, most tumor records are entered by highly trained, professional cancer registrars.

  -This approach may not work for documents generated as a result of routine medical care.
Key Considerations (Summary)

• Information Retrieval: how will input CDAs be gathered by system?

• Patient Matching: ensure the input CDAs are for the same person

• Semantic Encoding: Both DeDupIT and TRS Consolidator identified challenges with existing approaches involving ontologies; the CDC solution uses a highly specialized, custom data dictionary

• Workflow: where in the clinical workflow will systems (or people) fetch outside documents and require consolidation
Status and Future Direction

• Dr. Hosseini successfully defended his dissertation and graduated!
  – He continues his work on standards and interoperability

• Dr. Dixon continues his work in researching solutions for HIE
  – More work is necessary to enhance the capacity for EHR systems and HIE networks to identify, consolidate, and present information to end users where and when needed
  – Open to collaboration with ANY company, student or others with good ideas on how to enhance interoperability and workflow
The “E”s: Efficient, Exchange, Easy to Use

Reduce Information Overload
By consolidating data in C-CDA documents
Acknowledgements

• Research Committee
  – Josette F. Jones, RN, PhD
  – Daniel J. Vreeman, PT, DPT, MSc
  – Anthony Faiola, PhD
  – Huanmei Wu, PhD

• The work presented was completed in support of Dr. Hosseini’s PhD in Health Informatics at the Indiana University School of Informatics and Computing (SOIC)
  – http://doi.org/10.7912/C20W3J
For More Information


Questions

• Brian E. Dixon, MPA, PhD, FHIMSS
  – Associate Professor, IU Fairbanks School of Public Health
  – Research Scientist, Regenstrief Institute
  – http://tinyurl.com/fsphbed
  – Twitter: @dpugrad01

• Masoud Hosseini, MS, PhD
  – School of Informatics and Computing, IUPUI
  – Hosseini@iupui.edu