Third Party Vendor Perspective on Harmonizing Data

Dody Fisher, MD
Chief Clinical Officer
Forward Health Group, Inc
"From chaos to cohesion to clinical trials to cures"
Conflict of Interest - this is where I work
What is the definition of data harmonization?
Do we focus on the process?

The process of comparing two or more data component definitions and identifying commonalities among them that warrant their being combined, or harmonized, into a single data component.
Do we focus on the end use?

Data harmonization is defined as a process whereby an organization reviews separate but similar data and attempts to align the data elements in a way that can be combined and used for comparison or stratification.
Do we focus on previous efforts?

Data harmonization interprets existing characteristics of data and action taken on data and uses that information to transform or suggest subsequent data quality improvements.
Data harmonization refers to all efforts to combine data from different sources and provide users with a comparable view of data from different studies.
Making data from different sources comparable and usable.
EHR

Source of truth?
Colorectal Cancer Screening
A Case Study Within the EHR
Preventive Health Form
<table>
<thead>
<tr>
<th>Test Name</th>
<th>Results</th>
<th>Date</th>
<th>Previous Value</th>
<th>Full View</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPD</td>
<td></td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td>Hemoccult</td>
<td></td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td>PSA</td>
<td>negative</td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td></td>
<td>positive</td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td></td>
<td>negative x 3</td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td>Colonoscopy</td>
<td></td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td>Last Tetanus Booster</td>
<td></td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td>Last Flu Shot</td>
<td></td>
<td></td>
<td></td>
<td>FV</td>
</tr>
<tr>
<td>Last Pneumovax</td>
<td></td>
<td></td>
<td></td>
<td>FV</td>
</tr>
</tbody>
</table>
Enter Colonoscopy History...
Order a Test

The image shows a software interface for ordering medical tests. The interface includes options for selecting a specialty, entering order details, and selecting diagnosis-specific orders. The screen displays fields for entering test or procedure names and reasons for deferment.
...or GI History
...or Patient History

- Colonoscopy a few years ago

- Last update: reviewed - no changes required
Data is messy!!!

And harmonizing multiple sources is even messier!!
53215 File Summary

**WMS**
- 522 Physicians
- 522 Physician Name permutations
- Range -
  - 1 MD : 1 Name

**“BBD”**
- 1491 Physicians
- 3394 Physician Name permutations
- Range -
  - 1 MD : 1 Name
  - 1 MD : 18 Names
“BBD” - Medical Licenses

- 1491 Physicians
- 1095 Medical Licenses
- State(s):
  - WI  632
  - IL  108
  - CA  56
  - FL  36
  - MI  33
  - AZ, OH  24
  - TX  20
  - MN  16
  - MO, NY  13
  - VA, MD  11
  - MA  10
  - NE  8
  - CO  7
  - TN, IA  6
  - ID, NC, LA, OR, PA  5
  - UT, VT, KY, NV  4
  - GA, SC, NJ, WV  3
  - OK, AL  2
  - CT, RI, AK, DC  1
WMS Unique – 112 MDs

- 112 of 112 Verified as Physicians in Active Practice
- 112 of 112 Verified as 53215
- 112 of 112 Valid Wisconsin Medical License
4 1 10.89
4 1 10.90
2 1 10.91
4 1 10.92
2 1 10.94
1 1 10.95
1 1 10.96
1 1 10.97
1 1 10.98
1 1 10.99
1 1 100.93
1 1 102.65
1 1 102.87
1 1 102558.90
1 1 103.04
1 1 103.39
1 1 103.43
1 1 103.54
1 1 103.55

Something fishy with BMIs.
## Odd Ejection Fractions...

<table>
<thead>
<tr>
<th>1</th>
<th>31%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td>11</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>35-40</td>
</tr>
<tr>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>1</td>
<td>39%-45%</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>40% %</td>
</tr>
<tr>
<td>1</td>
<td>40-45</td>
</tr>
<tr>
<td>1</td>
<td>40-45 %</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>26</td>
<td>45%</td>
</tr>
<tr>
<td>1</td>
<td>45-50</td>
</tr>
<tr>
<td>9</td>
<td>45-55%</td>
</tr>
<tr>
<td>1</td>
<td>46</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>1</td>
<td>47%</td>
</tr>
<tr>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td>1</td>
<td>48%</td>
</tr>
<tr>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>23</td>
<td>50%</td>
</tr>
<tr>
<td>1</td>
<td>50-55</td>
</tr>
<tr>
<td>3</td>
<td>50-55%</td>
</tr>
<tr>
<td>1</td>
<td>54</td>
</tr>
</tbody>
</table>
Date of birth issues

1960-01-01 vs 1960-01-02 (fat fingered)
Trailing spaces - causes issues with mapping - untrimmed values “V70.02 " vs “V70.02”
Codes not having any punctuation

B218 vs B21.8
B18 vs B18.0
New line characters within fields
Headers not lining up:

9 headers, 10 columns
Why does this happen?
Humans
Errors, judgement, process issues
What can be done about this?
Data harmonization!
Machine learning
Eyes and Hands
Team effort
### Population Snapshot

**View:** All Clinics, All Teams, Overall

**Population:** Colon Cancer Screening, 37,058 Patients

**Groups:** All

#### Current Results - Q2 2016

- **Colon cancer screen**
  - Colonoscopy performed every 10 years: 30958, 5730, 84%
  - Colonoscopy due date recorded: 14548, 22140, 40%

**Data current through:** 2016-06-30

**Software Version:** 4.2.1

**Release Notes**

**FHG Internal Release Notes**

**Last Code Pull:** November 07, 2016
Thank you!!