P1752 Metadata Subgroup Group Meeting

Sponsored by IEEE Engineering in Medicine & Biology (EMB) Standards Committee

1 October 2019
Teleconference
Members/Attendance

- Subgroup chair: Ida Sim, Open mHealth / UCSF
- Subgroup secretary: Anand Nandugudi, U Memphis
- Call out your name in the following order if you’re here (so we can get familiar with your voice)
  - Pradeep Balachandran
  - Jakob Bardram
  - Daniela Brunner
  - Simona Carini
  - Paul Harris
  - Shivayogi Hiremath
  - Sean McConnell
  - Leonard Njeru Njiru
  - Henry Ogoe
  - Paul Petronelli
  - Udi Rubin
  - Anna T
  - Vishnu Ravi
Metadata Ecosystem
Action Items
From Last Meeting
Action Items from September 17

• Investigate further into DatapointID
• Investigate further into missing data in datapoint series
• Use AMA Blood Pressure use case as driving example after metadata minimum stabilized
Datapoint vs. Datapoint series
Datapoint Definition

• “A datapoint is a discrete observation or measurement on a single member of the unit of observation”
  • Datapoints are not about populations, but about individual units of observation, which can be a person, a run, a meal, etc
  • Discrete is discrete in time, value, and origin/source
  • An observation or measurement can be a calculated output
  • Datapoints are not restricted to being a single number. A datapoint can be e.g., an image, a vector.
Datapoint ID

• What are requirements?
• How are IDs assigned so that they can be used as expected and what are the expectations?

• Action item? -- Sean
  • Namespace? Uniqueness
  • Examples? HL7 FHIR resources.
  • How is this used?
  • Is this needed?
Datapoint versus Datapoint series

• Schema can be used for instances of arrays of observations (i.e. a series) not only a single datapoint

• Acquisition provenance often applies to a datapoint series not a data point (e.g., sampling rate)

• Metadata must be identical for every data point in the series.
Missingness in Datapoints and Datapoint Series

• Types of missingness, e.g.,
  • Device not worn
  • System outage, out of battery, etc
  • Obfuscated for privacy

• How to denote
  • sequence number (where to keep this?)
  • NULL values

• Periodic vs non-periodic data series

• Does this need to be in our minimum metadata standards? -- Anand
Missingness

• Source of missingness
  • Device failure (battery, software/hardware failure)
  • Break in the communication link
  • Privacy enabled for a time period.
  • Others?

• What needs to be captured
  • Missingness of data because enabling the privacy option is maybe the only thing we know for certain

• Who records missingness of data
  • The process that receives data from the sensor
  • How to capture failure of the process that receives the data
  • Can the receiver process record poor data quality also?
Missingness

• Other ways of inferring missing data
  • Algorithm that uses the sensor data could infer missing data or poor data quality.

• Example – currently the cStress algorithm can infer
  • No data
  • Insufficient data
  • Unusable data (poor data quality)
    • Motion above a threshold
    • Sensor not worn
Missingness

• How to represent missing data
  • Can a missing datapoint be represented with a datapoint (using the same schema that is used for the sensor values) with a value that corresponds to ”missing data”

• Capturing other information
  • Is it useful to capture whether a datapoint series’ sampling rate was regular or irregular?
  • Privacy
    • Should the sensor value be stored even if privacy is enabled.
  • Others?
Minimum Metadata

Use Case Example: Complex
Complex Example – Stress estimation from PPG

1. MotionSense HRV transmits PPG sensor data (IR, green, red channels) on Bluetooth
2. mCerebrum app on Android receives and stores sensor data
3. The sensor data is uploaded to cloud (CerebralCortex)
4. Algorithm on the cloud process sensor data
   1. Filter out unusable sensor data (movement from accelerometry and gyro, ‘bad’ data)
   2. Compute peak intervals from PPG sensor data
   3. Compute features from peak intervals
   4. Estimate and output stress level from features.
Stress Datapoint

{
    "stress_value": 0.75,
    "effective_time_frame": {
        "time_interval": {
            "start_date_time": "2019-08-01T07:00:00Z",
            "end_date_time": "2019-08-01T07:01:00Z"
        }
    }
}
# Datapoint: Complex Example

<table>
<thead>
<tr>
<th>Needs</th>
<th>Property (bold = required)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Which datapoint is this?</td>
<td>datapointID</td>
<td>See above slides</td>
</tr>
<tr>
<td>What does this value represent?</td>
<td>schema ID and schema metadata</td>
<td>Pointer to the stress datapoint schema</td>
</tr>
<tr>
<td>When was this datapoint created?</td>
<td></td>
<td>Moved to Acquisition</td>
</tr>
<tr>
<td>When is the effective time of this data?</td>
<td>[in the datapoint itself]</td>
<td></td>
</tr>
</tbody>
</table>
## Acquisition: Complex Example

<table>
<thead>
<tr>
<th>Needs</th>
<th>Properties (bold = required)</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>When was this datapoint created?</td>
<td>(source_)creation_date_time • When captured • When computed • When transmitted • When received/stored</td>
<td>Paul Petronelli: My proposal is that we consider that ‘data’ whether it is sensor or time stamp, has a ‘lifecycle’. For example State 1 is ‘measurement’ or ‘calculation’, State 2 is Transmission and State 3 would be recording or storing. The ‘state’ could be a metadata element. Will work more on this if the group finds it useful.</td>
</tr>
<tr>
<td>Was the datapoint sensed or self-reported?</td>
<td>modality</td>
<td>sensed</td>
</tr>
<tr>
<td>What was the periodic sampling rate, if any? (for datapoint series) – useful??</td>
<td>sampling rate</td>
<td>100Hz</td>
</tr>
<tr>
<td>Was the sampling periodic?</td>
<td>Periodic, Not-periodic</td>
<td>Periodic</td>
</tr>
</tbody>
</table>
Source: Complex Example

<table>
<thead>
<tr>
<th>Needs</th>
<th>Properties (bold = required)</th>
<th>Pointer system; access authentication, storage, maintenance, verification of datasheets; runtime access – all are features of the metadata ecosystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>What firmware/algorithm? What hardware? What app/product? Which person? Which study?</td>
<td><strong>Pointer(s) to</strong> <em>Software Datasheet, Hardware Datasheet (UDI), Product Datasheet, Personal Datasheet (User ID), Study Datasheet (Study ID)</em></td>
<td></td>
</tr>
</tbody>
</table>


Future Work
Outstanding Items

• Datapoint ID
• Missing data representation
• Data lifecycle
• AMA Blood Pressure use case to test
Future Meetings
Upcoming Meetings

- Metadata WG
  - October 15: 9:00 – 10:00 AM Pacific
Adjournment