IEEE STANDARDS ASSOCIATION

IEEE

P1752 Metadata Subgroup Group Meeting

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

- 27 August 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

Action Items
From Last Meeting

Action Items from July 9

- Define datapoint and datapoint series (was: data stream)
- Anna T: Work on a simple example, for e.g. Physical activity
- Anand: Work on a complex example, Stress calculation from PPG Sensor

Datapoint vs. Datapoint series

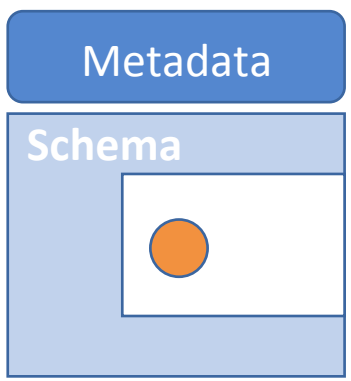
Datapoint Definition

- “An identifiable element in a data set” (Lexico.com / Oxford)
- “Any single, unique piece of information” (Computer Hope)
- “A discrete unit of information” (WhatIs)
- “A data point or observation is a set of one or more measurements on a single member of unit of observation.” (Wikipedia)
- Not in IEEE Computer or Tech Terms dictionary

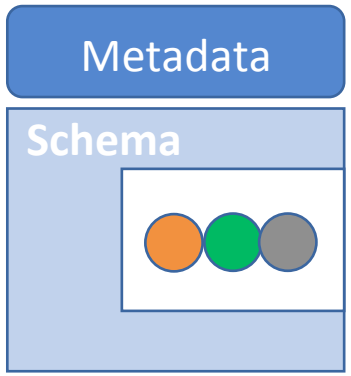
- For discussion: “A datapoint is a single discrete observation/measurement on a single member of the unit of observation”
  - Datapoints are not about populations, but about individual units of observation
  - Unit of Observation can be a person, or a run? a meal?

Datapoint versus Datapoint series

- Acquisition provenance often applies to a datapoint *series* not a data *point* (e.g., sampling rate)
- Schema can be used for instances of arrays of observations (i.e. a series) not only a single datapoint
- But metadata must be identical for every data point in the series.



Datapoint



Datapoint series

Sequence number?

Modeling example (I)

```
{
   "$schema": "http://json-schema.org/draft-04/schema#",

   "description": "This schema represents an array of numerical values with a unit of measure.",

   "type": "object",

   "properties": {
      "values": {
         "type": "array",
         "items": {
            "description": "The numeric value of the element.",
            "type": "number"
         }
      },
      "unit": {
         "references": [
            {
               "description": "The unit of measure of the element. Allowed values are drawn from the Common synonyms (non-UCUM) column of
[subset of] UCUM, SI and English units. ",
               "url": "http://download.hl7.de/documents/ucum/ucumdata.html"
            }
         ],
         "type": "string"
      }
   },
   "required": [
      "values",
      "unit"
   ]
}
```

Modeling example (II)

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "references": [
    {
      "description": "The NCIT code represents the time measurement between the R wave of successive heartbeats as measured in milliseconds.",
      "url": "https://ncim.nci.nih.gov/ncimbrowser/ConceptReport.jsp?dictionary=NCI%20MetaThesaurus&code=C0489636"
    }
  ],

  "definitions": {
    "unit_value_series": {
      "$ref": "unit-value-series-1.0.json"
    },
    "time_frame": {
      "$ref": "time-frame-1.x.json"
    }
  },

  "properties": {
    "rr_interval_series": {
      "allOf": [
        {
          "$ref": "#/definitions/unit_value_series"
        },
        {
          "properties": {
            "unit": {
              "enum": [
                "ms"
              ]
            }
          }
        }
      ]
    },
    "effective_time_frame": {
      "$ref": "#/definitions/time_frame"
    }
  },
  "required": ["rr_interval_series","effective_time_frame"]
}
```

Modeling example (III)


```
{
  "rr_interval_series": {
    "values": [
      0.895,
      0.896,
      0.885,
      0.896,
      0.885
    ],
    "unit": "ms"
  },
  "effective_time_frame": {
    "time_interval": {
      "start_date_time": "2013-02-05T07:20:00Z",
      "end_date_time": "2013-02-06T07:25:00Z"
    }
  }
}
```


Assuming metadata includes information that sampling interval is 1 minute
Question: how to represent missing data?

Metadata Landscape

Metadata Landscape

For research studies

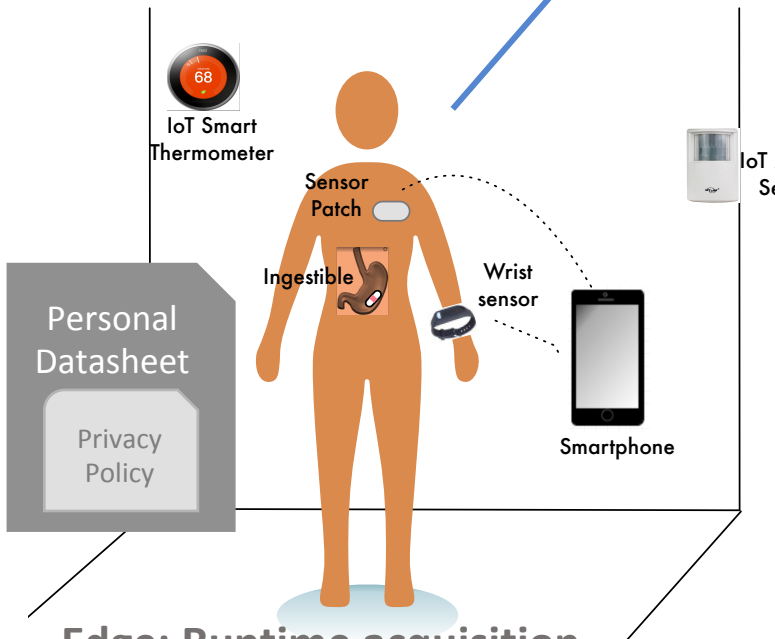
Study ID
Study Protocol
Data Sharing Policy

Software Datasheet
Data Data sheet

Data stream(s) → Firmware/ algorithm/model → { Header/min metadata } { Body: datapoint }

Cloud: Runtime acquisition and privacy metadata

Open mHealth
data and metadata standards



Personal Datasheet
Privacy Policy

Hardware Datasheet
UDI

Edge: Runtime acquisition and privacy metadata

EHR

Solutions Datasheet
Privacy Policy

Minimum Metadata

Use Case Example: Simple

Step count datapoint body – Regular

Example: 2 samples/hour

```
{
   "activity_name": "walking",
   "effective_time_frame": {
      "time_interval": {
         "start_date_time": "2018-12-27T06:25:00Z",
         "end_date_time": "2018-12-27T10:45:00Z"
      }
   },

  "distance": {

      "value": 2, 2, 1, 0.5, 2,  0.5, 0.5, 0.2, ← for discussion only; data does not comply with schema; must be single value

      "unit": "km"
   },
   "base_movement_quantity": {
      "value": 2476, 2476, 1238, 619, 2476, 619, 619, 250, ← as above, for discussion only
      "unit": "steps"
   "cadence": {
      "value": 82.5, 82.5, 41.2, 20.6, 82.5, 20.6, 20,6, 8.3, , ← as above, for discussion only
      "unit": "steps/min"
   }
}
```

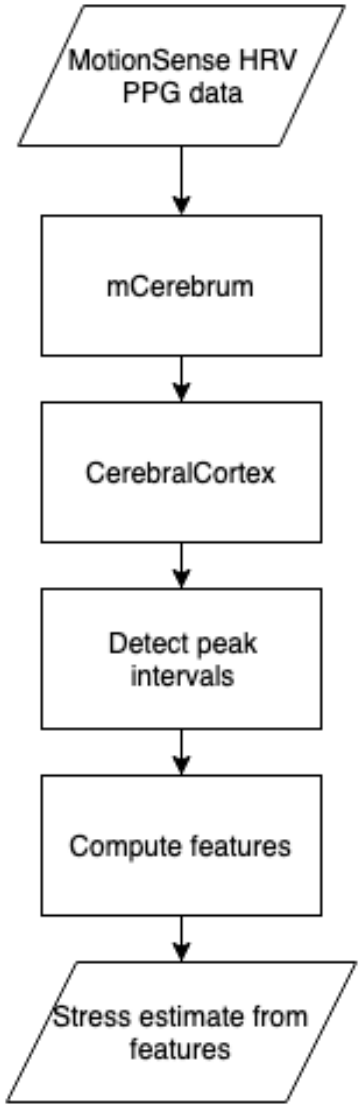
Step count datapoint body - Irregular

```
{
   "activity_name": "walking",
   "effective_time_frame": {
      "time_interval": {
         "start_date_time": "2018-12-27T06:25:00Z", "2018-12-27T08:35:00Z", ← for discussion only; data does not comply with schema; must be single value
         "end_date_time": "2018-12-27T08:25:00Z", "2018-12-27T10:45:00Z", ← as above, for discussion only
      }
   },

  "distance": {

      "value": 6.7, 2,

      "unit": "km"
   },
   "base_movement_quantity": {
      "value": 10000, 773, ← as above, for discussion only
      "unit": "steps"
   "cadence": {
      "value": 83, 6, ← as above, for discussion only
      "unit": "steps/min"
   }
}
```

Use Case Example: Complex

Complex Example – Stress estimation from PPG

1. MotionSense HRV transmits PPG sensor data on Bluetooth
2. mCerebrum Android app receives and stores sensor data
3. The sensor data is uploaded to cloud (CerebralCortex)
4. Algorithm on the cloud processes the sensor data
    1. Filter out unusable sensor data (movement, ‘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

Needs	Property (bold = required)	
Which datapoint is this?	**datapointID**	**The algorithm that computes this datapoint should assign datapointID. How to generate unique datapointID?**
What does this value represent?	**schema ID and schema metadata**	**Pointer to the stress datapoint schema**
When was this datapoint created?	**creation_date_time**	**The time when the datapoint was computed in the cloud?**
When is the effective time of this data?	[in the datapoint itself]	

Acquisition: Complex Example

Needs	Properties (bold = required)	
When was this datapoint created at the source?	**source_creation_datetime** date-time schema represents a point in time (ISO8601). Timezone is UTC unless otherwise specified	Not sure about this field
Was the datapoint sensed or self-reported?	**modality**	**sensed**
How often was data sampled and was the sampling regular?	**sampling rate** and **regular or not** (Boolean) / *sequence number*	*Should this be split into two fields?* *Sampling rate – 100Hz* *Regular sampling – not sure what this means.*
Type of filtering, if used	e.g., values averaged	Pointer to the algorithm? For stress, we filter out data where periods when vigorous movement is detected or when the sensor is not worn.

Source: Complex Example

Needs	Properties (bold = required)	
What firmware/algorithm?	**Firmware name, firmware version**, *Software Datasheet*	Is this the firmare of MotionSense HRV and/or algorithm/sw versions of the cloud and mobile app
What device/app?	**name, manufacturer/publisher, model**, *Solutions Datasheet*	**Mobile app version and/or phone model**
What OS platform?	**OS version** (e.g., iOS, Android), *Hardware Datasheet*	iOS / Android OS version
Which individual device/app?	ID, ID Type (e.g., UDI)	Is this the UDI of MSHRV/Phone ?
Which person?	**User ID,** confidence, *Personal Datasheet*	We use uuids in our studies for each user.
Which study?	**Study ID??** Ref Study protocol	Pointer to the study ID.

Future Meetings

Upcoming Meetings

- Metadata WG
  - September 17: **9:30 – 10:30** AM Pacific

Adjournment