**Title of session**: Common Aspects of Data Description

**Day**: Tuesday (2b)

**Participants**: Klas Bloomqvist, Flavio Rizzolo, Eric Prud’hommeaux, Barrie Nelson, Gregg Kellogg, Larry Hoyle, Steve McEachern, Deirdre Lungley, Wendy Thomas, Achim Wackerow

**Chair**: Achim Wackerow

**Note taker**: Wendy Thomas

**Background information**

Where are the common aspects in terms of data and the metadata between standards? What level do we wish to map to? It is easiest to look at the logical model first

*[Links to any relevant documents*]

[High-Level Specification/ Data Type Matrix](https://ddi-alliance.atlassian.net/wiki/download/attachments/31260690/SpecificationFeatureMatrix.docx?api=v2)

Piratepad.net/ddib

**Decisions**

The initial terms were from the original document High-Level Specification/ Data Type Matrix. The highlighted term seems to be the clearest consensus.

* Microdata – Unit level, observation, single attribute, has a relationship with the unit but no structural relationship with other attributes of the unit, **unit data**
* Aggregate data – matrix relationship of the attributes of unit; **dimensional data**, hyper-cube, data may have a dimensional description commonly the result of some tabulation which is generally understood to result for aggregation
* **Event data** – The unit is the event as opposed to the object. Capture is triggered by an event and associated to an object. In a survey/census/non-trigger capture the data collection is triggered by a choice and the collection event is common to the group of units. An event which triggers data capture generally relates to one or a subset of the objects (one : birth, subset : weather event). Either can be periodic in nature. Events and interventions are similar. Time is a critical attribute of the data. Findings are a data set instantiation of a findings class of the attributes of an event or intervention types. Events, Interventions, and Findings are all analyzed separately. An event can be characterized as a state transition in a sequence of observations. A single finding (a unit record) is an observation which may have multiple variables that describe it and may be associated with events and/or interventions. For each you record time and date and the object identifier.

The differentiation of microdata, aggregate data, and event history data is not the way to approach this. Basically, we collect things about units and roll it up in different ways. Therefore, we should be looking at this as units, variables, records, and data point. The differentiation between data types is what the unit is and how data points relate to each other.

* Unit – The person, place, thing, or event that variables refer to
* Variable – An attribute related to the unit (semantic description)
* Record – A set of attributes related to the unit
* Data Point – The specific value/content of a specific variable related to a specific unit

**Issues requiring further discussion**

*[Please note pros and cons for different positions and the reasoning for them]*

**Is there a need for a continuation of this discussion?**

*[Possible options are further session, plenary, or evening session]*

**Is there a need for a longer document to continue this discussion?**

* Barry and Eric write down how Variable, Record, and DataPoint are expressed in CDISC and FHIR (see piratepad.net/ddib)
* Need to map in terms of terminology