

## Key documents:

- Spreadsheet of terms: <https://docs.google.com/spreadsheet/ccc?key=0Aoy0zfdRviKsdFJWTDFpbINXc3BtelhrdEpNYTdvbXc#gid=1>
- Data model diagrams:
  - Abstract: <https://docs.google.com/drawings/d/1e6gsxPkc-gKecVTJGJePE1Nuy2sD8Puu-FsEtUtGC-o/edit?disco=AAAAAFWHcnw>
  - ChEMBL: <https://docs.google.com/file/d/0B4y0zfdRviKsS1I2NEttN3pfc1k/edit>

## Dataset statistics

- Bio2RDF have a very comprehensive set of pages giving the statistics of each of their datasets  
<http://download.bio2rdf.org/release/2/release.html>
- Statistics are very RDF oriented. Question of where we attach them to: RDF versioned so that it is available to all serialisations but not to a database serialisation.
- Looking at the vocabularies to represent these rather than require these to be provided. Provide recipes (SPARQL queries) for creating these statistics.
- Statistics are currently on a per graph basis. Nothing stopping you computing statistics across multiple graphs. Added number and list of named graphs property. Added a property for capturing graph name pattern.

Dataset Indicators: require tracking over time, e.g. availability

- Dataset statistics are a subset of the size property: The size property is about any format whereas the statistics are really only for an RDF serialisation. Of course, other statistics could be captured about relational databases.
- Rate of change would need to specify the unit of time that it relates to.
- Extent of use could be provided by the users themselves rather than the original data providers
- Language coverage: per language what percentage of literals are available.

Next meeting: start getting agreement for the language constructs used for each property  
Next two calls will be on Tuesdays instead of the Monday slot.