Michel and Alasdair have model proposal for a dataset version and formatting: thrashing out the final details and will present as the first item at the next call.

Focus on provenance and origin:

- Version: literal
- Dates: Full date time with timezone
- Modified by: URI for a person
- Source: Needs to be present and specify the version/date of the source
- Prior version: point to URI
- Superseding version: Maintenance issue; can be inferred; provenance vocabularies only point backwards
- Subset/superset: only assert that you are a subset of another one
- Frequency of change: estimate; use a URI for the value, e.g. dublin core
- Latency of change: time it takes for changes in the raw data to appear in the derived dataset. Very specialised and probably not to be included
- created with: realised we need to point to a tool that was used to generate the dataset; particularly for D2R or Bio2RDF point to versions of scripts

Aggregation of datasets is covered by source/derivation: simply include multiple sources

Availability

- Availability: raises a maintenance issue; could capture available until X if it is known that it is no longer going to be available. Good for registry use case, but not necessarily for data publishing. It is a monitoring property.
- Publisher: need to decide a value set: literal/URI
- Format: mime type of the file, not the vocabularies used; EDAM, biosharing as candidates
- data item HTML template: to automate access
- RDF dump: available in multiple formats
- SPARQL endpoint
- API: point to a top level page about the API rather than each individual method
- Catlog/registry: point to records in registries. Inverse relationship with the registries