Registering (mainly data) services

- Want to access smaller data collections too
  - part-time/spare-time archivists
- Many early Registry entries are incomplete
  - coverage undefined/incomplete
  - data description missing
  - sometimes no date, no version, etc. etc.
- Got away with it so far
  - test queries built using prior knowledge
- Why is it too hard even for 'us' to make proper entries?
  - no low-effort interface
  - data providers give too much or too little detail
  - coverage takes effort to compute
  - we forget users/tools lack prior knowledge
  - element names not unique
  - not documented for astronomers etc. etc.

Prioritise

- Define elements as *must, should, may* etc.
  - *must* implies Registry can't describe resource if element is missing
- Essential *must's* (defined as Required in RMv1.0)
  - Title, Identifier
  - Publisher
    - could be VO if otherwise unknown
  - Subject, Description
    - highly desirable but maybe not essential
  - Reference URL
  - Type (controlled list - Archive, Registry etc.)
- Add Date and/or Version to Required
  - suggest Date, default date Registry entry made
- ServiceMax(etc.) - user must know any restrictions?
- Element names unique to aid human checking
- Flag for accuracy/completeness of metadata?
  - Bob's proposed incremetal quality indicator!

Need tools to support registration

- Web forms - keep short!
  - Scripts to help data providers (a là Sébastien's talk?)
  - Validate entries (against schemata - and data?)
    - helpful error messages
  - **Identity, Curation** not too bad
  - Type determines which elements are needed?
  - Subject, Description
    - *metadata harvesting* from data resource?
    - from keywords and abstract if Source=bibcode
- Relationship, RelationshipID
  - Needs careful definition wrt use of extensions
  - Thereafter could use to copy or link common elements
    - e.g. observing log and source list share Facility etc.
- Coverage
  - If not filled, extract from data or keywords
  - Initially simple, probably over-inclusive
    - e.g. 'radio' data given $\lambda \leq 10$ mm if not specified

Spatial coverage

- Crude approximation 360° by 180°
- For some data best description is in Galactic coordinates even if CFRAME is Equatorial
  - could default e.g. to FK5 with warning if CFRAME undef
- Eventually refine coverage calculations
  - sky indexing?
  - select small number of regions with different filling factors?
- Similar (simpler?) process for spectral, temporal