Notes from ACRID meeting, 14ᵗʰ December 2010, RAL

Attending: Jeremy Tandy (Met Office), Tim Osborn (UEA), Colin Harpham (UEA), Andrew Woolf (STFC), Arif Shaon (STFC), Sarah Callaghan (STFC), Bryan Lawrence (STFC), Ag Stephens (STFC)

Next meeting: to be confirmed, but will be a teleconference in mid-late Jan/early Feb.

Actions:

- Finalised project documents (project plan, work package structure, budget and signed consortium agreement) need to be sent to JISC – **Sarah**
- Set up a doodle poll for the best date for the next project meeting/teleconference –**Sarah (done)**
- Set up a mailing list for the group - **Sarah**

Issues:

- The analysis of the workflow to be done by UEA needs to be in a language that everyone understands. Possible methods for describing the workflow could be through user stories/cases, plain text descriptions or sequence diagrams.
- We should be aware of what other people have done with regards to software for rendering linked data in different forms.
- This project is effectively a prototype – we should be careful not to over-commit, and should define the project scope carefully.
- CRUTEM and the tree rings data set are considered higher priority than CRU TS. The BADC has CRU TS in the archives and could expose it as part of the STFC contribution.
- Linked data graphs and versioning are an issue that needs to be addressed.
- Met Office issues (see also Jeremy's notes from the same meeting):
    - Version control on the operational version of CRUTEM
    - How can we make sure ideas produced by ACRID are cost-effective in operational cycles?

Decisions:

- Sarah Callaghan will be taking over as project manager from Andrew Woolf, effective immediately. Andrew is due to leave STFC in January 2011 to move to the Australian Bureau of Meteorology.
- The project is officially starting now. The end date is still set for July 2011, though Simon Hodson at JISC has indicated that a no-cost extension might be possible. Given that Arif will be able to work 100% on this project from January, the extension might not be needed. Colin will be working on the project from 1 December until the end of May.
- The work done on GeoTOD forms part of the STFC contribution in kind, and therefore should be part of the project plan and fully reported on in the documentation.
- The main JISC programme meeting will be held on the 28ᵗʰ and 29ᵗʰ March, 2011. Sarah will be attending this to represent the project.
- The surface databank project could use the same methods as ACRID.
- We will do CRUTEM first, and then can use CRU TS as a test for the process. We should therefore timetable some time in for CRU TS.
- The ACIRD processes will be lightweight, fitting in with the research environment of UEA.

- There is no need for Met Office to validate implementation costs for the processes developed to support UEA - our interest is purely in understanding cost implications of the proposal when transferred into the operational environment at Met Office (for example, if Met Office see value in re-publishing a new linked-data graph with each monthly update to CRUTEM. The critical 'validation question' posed is "will the proposal help merge development and operational branches of CRUTEM to create a new CRUTEM4 baseline.
- Met Office will also use HadCET to review the process as fit-for-purpose to support a wider genre of climate data management - such as the Surface Temperatures Databank. Again, a formal cost assessment is unlikely as a result of this activity.
- The following is a rough roadmap of the next project steps:
    - Description of processes
    - Data modelling
    - Choice of workflow and granularity

Information points:
- UML can be used to autogenerate rdf (and the BADC has experience with this).
- The CRUTEM dataset is in two versions, the operational version and the research/development version. The research version is only used by researchers at UEA to inform the development of the next operational version. Everyone else uses the operational version.
- TRiDaS is an xml schema for tree ring data.
- CRU TS is updated monthly, yearly and decadally.
- MOLES can conveniently package thousands of observations in a html page, but we can also page it.
- Virtual datasets (e.g. all the places that use a particular instrument in the past 10 years) can be generated by query and can be usefully encapsulated.
- DOIs can only be issued to datasets which are "frozen" i.e. not going to change.

First version, Mon 20th December 2010
Second version, Tues 21st December 2010 (Amended after a clarification of one of the points)