JISC

Progress Report Template -

Text in italics is explanatory and should be deleted in completed documents.

Project Name	Advanced Climate Research Infrastructure for Data (ACRID)- *University of East Anglia*
Project Website	http://www.cru.uea.ac.uk/cru/projects/acrid/
Report compiled by	*Sarah Callaghan*
Reporting period	*August 2010 – February 2011*

Section One: Summary

Please provide a short overview (1-2 paragraphs) of project progress during this reporting period, which could be disseminated to programme stakeholders.

The ACRID project got off to a delayed start, mainly due to problems with the availability of key staff. Work started in earnest in December 2010 and has been progressing well.

Two deliverables, D2.1 (Description of scientific workflows) and D2.2 (Information architecture) have been produced.

Section Two: Activities and Progress

Report on activities as outlined in your work packages for the period covered by this report and describe any changes to this, including the reasons for these. Do include any additional activities undertaken that are not in your work packages, providing the background to their inclusion.

WORKPACKAGE 2: Requirements analysis

2.1 : *Analyse workflows*

Initially, detailed text explanations were produced of the workflow processes involved in the construction of two climate datasets, (i) the CRUTEM weather station monthly temperature database and the gridded and global temperature records constructed from them; and (ii) tree-ring measurements and the chronology of regional tree-growth constructed from them. From these, more formal descriptions of the workflows have been developed which identify the data, metadata and processes/software associated with the workflow in each case, including identifying the components that need to be captured to provide a full record of these datasets. Preliminary work on a third CRU dataset, the CRU TS high-resolution climatological dataset encompassing multiple climate variables, was also begun. Overall, this work forms the basis for deliverable D2.1.

2.2 *Develop information architecture*

  a. An information model has been developed in UML to generalise the concepts identified by the analysis of the scientific workflows associated with the prototype climate research datasets (WP 2.1). In essence, this information model specialises the ISO 19156 Observation and Measurement model to enable detailed and structured description of the observations represented by the prototype climate research datasets. In addition, ISO O&M specialization would enable the model to be interoperable with other ISO O&M based information models and thereby enabling the CRU datasets to be shared with a wider Geospatial community, potentially through a global Spatial Data Infrastructure, such as the INSPIRE SDI. An RDF ontology representation of the model (using a suitable UML to RDF

conversion mechanism) will be produced in the next phase of the project in order to facilitate publishing the scientific workflows according to the linked-data principles.

b. A data management infrastructure is being developed to capture various aspects of managing the CRU datasets. In particular, this infrastructure will help incorporate the capturing of the metadata represented by the aforementioned information model within the current CRU data management architecture in an efficient manner. It should be noted that the completion of this item, which was originally scheduled for the current phase of the project, has been delayed due to the deferred start of the project. The development of this data management infrastructure is expected to be completed in the next phase of the project.

c. A deployment architecture indicating the software components that need to be developed to enable capturing the workflow related metadata (represented by the information model) has not been completed due to the delayed start of the project. This item is expected to be completed in the next phase of the project.

WORKPACKAGE 3: Research data management

3.1 Software management

A version control system has been implemented using 'Subversion' software and a repository provided to cater for the three CRU datasets that are ACRID case studies. The CRU observed station temperature archive has been traditionally maintained in a single file, which is updated as part of the development of the next major version (CRUTEM4). However, the operational version (currently CRUTEM3) which is maintained by the Met Office Hadley Centre (MOHC) uses a different storage structure and file format. To provide consistency with MOHC, CRU now follow their structure and format of individual station files. Software tools have been developed to decompose the large CRU file to MOHC format individual files and vice-versa. The individual file format is now considered to be the master copy for the CRU development version and is put under version control; the data are updated by checking out these files and then assembling them into the original CRU format for compatibility with existing software/scripts used for updating and analysis. The reverse process is used for committing an update to the repository. This common file format will facilitate improved interaction with the MOHC. All other key CRUTEM files, including software, are in the process of undergoing version control, as are the data/software related to the tree-ring and CRU TS datasets. An introduction to the 'Subversion' software and to version control has been provided to a number of CRU staff not directly involved in ACRID, with the aim of using the ACRID project to stimulate improved data/software management practice more generally within CRU.

WORKPACKAGE 4: Data linking, citing, integrating

4.1 RDF Vocabularies

a. A comprehensive review of the existing data citation technologies and specifications, such as the DOI and OAI-ORE specification has been conducted to help determine an efficient approach to publishing the CRU datasets in linked-data format. Based on the outcome of this review, it has been decided to adopt an approach that involves encapsulating an instance of the information model produced by **WP 2.2a** within an instance of the OAI-ORE specification that is accessible through a DOI.

4.2 Linked-data Server

a. A linked-data server has been developed as part of another project called Geospatial transformation with OGSA-DAI (GeoTOD), which will be configured to serve up linked-data representations of the CRU datasets. This is an in-kind contribution to ACRID by STFC. Notably, the GeoTOD linked-data server is a full implementation of the guidelines provided in the paper entitled "Designing URI sets for Location" by the UK Cabinet Office, a part of the current UK government's "data transparency agenda" to make the public sector data freely accessible to the general public.

Section Three: Institutional & Project Partner Issues

Report on any changes or issues in your institutional context and/or any progress or issues with your project partners (where applicable).

The original project manager, Dr Andrew Woolf has left STFC, and Dr Sarah Callaghan has joined the project as his replacement.

Section Four: Outputs and Deliverables

Are there any outputs or deliverables you would like to share (e.g. presentations, studies)? Please describe, provide URLs or attach documents etc.

Two deliverables, D2.1 Description of scientific workflows and D2.2 Information architecture are submitted at the same time as this report, and are attached to this document.

Section Five: Outcomes and Lessons Learned

Outline any emerging outcomes or lessons, if any that have been learned during this reporting period that could be passed on to other projects or JISC.

Work is progressing well, given the late start of the project. The delay in the project start and the loss of the original project manager provide a lesson in the need to be flexible and produce contingency plans. However the hand-over to the new project manager has gone smoothly.

Section Six: Evaluation

Provide brief details of progress in terms of the development and implementation of the project evaluation plan. Detail any interesting findings or emerging evaluation issues of interest.

Project outputs are being produced as planned, given the late start to the project, and stakeholders have been engaged. Jeremy Tandy (Met Office) is a regular participant in the project meetings and teleconferences.

Section Seven: Dissemination

Report on any communication or dissemination activities with project stakeholders or the wider community which have taken place during the reporting period. Attach or provide URLs for any relevant dissemination or presentation materials. Include details of any publicity the project received during the reporting period.

Information about the project was presented at the NERC SIS Data Citation and Publication project workshop, held at BGS Keyworth on the 11th January 2011.

The project will be presented at the JISC MRD programme workshop, to be held in late March 2011.

Section Eight: Risks, Issues and Challenges

Report on any issues or problems that have impacted on the development and implementation of the project during the reporting period. Detail what impact any issues may have on the achievement of project targets, and set out how you plan to tackle these issues. Report on any unexpected project achievements. In this section you can list whether there has been changes in risks, whether they have become issues and whether new risks have been identified.

There is a risk associated with the delayed start of the project, in that if further delays occur the project may not be completed by the planned finish date. This is being mitigated by taking advantage of the fact that key staff are now working full-time on this project, rather than part time as was previously planned. The situation is being monitored regularly and the JISC programme manager will be contacted if a no-cost extension to the project is anticipated.

Section Nine: Collaboration and Support

What areas of work would you like to discuss with other projects?

None identified as yet, though the forthcoming JISC MRD workshop will provide an opportunity to network with other projects.

ACRID has formed links with the NERC SIS Data Citation and Publication project that is currently ongoing. (BADC are key partners in both projects.)

Is there anything that you would like advice and support on? Do you have any specific training needs, requests or suggestions for supporting workshops for the programme?

None identified at this time.

Section Ten: Financial Statement

In this section you should detail the expenditure of the project so far. Against the budget headings you should set out the expenditure for the reporting period, noting any significant over/under spend giving reasons for this. You should also state the total expenditure to date against each budget heading. The table below is designed to help this reporting process. Additional budget headings may be added to fit an individual project's budget. Projects may find it more appropriate to use a spreadsheet to report financial information.

See table below for financial information. Please note that the financial information is provided for the period ending 31 January 2011. Financial information up to the end of February 2011 will not be available until 7 March 2011.

Section Eleven: Next Steps

In this section you should very briefly list the activities planned and/ other information of relevance for the next stage of the project.

The next steps in the project will focus on research data management and will design and develop the required data management infrastructure to ensure capture of relevant metadata for climate research data, software, and workflows.

Similarly, work will begin on the topic of data citing, linking and integrating and will focus on publishing climate research data in citeable linked-data form, enabling integration with other datasets through use of common data models.

Please see the full project work plan for a more detailed explanation of the project's next steps.

Total Grant	119,773.00		Duration of project	12 months
Reporting Period	01.08.10 – 31.01.11			

UEA (period ending 31 January 2011)

Budget Headings	Total budget allocated	Expenditure this reporting period	Total expenditure to date	Further information
Staff – research	21,411.00	7,307.16	7,307.16	
Staff – secretarial & clerical	4,054.00	2,026.08	2,026.08	
Travel & Subsistence (T&S)	1,000.00	172.64	172.64	
Overheads	35,382.00	3,985.47	3,985.47	
Sub contract Salary	29,937.00	0.00	0.00	Expenditure has been incurred by the subcontractor (STFC), but no claims have yet been paid by UEA. Please see below for information on the costs accrued by STFC.
Sub contract T&S	800.00	0.00	0.00	
Sub contract Overheads	27,189.00	0.00	0.00	

STFC (period ending 25 February 2011)

Budget Headings	Total budget allocated	Expenditure this reporting period	Total expenditure to date	Further information
Staff	£64,833	£5,054.50	£5,054.50	Note that some of the staff bookings for time spent in February 2011 on this project have not yet come through the financial system to be reported here.
Travel & Subsistence	£1,000	£0	£0	
STFC Overheads	£60,129	£3,407.43	£3,407.43	

Checklist:

Before you return this report:

- ✓ Ensure that your project webpage on the JISC site is up to date and contains the correct information. Attach details of any required amendments to this report. Project webpages can be found from: http://www.jisc.ac.uk/whatwedo/projects.aspx

- ✓ If there have been any changes to the original project plan and/or work packages, ensure that amended copies of the relevant sections of your project plan are attached to this report.