The QUEST Earth System Data Initiative

Implementation Plan

1. Introduction

QUEST projects both use and produce an immense variety of global data sets that need to be shared efficiently between the project teams. These global synthesis data sets will also be a key part of QUEST's legacy, providing a powerful way of communicating the results of QUEST among and beyond the UK Earth System research community.

To facilitate this data exchange, and to avoid replication of the often labour-intensive efforts to source and visualize data, QUEST in its full operational phase will implement the QUEST Earth System Data Initiative – **QESDI** – a mechanism for *easy, centralized access* with *flexible statistical and visualization tools* for consistent processing and presentation of global data sets.

The project will last 18 months, from Feb. 2009 to July 2010, with a total cost of £252,093.

2. Objectives

The objectives of QESDI are to:

1. **Clearly document the data sets in use across QUEST, providing full metadata and information about their interconnectivity;**
2. **Facilitate the flow of data across QUEST, using agreed conventions and prescribed data formats;**
3. **Provide a mechanism for identifying and resolving emerging data needs;**
4. **Provide central coordination of selection, procurement and quality assessment of key external data sets;**
5. **Provide a common platform for data-sharing among QUEST project scientists and Earth system researchers in the wider community;**
6. **Provide bespoke community statistical analysis and visualisation toolkits;**
7. **Enhance the visibility of QUEST research activity;**
8. **Exploit NERC's latest web-based portal and archiving technologies for enhanced value of QUEST outputs.**

3. Rationale

The QUEST Core Team, project teams, and BADC, the designated data centre for QUEST, have held ongoing discussions over the past year about programme data management needs and about ensuring the best impact and legacy of QUEST science. Global synthesis data sets – a term that includes both the collation of local/regional data sets into a global product and the creation of derived or aggregated data from other global data sets – are both a key research tool and a high-value output of QUEST projects. By coordinating and facilitating the exchange of these data sets, QESDI is designed to maximise both the research efficiency through the remainder of the programme and the end value of QUEST.

A characteristic feature of the data management of QUEST, a strongly interdisciplinary programme, is that the producers and users of the data are often from very distinct research communities (with different data use conventions), extending the requirements of data services beyond normal needs. Projects primarily contributing data include PalaeoQUMP, QUAAC, QUERCC, Deglaciation and Quaternary QUEST. Synthesis data sets from these projects are already well advanced, and model output data sets are in development. Projects primarily using/exchanging data include QESM, GSI, QUEST-FISH and QUATERMASS.

The data needs of the QUEST projects have been scoped by BADC, who are already contracted to take project data and archive it with browsing and discovery services. QESDI will extend this basic data service by providing information internally in the programme about accessing data sets in co-development stages; delivering detailed meta-data with the peer-review/critique information that is needed for greater confidence in cross-community sharing; and developing a community visualisation toolkit based on R, a versatile programming language and software environment. Demonstration products showing how QUEST data can be explored using the toolkit will both provide a community testing-ground for some of QUEST's synthesis activity and be a means for the wider dissemination of the QESDI portal.

By making synthesis data sets readily available with a suite of analysis, mapping and visualisation tools, QESDI will deliver benefits to the QUEST community in its synthesis phase; the wider UK Earth system science (ESS) community and other NERC researchers with an interest in global change; and of course ultimately to global change scientists in academic and policy contexts worldwide. There are particular opportunities to be exploited in the context of Living With Environmental Change, where researchers across the UK will benefit from this central focus for global synthesis data.

Although the strong priority will be for exchanges among the QUEST community, interfaces will be designed to be accessible to external users. It is important to extend the use of QUEST data products and analysis tools, firstly so that future programmes do not need to start "from scratch" in collating data sets, often a rate-limiting step in research that bridges global modeling and observation. Also, sharing data sets through QESDI with analysts beyond the programme (e.g., in UKCIP and CEFAS) is a powerful means for driving community scientific development, and for continuing/improving science-policy links. Another aspect of this latter goal is that QESDI will be a tool to help QUEST scientists across the programme to engage better with stakeholders, by making authoritative quantititative information readily accessible.

4. QESDI management and support

BADC will manage the QUEST Earth System Data Initiative (QESDI) and develop the associated web services. BADC is already the provider of QUEST's data management services. This initiative adds substantial value to the basic service, by providing an incentive for project teams to lodge data before the end of the projects, and raising demand for access to BADC and its services.

Subcontracts will be issued to Tessella (WP1 – portal) and CEFAS (WP2, tools library). WP3 will be a cooperative activity involving all partners, including the QUEST Core Team.

A QUEST data users' group already exists (referred to as QDUG on workplan), with representation from the QUEST Core Team and from each of the QUEST projects. QESDI

will be a discussion focus for this group.

Similarly, QUEST scientists with an interest in visualisation, mapping and analysis have already been engaged by the Core Team in discussions about what is needed and what can be achieved. This input has resulted in a skeleton tools library, giving proof of concept and a good indication of the degree of uptake across QUEST and more widely in the UK global change research community (e.g., strong synergies can be delivered with the Tyndall Centre and the DECC AVOID activity). The Visualisation Working Group (VWG) members will continue to be engaged in QESDI in an advisory/consultative role. Because the Tools Library will be developed in R, an open-source language with a rapidly growing research and development community online, its progressive co-development will ensure its lasting viability and usefulness even after the QUEST investment is completed.

5. Workplan

The QESDI activities will operate in three work packages, described more fully below. The work packages are

1. Data collation and documentation (BADC lead, Tessella subcontract)
2. Visualisation and analysis tools (CEFAS subcontract to lead)
3. Coordination of activities and integration (BADC, with assistance from QUEST Core Team).

6. Deliverables

QESDI will provide a QUEST Core Archive (QCA) of data with enhanced metadata, a library of software tools configured to analyze this data, and the QUEST Earth System Portal (QESP) to give flexible access to the data, software and documentation.

The deliverables from the work packages are phased through the remaining lifetime of QUEST, with the goal of providing:

- baseline operability for both data handling and analysis tools within the first six months of the initiative,
- progressive improvement to a first order/prototype version for internal use before the main synthesis phase of the programme (i.e., by Sept 2009)
- a handover version for retention within the UK Earth system science community, by the end of QUEST (Sept 2010), to remain accessible through BADC at no further cost to QUEST (the data will be maintained indefinitely; BADC will continue support for the portal at a basic level, and will review the possibilities for further development with the community – a first review will be provided 3 months after the end of the contract).

Work Package (WP) 1 - Data Collation, Documentation, and portal development.

Deliverable	Project Month	Description

1.1	3	Report on data to be included in QCA, roadmap for data and metadata collection, provisional portal design specification.
1.2	6	Provisional portal
1.3	9	Prototype portal giving access to version 0.5 QCA
1.4	12	Progress report on QCA, and roadmap to definitive version
1.5	15	Portal documentation
1.6	18	Handover version portal - data will be held centrally, accessible by the QESDI website

WP2 - Visualisation and Analysis Tools.

Deliverable	Project Month	Description
2.1	2	Tools library v0 made available to Quest community - core tasks offline: plot gridded data; evaluate differences.
2.2	5	Library specification
2.3	9	Tools library v0.5 made available to Quest community - linked to WP1 first-order database (Deliverable 1.4)
2.4	12	Draft documentation
2.5	15	Tools library v1 made available to Quest community - fully documented handover version
2.6	18	Documentation updates

Figure 1 shows the dependencies between the deliverables in the form of a Gantt chart.

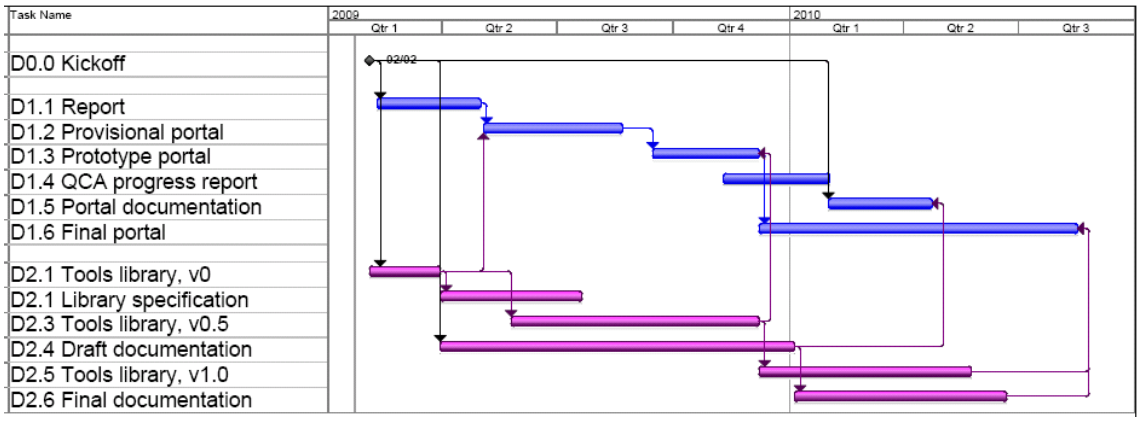

Figure 1 Gantt chart showing dependencies among deliverables

WP3 - Coordination and Integration

The deliverables for the coordination work package will be reports to NERC at 3 monthly intervals.

7. Tasks and activities

7.1 WP1 - Data Collation and Documentation, portal development [led by BADC]

Effort: BADC:, BADC subcontractor (Tessella). The BADC effort will be spread evenly throughout the period. The subcontractors will be engaged for 3 periods of 2 months, allowing for testing and gathering of feedback from the QUEST community in the intervals.

T1.1 The portal service specification and the associated metadata requirements will be determined through a consultation process targeting QUEST PIs through a workshop and email exchanges.

T1.2 A core collection of data, the QUEST Core Archive (QCA), will be identified, starting from the list in the Annex. Data within the QCA has the purpose of showcasing QUEST research and should therefore be open access. The provisional list of data sets for inclusion in the QCA will be circulated to PIs and discussed at a workshop to identify potential omissions. The data will be subject to the QUEST Data Policy with the additional condition that the data owners grant open access once they have been launched within the QUEST Earth System Portal (QESP).

T1.3 The QUEST Earth System Portal User Interface (QESPUI) will provide a flexible graphical access to data. It will be deployed in 3 stages: (1) a provisional portal to give basic access to the QCA; (2) a prototype portal with graphical interface and prototype versions of services identified in T1.1; (3) a final version. The final version will be made live at least one month before the end of the project.

T1.4 Documentation explaining access to QUEST's data and metadata will be prepared and included in the QESP.

T1.5 In order to ensure smooth implementation of the services, there will be frequent contact with data suppliers and with the QUEST core team.

T1.6 Three regional/thematic workshops will be held to gather user feedback as the project develops and to provide continuing interaction with users and data providers.

T1.7 Support will be provided to data producers to facilitate production of standard compliant data files. In addition to direct communication, advice will be provided in a FAQ document, users will be given guidance on the use of the CF-checker, and tools for simple attribute editing will be developed.

T1.8 In collaboration with WP2, an imaging package will be developed to provide the QESPUI with data in a graphical form. The package should ensure that appropriate and meaningful labels and colouring are provided automatically for all fields in the QCA.

T1.9 In addition to providing graphical access to the QCA, the QESP will give access, subject to the access controls described in the QUEST DMP, to the entire BADC QUEST archive.

T1.10 Preparation and publication of QESDI briefings in QUEST update communications

7.2 WP2 - Visualisation and Analysis Tools

Effort: Staff: CEFAS.

T2.1 Implement and package core/programme-critical tasks identified in VWG workshop and QESDI workshops.

T2.2 Set up online R toolkit developer to facilitate community input into the QESDI Software Library.

T2.3 Provide detailed documentation of the software library and work with WP1 to make these available through the QESDI Portal.

T2.4 Organise developer and user workshops, with support from WP3.

T2.5 Devise tutorials for Software Library users and work with WP1 to make these available through the QESDI Portal

7.3 WP3 - Coordination

Effort: BADC:; the QUEST Core team will also contribute effort.

The coordination will be carried out by BADC in collaboration with the QUEST core team: the latter will be funded for this work from their existing QUEST budget. The successful execution of WP3 depends on the pooling of BADC's specialist technical expertise with the core team's scientific synthesis expertise and knowledge of programme priorities. The division of labour in the tasks reflects this division of expertise.

T3.1 WP3 will coordinate reporting to NERC and ensure that there is coordination between WP1, WP2 and related QUEST activities [BADC].

T3.2 Administrative support of workshops: 3 regional/thematic workshops and, from WP2, 2 user developer workshops [BADC].

T3.3 Programme preparation, promotion and running sessions for the 3 regional/thematic workshops [Core Team/BADC].

T3.4 QESDI activity and products will be promoted through QUEST meetings and other NERC events [Core Team].

WP3 deliverables will be reports to NERC every 3 months, including reports from workshops.

8. Dependencies and critical path

The timeline for the QESDI workplan has been designed to ensure that QESDI is developed and delivered within the timeframe of QUEST, supporting the projects in the run-up to their completion, and supporting the collective synthesis activity of the whole programme. For greatest cost-effectiveness, deliverables must fit within this timeframe in order to exploit the opportunities presented by existing QUEST activities, such as the

Annual Science Meetings.

The deliverables in WP1 and WP2 are constructed to ensure that each work package has timely input from the other.

Annex: Initial data list for QUEST Core Archive

Candidate datasets for inclusion in the QCA are listed below. This list may be modified during the project.

A1 External data:

A1.1 Spatio-temporal observations

- gridded monthly climate observations (CRU)
- synthesized, medium-resolution EO data (examples: land/ocean greenness, burned area; drawing especially on NERC EO centre outputs)

A1.2 Geographic data

- land cover/land use
- soil types and propeties
- protected areas, diversity hot-spots

A1.3 Spatio-temporal scenarios

- climate change (derived from CMIP3 and Hadley Centre)
- population change
- vegetation change
- land use change

A1.4 Non-geographic time series

- CO_2 concentrations at monitoring sites
- selected flux measurements at eddy covariance sites
- GHG and reactive gas emissions
- yields (crops, forests), fish catches
- river flows
- leaf carbon isotope measurements
- biodiversity indicators
- socio-economic indicators (countries/regions)
- water extraction

A1.5 Non-geographic scenarios

- GHG emissions scenarios
- socio-economic scenarios (e.g. trade, population)

A2 QUEST outputs:

A2.1 Spatio-temporal scenarios

- outputs from QUEST Earth System model (QESM)
- impact assessments: water, food, biodiversity, and other sectors (GSI)
- biofuel use, forestry carbon sequestration scenarios (QUATERMASS)
- carbon balance scenarios (CCMAP)

A2.2 Palaeodata and reconstructions

- palaeoclimate reconstructions (PalaeoQUMP, Deglaciation, PETM)
- synthesized pollen data from long time series (Quaternary, QUAVIDA WG)
- transient palaeoenvironment simulations (Deglaciation, Quaternary)
- sedimentary charcoal, peat (QUEST Deglaciation, Fire WG)

A2.3 Non-geographic scenarios

- food supply/demand (GSI, QUEST-Fish)

The data will be provided with comprehensive metadata including literature references, references to authorship, creation data, revision history, and reference to (or documentation of) detailed methodology.