Customer	**:**	**University of Reading**	**Document Ref**	**:**	**EO4CDS.REP.001**
Contract No	**:**	**RE006432**	**Issue Date**	**:**	**17 March 2016**
WP No	**:**	**5**	**Issue**	**:**	**1.1**

Title : **UKSA CDSSG – Task: 5 Cost Estimation of routine climate processing system activity**

Abstract : This document is the deliverable associated with Task 5 of Phase 2; “Assessment of costs of implementing proposed system”. The purpose of the document is to estimate the costs associated with the implementation of an operational climate data production c

Author :

Kevin Halsall
Alex Hayward
Telespazio VEGA UK Ltd

Approval :

Accepted :

Distribution :

Hard Copy File:
Filename: EO4CDS.REP.001_1.1.doc

Telespazio VEGA UK Ltd
350 Capability Green, Luton, Bedfordshire LU1 3LU, United Kingdom
Tel: +44 (0)1582 399 000 Fax: +44 (0)1582 728 686
www.telespazio-vega.com

TABLE OF CONTENTS

AMENDMENT POLICY

This document shall be amended by releasing a new edition of the document in its entirety. The Amendment Record Sheet below records the history and issue status of this document.

AMENDMENT RECORD SHEET

ISSUE	DATE	DCI No	REASON
1	15 Mar 2016	N/A	Initial Issue
1.1	17 Mar 2016	N/A	Updated following University of Reading review

1. INTRODUCTION

1.1 Purpose and Scope

The Climate Data from Space Stakeholder Group (**CDSSG**) has a long term goal to develop and demonstrate a "seamless supply chain for climate data from space", exploiting existing infrastructure to develop better multi-layer, multi-data services.

The focus of Phase 2 of this activity is to develop a professional system design for the seamless supply chain, using input from the CDSSG commissioned implementation plans produced by EO researchers to show how their mature "essential climate variable" products could be implemented in an operational framework.

This document is the deliverable associated with Task 5 of Phase 2; “Assessment of costs of implementing proposed system”. The purpose of the document is to estimate the costs associated with the implementation of an operational climate data production capability at the Climate and Environmental Monitoring from Space (**CEMS**)-JASMIN national infrastructure, based on the outputs of Task 2.

1.2 Referenced Documents

The following is a list of documents with a direct bearing on the content of this report. Where referenced in the text, these are identified as RD.n, where 'n' is the number in the list below:

RD.1 UKSA CDSSG Task 2: System design for routine climate data processing at Jasmin-CEMS, 01_UKSA_CDSSG_SystemDesign_V1.0, v1.0, 10th March 2016

1.3 Definitions of Terms

The following terms have been used in this report with the meanings shown.

C3S	Copernicus Climate Change Service
CCI	Climate Change Initiative
CDSSG	Climate Data from Space Stakeholder Group
CEMS	Climate and Environmental Monitoring from Space
ECV	Essential Climate Variable
IEA	Institute for Environmental Analytics
KPI	Key Performance Indicators
RD	Reference Document
SLA	Service Level Agreement
UI	User Interface
VM	Virtual Machines

2. EXECUTIVE SUMMARY

This document provides an estimation of the cost for the development an operational climate data production capability at CEMS-JASMIN, based on RD.1. CEMS-JASMIN was established as an R&D rather than operational infrastructure, and the UK currently has no community operational processing facility. The scope of the estimation covers the necessary software development, the required upgrades to the archive capacity & processing capability and the additional requisite support personnel to upgrade CEMS-JASMIN.

The overall total cost for all these elements is:

Hardware & software development costs:	**£825k**
- **Software Build:**	**£335k**
- **Software Maintenance (to KO +2 years)**	**£60k**
- **Additional storage**	**£400k**
- **Additional processing cores**	**£30k**
Additional support personnel:	**3 FTE**

Provision of this funding to develop CEMS-JASMIN would enable the development of a leading operational processing facility, providing a clear advantage for UK companies in bidding for opportunities from the ECMWF in their ECV procurement, the CCI+ programme and the exploitation of Senintel-3 data.

2.1 Background

A long term goal for the UK space sector, as expressed in the Space Innovation and Growth Strategy Action Plan, is to “Position the UK at the leading edge of exploitation of a wealth of institutional Earth observation data by creating a Climate Services Centre for Europe in the UK”.

To realise this goal then the UK must develop and demonstrate the "seamless supply chain for climate data from space", exploiting existing infrastructure to develop better multi-layer, multi-data services for specific sectors. This will generate growth and export UK capability to a large (£12.3bn in 2010/11) and strongly-growing (estimated 9.8% pa in 2015) world-wide climate services market. The concept of “seamless supply” implies sustained, trusted, robust, accessible, timely, highly usable flows of climate-quality data to multiple types of users, including commercial climate services.

2.2 Existing Achievements

The world leading climate science, meteorological and industrial capability in the UK, together with previous HMG investments in the JASMIN-CEMS facility at Harwell, has resulted in significant national achievement:

- ESA Climate Change Initiative (**CCI**) – the UK leads the Sea Surface Temperature, Ocean Colour and Data Access Portal projects as well as the Climate Modelling User Group.

- Institute for Environmental Analytics (**IEA**) – funded by the HEFCE Catalyst Fund, the IEA is a partnership of organisations across the whole supply chain. It addresses wider environmental challenges, but has already initiated case studies specifically concerning climate with end users such as Sainsbury’s.

- Climate Data from Space Stakeholder Group (CDSSG) – funded by the UKSA this group is made up of UK experts from academia, government and industry. Group outputs include addressing key parts of the supply chain in the form of case studies e.g. climate data for environmental consultancies and seasonal forecasts.

- EC funded Copernicus Climate Change Service (**C3S**) – the UK is leading four of the seven Sectoral Information System projects recently procured by the C3S-entrusted entity ECMWF, in the sectors of Water, Energy, Agriculture & Insurance.

2.3 What is Needed?

Operational production of climate data will be funded by the C3S through competitive Europe procurements, with a separate contract / project for each Essential Climate Variable (**ECV**). ECV projects will allow specific processing chains and algorithms to be "operationalised" to allow reliable and robust data feeds into the C3S. There are shared tools that will be needed across multiple ECV projects and the UK has the opportunity to put these in place alongside the JASMIN-CEMS infrastructure. On the other hand no one ECV project will have the capacity to fund the implementation of such tools and therefore there is a compelling opportunity for the UK to take a lead so that shared benefits and economies of scale can be realised across the ECV provision activities.

2.3.1 Software Development

Table 2-1 identifies particular software modules that could be developed that would support the provision of an operational processing chain. They have been sub-divided into the different modules in order to provide a view on the granularity of the proposed development activity, and identify areas that could be developed at a later data should sufficient resources not be available in the first instance. **The Orchestrator and its Dashboard are the key core developments** required; however, it is strongly recommended that they all be developed as part of a single activity.

Table 2-1 Software Module Costs

Module	Description	Estimated Cost
Orchestrator	This will involve software tools to schedule and execute the data processing chains, automatically scanning for input and output data	£190k
Orchestrator Dashboard	To allow engineers to visually check on processor and data output status, and specific execution / trial of a processing chain on a set of virtual machines	£55k
Comment collection	For collecting comments on data and its use from data users. For the purposes of this analysis ,it is assumed that CHARMe could fulfil this criteria, with some adaptations	£35k
Provenance generator	This would compile a provenance file when a processor completes, listing the data inputs, the algorithms used and their version numbers	£35k
Visualisation Tool	The visualisation tool would allow the project teams to undertake a basic visual check of the datasets to identify any issues.	£20k

Software Maintenance

Software maintenance costs are estimated at 12%, or £40k (£25k-£60k) annual average for bug fixes and updates required by changes to other software

The resulting total cost of the development activities over two years (9 month development schedule) is £395k, inclusive of maintenance.

2.3.2 CEMS-JASMIN Resources

No major modifications are planned for the CEMS-JASMIN infrastructure in order to accommodate the operational processing chain infrastructure; however, the existing facilities will need to be upgraded to accommodate additional archive capacity and processing cores. There will also be additional manpower needed to help support the projects and undertake the necessary data ingestion activities for new datasets.

Additional Storage (1PB on-line, 2PB off-line): ~£400k

Additional Processing Cores (100/ 2 years): ~£30k

Additional manpower for data ingestion and project support activities: 2 FTE

2.3.3 Operational Support

In addition to the support provided by CEMS-JASMIN, the Processing Chain Providers will require support for operational activities, such as upgrading the processing software, scheduling, cataloguing, testing and code integration. Given the operational nature of the activity, this support could also optionally extend to QA of the output data, helpdesk provision and monitoring of any associated Service Level Agreement (**SLA**) Key Performance Indicators (**KPI**).

Operational Support Provision: 1FTE

2.3.4 Up Front Costs

One of the key issues to note is that even though the individual projects running the operational chains will pay the costs related to hosting and manpower support, an initial up-front investment will be needed in order provide that capacity in the first instance. It is also difficult for academic institutions in particular to split staff across different roles; if supporting a new project team requires 0.25FTE, then often that role will need to be fulfilled by a full time member of staff brought in in advance, with the expectation that further projects will subsequently increase utilisation of that resource.

2.4 Benefits

Funding this initiative would allow the UK to:

- Offer higher value bids to ECMWF in their ECV procurement, giving the edge over European competitors; the UK could potentially win work across 6 of the 9 ECV projects bringing in UK revenues of €3M by mid-2018, with potential follow on revenues of €1M per year when C3S operations commence.

- Offer higher value bids for the CCI+ programme in the period 2017-2021, giving powerful, efficient, and flexible processing capability for CCI+ ECV projects in the “big data” era of Sentinel data.

- Take a European lead in the operational provision of ECV data, to be used in support of commercial climate services, policy decision making and wider research concerning the impact of climate change.

- Prepare for efficient exploitation of Sentinel 3 data, where the UK has world-leading science expertise

3. METHODOLOGY

Based upon the requirements and system design defined in RD.1 from Task 2, the cost estimate of the system can be divided into the following categorisations:

- Software module development

  Dedicated software modules may developed and deployed within the CEMS-JASMIN infrastructure in order to facilitate the teams (*Processing Chain Providers*) running the operational processors. These are not individual processor specific tools, but rather communal software modules designed to support the teams and reduce the load related to common activities.

- Additional infrastructure & manpower resources for CEMS-JASMIN

  The introduction of operational processing chains to the CEMS-JASMIN infrastructure will place a potentially significant additional load on it outside of its original specification. Additional resources are likely to be required in order for it to be able to support an operational climate production facility

- Operational manpower support for the Science Teams

  In addition to the direct support the Processing Chain Providers will require from CEMS-JASMIN, there are additional services and levels of support that could be made available by a dedicated operational support team that, like the development of specific software modules, may reduce their load by fulfilling common activities.

- Processing Chain Providers

  This document will also give an indication of the load that will have to be borne by the science teams themselves, who have responsibility for the development of the processors and maintenance of their operation.

A matrix of these categories against the system design requirements may be found in Appendix A.

3.1 Software Development Methodology

The initial development estimation identified in Section 4 followed the four following steps:

- Form representative assumptions of the most important unspecified details of design and requirements, and of the development team and environment. These serve to make consideration of software size more concrete and to parameterize the cost model, and are not intended to constrain design or expected to be completely accurate.

- Split the software in to assumed components and estimate the size of each in thousands of lines of code (kSLOC), by using judgement and comparison to other projects.

- Use the COCOMO II cost model to turn size estimates in to effort (person-month) estimates.

- Multiply by an assumed average software-developer day-rate.

The range given uses the Cone of Uncertainty at the requirements stage.

Maintenance is a more poorly researched area, and was estimated using typical software industry maintenance costs (75% of total costs, 60% of which is for enhancements) and software lifetimes (10 years). This produces a total maintenance cost of three times development costs, or 30% per year, from which enhancements were deducted.

4. SOFTWARE DEVELOPMENT COSTS

Based upon the system design defined in RD.1, the following components have been identified as common modules required to facilitate the development of the operational climate production facility in CEMS-JASMIN.

- Orchestrator

  The core software to schedule and execute the data processing chains, and to automatically scan for input and output data

- Orchestrator Dashboard

  Software to allow engineers to visually check on processor and data output status, and run one-off processing

- Comment Collection

  Collecting comments on data and its use from data users

- Provenance Generator

  Compiling a provenance file when processors complete, detailing data inputs and the processing software used

- Visualisation Tool

  A basic visualization facility for available data to check for obvious problems

Total development effort of the defined components is estimated at 32 person-months, a cost of £335k

Considering typical errors at this stage a range of £225k-£505k is given.

4.1 Component Cost Breakdown

Table 4-1 Component Cost Breakdown

Component	Effort (person-months)	Cost (£1000s)
Orchestrator	18	190
Orchestrator dashboard	5	55
Comment collection	3	35
Provenance generator	3	35
Visualisation Tool	2	20
Total	**32**	**335**

The above costs:

- **Includes** detailed software design work, software development, software testing, configuration control and project management of each of those pieces.

- **Excludes** any further analysis of requirements, tender processes, and future upgrades.

4.2 Maintenance

Software maintenance costs are estimated at 12%, or £40k (£25k-£60k) annual average for bug fixes and updates required by changes to other software.

This will rise if enhancements increase the software's size, and may be larger at the beginning of the software's life or after changes (15%, £50k, or £30k-£75k) and lower mid-life.

With a nine-month development schedule, this gives a **two year cost of £395k (£265k-£590k) –** development and maintenance inclusive.

If the maintenance was to be upgraded in order to include enhancements as well, £100k/year is predicted, but can't be estimated well. Total maintenance costs may rise again towards the end of software life as functionality diverges from original design criteria, code quality falls and software size rises.

4.3 Ownership

The software is intended to be developed as open source, enabling anyone to use it for any type of activity (research, commercial etc.).

With regards to the integration of external additional modules that form part of the software development (comment collection, visualisation tool), any separate IP issues will need to be determined, along with identification of the most appropriate module, though open source options for these are also available.

4.4 Assumptions

The following assumptions were made in the above calculations:

Functionality

- Input and output data comes with or is given a unique identifier for each data stream
- Given an identifier and a date range, the location of data files can be found
- Input data can be accessed on a shared file system with no additional action
- Processing can be run for fixed sized windows (eg, 1 day or every six hours)
- Processing for a window does not begin until all previous windows have been completed or cancelled
- Scheduling, especially in the case of missing or late input data, is done by asking processing chains to make their own go/not-yet decision based on available data, and that decision is based only on whether the input data covers the window or not
- Output can be delivered to a shared file system location, and delivered almost as-is from the processing chain

- Onward transmission of the data to e.g. ECMWF has not been included (but could be configured at minimal additional cost)

- Processing chain software quality or process requirements are enforced by contractual requirements, not by delivered software

- CHARMe can be used with new-line-equivalent modifications equal to a third of its size (which is 6kSLOC); an assessment of the suitability of CHARMe was not carried out

- The visualisation tool was costed on the assumption of using the University of Reading's Godiva tool, or TVUK's Peep tool (based on Godiva), the latter of which is currently employed for a similar role in ESA's CCI Portal. Implementation of these tools within CEMS-JASMIN includes the appropriate configuration of WMS and web servers in order to be able to serve the data to users.

Development process and environment

- The developers and development organization have experience with similar software, understand the goals of the project well and have experience with the platform, language and tools used

- Good and mature software and project management processes are used

- Average developers and designers are assumed

- Stakeholders come from multiple organizations and environments and may have divergent priorities

- Reliability requirements are not high because failures are not immediately visible to end-users and can be recovered internally

- None of the software is intended to be reused elsewhere

- Development will involve participants and stakeholders across multiple organizations and locations

- An unusually compressed schedule (less than nine months) is not required

Design

- The orchestrator is configured with the identities of relevant data, a schedule of expected data arrivals and with processing chain locations (test and live Virtual Machines (**VM**)). Each chain has a set of data of interest.

- This configuration is simple to implement, for example in files or directly in a database, and not via a user interface.

- When an event of interest occurs (data arrival or missing scheduled arrival), the processing chain, not the orchestrator, decides whether it should run or not.

- When a decision to run processing is made then attempts to do so begin immediately; no special scheduling (e.g. to smooth resource use) is used.

- Running a processor or passing an event of interest to it is simple, for example by using ssh and a wrapper to run a command which is part of the processor

- Retrying after failures is simple (e.g., if no output is received within <x> time then retry up to <n> times with a fixed delay)
- Once begun, no further attempts to run processing for a particular time period occur except for retries. For example, if preferred input data arrives late then no automatic reprocessing occurs.
- Alerts to operators are via e-mail
- The orchestrator dashboard is a web interface, accessible from managed VMs
- Authorization is simple (e.g. by IP or a manually maintained password list) and there is no User Interface (**UI**) required for managing users
- Processing chains will ask for all of their inputs via the system and can produce a version or commit number for themselves and their components
- Combining this data with provenance data from the inputs is sufficient to generate provenance data for the output
- Monitoring of resource use uses existing CEMS facilities, not new software

5. CEMS-JASMIN INFRASTRUCTURE/RESOURCE COSTS

The existing CEMS-JASMIN infrastructure was not designed to run operational processing chains, and as such, additional resources to supplement the existing system will be required.

- The operational processors will require a number of additional input datasets not current in CEMS-JASMIN, for which additional storage capacity will be needed. In a number of cases, given their operational nature, these datasets will by necessity need to reside on a short time-delay retrieval system.

- The additional datasets, if not already stored in CEMS-JASMIN, will have to be appropriately characterised by the CEMS-JASMIN team for ingestion and therefore there will be additional effort required in order perform this activity.

- In addition to storage and dataset management, there will very likely be a need for additional processing power within the cluster in order to support the regular activities of the chains.

- The CEMS-JASMIN team are responsible for providing a certain level of support to each of the projects, both with respect to the system (electricity, on-going operations and maintenance) and the projects themselves (reporting, meeting attendance etc.)

There are a large number of variables and unknowns related to both the storage and processing issues, which makes accurate estimates of costs difficult to calculate.

5.1 Storage

Estimates for the additional storage capacity required for the CEMS-JASMIN archive is based upon previous experience with the CCI ECV teams that have used the infrastructure for their processing.

Workspace areas of each of the teams can range from 100-200TB (Cloud, SST) to 800TB (QA4ECV). The corresponding output from each team is in the order of 100TBs in size. In an operational scenario, it is assumed that the Sentinel data will also become part of the input data for many teams (through shared access to the single archived version on CEDA), and their corresponding output.

Based on this, an additional 2-3PB of archive capacity would be a modest estimation of the required upgrade to the CEMS-JASMIN infrastructure. This capacity would be split between disc (for fast retrieval storage) and tape (lower cost, but lower speed retrieval).

The cost of this is estimated to be: £400k

[1PB disk: ~£300k, 2PB tape: ~£100k]

5.2 Dataset Ingestion

For each additional dataset to be stored on the CEMS-JASMIN archive, it is required to undergo an ingestion process to be catalogued and archived on the system. This is a single one off cost done in advance of the dataset be archived.

Typically, the effort required for each dataset is roughly 0.1FTE

As an indicative value, 6 of the CCI ECV teams who might feasibly use CEMS-JASMIN for operational purposes were consulted, and a combined estimate of the effort required to ingest all of the input datasets for these teams was calculated.

The additional resources needed by CEMS-JASMIN for these projects would be in the region of 1.5-2 FTE.

5.3 Processing

The large numbers of variables an unknown make predictions of the required additional processing power particularly difficult at this stage. The estimates therefore, are based on experience on previous projects, specifically the CCI projects that utilised CEMS-JASMIN from 2012 (SST, Cloud).

As an initial estimate, **100 cores for 2 years** seems a reasonable figure that also allows for a flexible configuration allowing the processing to increase beyond the baseline figure for short periods of time.

The cost of this is estimated to be: ~£30k

As this is simply an upgrade to existing hardware, these expansions can be implemented within a very short lead time.

5.4 Support

Each project team will incur additional costs related to support:

- System Support

  Each project contributes to system support and electricity costs for running the storage system. **£4k / year / 100TB**

- Project Support

  CEMS-JASMIN personnel are required to provide support to the projects such as for attending meetings, reporting etc. **Typically £3k / year per project**.

In supporting the projects, CEMS-JASMIN will also require the appropriate manpower to attend meetings, undertake reporting etc., so the above figures will also equate into additional FTEs needed by CEDA.

5.5 Up Front Costs

One of the key issues to note is that even though the individual projects running the operational chains will pay the costs related to hosting and manpower support, an initial up-front investment will be needed in order provide that capacity in the first instance. It is also difficult for academic institutions in particular to split staff across different roles; if supporting a new project team requires 0.25FTE, then often that role will need to be fulfilled by a full time member of staff brought in in advance, with the expectation that further projects will subsequently increase utilisation of that resource.

5.6 Economies of Scale

Should additional major data streams (OLCI, S5-P) be made available through the CEMS-JASMIN infrastructure, then additional Processing Chain Providers may wish to make use of the provided facilities, taking advantage of infrastructure established based

on the design defined as part of Task 2. Some configuration of the orchestrator may be required, and the datasets will need to be ingested as any other dataset described above (these additional tasks to be carried out by the required manpower resources identified in Section 5.2), but no major modifications will be needed to take account of the new data streams.

Having additional major data streams will not only encourage the use of the infrastructure by Processing Chain Providers wishing to make use of them, but may also encourage the development of synergy products through the availability of so much data through the facility.

6. OPERATIONAL SUPPORT COSTS

In addition to the support provided the each of the Processing Chain Provider teams by CEMS-JASMIN, an additional level of support is identified in the design requirements related to processing SW upgrades, coding, scheduling etc. (see 'Ops Team' column in Appendix A).

For the purposes of this analysis, this has been separated from the CEMS-JASMIN system support activities, and could be offered by the providers of the Orchestrator software modules, or a separate service provider.

The requirements identify the following support to be provided as a minimum to the Processing Chain Provider teams:

- upgrades to processing software
- assistance with scheduling
- assistance with creating correct catalogue entry information for products
- testing at all levels
- code integration and dependencies

Additional common services that could be provided to the teams, given the operational nature of the activities, might also include:

- Helpdesk
- QA of output data products
- Service Level Agreement support

Based on TVUK's experience in offering a similar service to ESA's EO satellites through the IDEAS+ service, the estimated manpower effort for these is 0.5-1 FTE depending on the particular project specification.

7. PROCESSING CHAIN PROVIDER COSTS

By necessity, some of the costs resulting from the requirements of implementing an operational processor chain on CEMS-JASMIN will be borne by the project teams (Processing Chain Providers).

This section provides an indicative overview of these costs (where possible).

The support costs with regards to the CEMS-JASMIN infrastructure have been described in Section 5.4. Based on a project requiring 400TB storage, this roughly translates to a cost in the region of £20k/year.

Hardware storage costs (as distinct from the support) are:

- £300 /TB for on-line storage
- £50 /Tb for offline tape storage

Using the above example (estimating 400TB split evenly between on-line and off-line storage), this would give a cost of £70k.

In addition to the costs resulting from CEMS-JASMIN usage, there will also be certain software design requirements to fulfil (see 'Processor Design' column in Appendix A), as part of the development of the processor; however, at this point these cannot be easily estimated.

The requirements relating to particular software modules from Section 4, such as that relating to provenance, if not funded as part of a central provision of services to the Processing Chain Providers, will need to be absorbed by each of the teams individually. This can roughly be estimated to equate to a significant proportion of the overall cost of the software module provided in Section 4.

Summarised total costs per project:

- **One-off £70k up front**
- **£20k per year support from CEMS-JASMIN**
- **Software redesign/reconfiguring – project-dependent**

Note: the figures for project teams are indicative and will need to be confirmed with CEDA on a case-by-case basis for inclusion in any bids

APPENDIX A. DESIGN REQUIREMENT CATEGORISATION

				Software Modules					CEMS-JASMIN			Science Team		
Function	**Sub-function**	**#**	**Requirement**	**Orchestrator**	**Orchestrator dashboard**	**Comment collection**	**Provenance Generator**	**Visualisation Tool**	**Existing i/f / service**	**Additional i/f**	**Ingestion of new data**	**Processor design**	**Ongoing**	**Ops Team**
Data Ingest	Feedback for issues/quality	1.1.1	the system must provide a mechanism for collecting comment against dataset or specific data file within a dataset for onward transmission to producer. (also for products - see later).			x								
Data Ingest	Feedback for issues/quality	1.1.2	ingested data must meet quality standard and/or has associated provenance to demonstrate quality.				x							
Data Ingest	Monitoring Input	1.2.1	the system must alert monitoring personnel of failure/delay of expected data collection within specifiable time and/or retries.	x										
Data Ingest	Monitoring Input	1.2.2	the system must alert dependent processing chains.	x										
Data Ingest	Monitoring Input	1.2.3	the system should prompt processing chains to optionally retry when data becomes available.	x										
Data Ingest	Monitoring Input	1.2.4	the system could provide an interface for monitoring processed data and output files produced.		x									
Data Management	Archive Retrieval	2.1.1	the system must provide a mechanism to specify the data file required and to automatically load it or otherwise make it available.	x										
Data Management	Archive Retrieval	2.1.2	the system must provide a facility to discover/search the data archive for specific data using keyword/source id/DTG.	x										
Data Management	Archive Retrieval	2.1.3	the system should offer alternate processing chain options dependent on dataset availability where the software offers this. The processing software could offer either an automatic or manual mechanism.	x										
Data Management	Archive Retrieval	2.1.4	the system must make the continuity of data clear - i.e. rolling data is usually required and available but if not, the system stops the operational run, and would instead use the data at a later time when available in a reprocessing run. This provenance must be recorded and additionally, flagged up to monitoring personnel.				X							
Data Management	Quality Assurance of	2.2.1	output datasets must have a suitable marker to indicate maturity of production method; this will be linked to the CORE-Climax maturity matrix score in sections:									x		

	Output - maturity matrix		software readiness, metadata, some of user documentation, some of uncertainty characterisation and some of public access, feedback and update(see annex).											
Data Management	Storage Needs	2.3.1	the system must provide sufficient disk storage space. The rates of increase estimated from above table and trial implementation.						x					
Data Management	Storage Needs	2.3.2	the system must provide disk space warnings. This will require personnel to monitor and upgrade as needed.						x					
Data Management	Storage Needs	2.3.3	the system must accommodate data preservation for an extended period.						x					
Data Management	Processing Needs	2.4.1	the system must provide sufficient processing capacity.							x				
Data Management	Processing Needs	2.4.2	the system must provide processing statistics. This will require personnel to monitor and upgrade as needed.	x										
Scheduling	User support for optimised processing	3.1.1	the system already provides batch queueing system and options to control when jobs are run - the system should make this more straightforward by one or more of:											
Scheduling	User support for optimised processing	3.1.1a	simplification of job submission options;	x										
Scheduling	User support for optimised processing	3.1.1b	provision of a simple submission interface;	x										
Scheduling	User support for optimised processing	3.1.1c	provision of technical support.											x
Scheduling	System Resources	3.2.1	depending on loading figures for trial implementation, full system may well need increase in processing cores.							x				
Scheduling	System Resources	3.2.2	the system's scheduling must support an optimal loading model whereby processing and other resources are balanced during operation.	x										
Workspace	VM Creation and Update	4.1.1	the system shall provide dedicated managed VMs, one for each live processing chain rather than per project.						x					
Workspace	VM Creation and Update	4.1.2	the system shall provide a dedicated managed VM for testing code baseline changes after update releases for each processing chain.						x					
Workspace	VM Creation and Update	4.1.3	the system shall enable a straightforward method to alternate live/test VMs so that the VM with tested latest release software becomes the one taking responsibility for the processing chain.						x					
Processing Chains	Testing; Standards for the processing chains' software	5.1.1	the system shall accept processing chain software only with a minimum set of tests to validate and verify that software including unit, module, system and regression tests: each of these will grow as new features are brought online.									x		
Processing Chains	Testing; Standards for the	5.1.2	the system shall accept processing chain software error fixes only with a concomitant library of error tests which will grow as errors are fixed.									x		

	processing chains' software													
Processing Chains	Testing; Standards for the processing chains' software	5.1.3	the system shall accept processing chain software only where coding standards are adhered to. These can be a public standard or documented in-house by the research group.									x		x
Processing Chains	Testing; Standards for the processing chains' software	5.1.4	the system shall accept processing chain software only where there is evidence of code review against the standards.									x		
Processing Chains	Testing; Standards for the processing chains' software	5.1.5	the system shall accept processing chain software only where all code is version controlled (e.g. using GitHub or BitBucket).									x		
Processing Chains	Testing; Standards for the processing chains' software	5.1.6	the system shall record provenance details (as detailed above) and provide them as part of the QA accreditation of output data products.				x							
Processing Chains	Sharing Common Functionality	5.2.1	the system should facilitate sharing common library code by multiple processing chains.	x										
Processing Chains	Sharing Common Functionality	5.2.2	the system shall record code modules used and provide them as part of the QA accreditation of output data products.	x										
Processing Chains	Compilers, SW & licenses	5.3.1	the system shall use open source compilers wherever possible.	x										
Processing Chains	Compilers, SW & licenses	5.3.2	the system shall provide compilers and libraries at the same version across all processing chains; plans for upgrade of processing chain software must be in place if they cannot use the system configuration.	x										
End Products	Rudimentary Visualisation	6.1.1	the system products shall be in NetCDF compliant format wherever possible.									x		
End Products	Rudimentary Visualisation	6.1.2	the system shall provide a basic visualization facility for available data to check for obvious problems. (?use the Godiva platform?, Barry's viz.)					x						
End Products	Cataloguing	6.2.1	the system shall provide a mechanism to assign indexing information to output data products.	x										
End Products	Cataloguing	6.2.2	the system shall provide a facility to add indexed output data products to its searchable data archive (see Data Management above).	x										
End Products	Cataloguing	6.2.3	the system's cataloguing subsystem should provide indexing and search facilities for users to find required data product.	x										
End Products	Data availability	6.3.1	the system should provide its catalogued searchable data to a wider user community: this may include public access.						x					
Accessibility	Create login and	7.1.1	carry out review of the current method for accessing the Jasmin system: some users							x				

	set up secure connection		find steps to use Jasmin facilities off-putting due to the seeming complexity. Investigate whether there are methods to (partially) automate the process, and certainly provide more training and/or support for new users.											
Accessibility	Request resource or use common science VM	7.2.1	It is envisaged there will be dedicated VMs for the processing chains producing the standard output datasets but the system should also provide mirror VMs available for other users to set off processing chains: resources must be allocated carefully for this as it would be lower priority.						x					
Accessibility	Support for users and system	7.3.1	support for research groups and users will be essential for the following tasks (as a minimum):											
Accessibility	Support for users and system	7.3.1a	upgrades to processing software											x
Accessibility	Support for users and system	7.3.1b	assistance with scheduling											x
Accessibility	Support for users and system	7.3.1c	assistance with creating correct catalogue entry information for products											x
Accessibility	Support for users and system	7.3.1d	testing at all levels											x
Accessibility	Support for users and system	7.3.1e	code integration and dependencies											x

This Page Is Intentionally Blank