

a. Proposal Cover Sheet

JISC Grant Funding 10/09

Cover Sheet for Proposals <i>(All sections must be completed)</i>		Rapid Innovation Programme	
Name of JISC Initiative:		JISC Rapid Innovation Grants	
Name of Lead Institution:		University of Reading	
Name of Proposed Project:		SocialGodiva	
Name(s) of Project Partner(s):		University of Southampton (sub-contractors)	
Full Contact Details for Primary Contact:			
Name: Dr J.D. Blower			
Position: Technical Director, Reading e-Science Centre			
Email: j.d.blower@reading.ac.uk			
Tel: 0118 378 5213			
Fax: 0118 378 6413			
Address: Environmental Systems Science Centre, Harry Pitt Building, 3 Earley Gate, Whiteknights, Reading RG6 6AL			
Length of Project:		10 months	
Project Start Date:		Project End Date:	31 st October 2010
1 st January 2010			
Funding Requested	from	JISC	£38,475.04
(Jan10-Jul10):			
Funding Requested	from	JISC	£17,452.16
(Aug-Oct10):			
Total Funding Requested from JISC		£55,927.20	
(Jan10-Oct10):			
Total Institutional Contributions:			
Outline Project Description: This project will combine two existing systems to create a new Virtual Research Environment for environmental scientists. By combining the capabilities of an online environmental data visualization system with a sophisticated blogging and collaboration engine, we will rapidly create a new facility for scientists to collaborate on important data analysis tasks, focussing here on the diagnosis of the latest cutting-edge numerical simulations of the climate and oceans.			
List of priority areas, highlight each that applies:			
<input type="checkbox"/>	Discovering, Representing and Publishing Information and Results		
<input type="checkbox"/>	Analysing Data and Performing experiments		
<input type="checkbox"/>	Communication and Collaboration		
<input type="checkbox"/>	Managing research and research projects		
<input type="checkbox"/>	Actively engaging Businesses and the General Public (BCE)		
I have looked at the example FOI form at		YES	

Appendix A and included an FOI form in the attached bid (Tick Box)		
I have read the Funding Grant and associated Terms and Conditions of Grant at Appendix B (Tick Box)	YES	

b. Appropriateness and Fit to Programme Objectives and Overall Value to the JISC Community

b.1 Background

b.1.1 The **Godiva2** online data visualization system is developed and maintained by the Reading e-Science Centre at the University of Reading (<http://www.reading.ac.uk/godiva2>). It provides a means for scientists to browse interactively in a “Google Maps-like” fashion through large environmental datasets, including numerical model outputs and high-resolution satellite imagery, using only a web browser. Scientists can produce maps, timeseries and other plot types. This system completely removes the need for the scientist to understand the technical details of how and where the data are stored. It is based upon open standards, particularly the Open Geospatial Consortium's Web Map Service specification (OGC WMS) and is released as open source software under a liberal (modified BSD) licence (<http://ncwms.sf.net>). Many research institutes, forecast centres and other data providers around the world have installed their own Godiva2 system for data visualization and it is used as the dynamic visualization system for the major European operational oceanography projects MERSEA, ECOOP and MyOcean. The users of these various Godiva2 servers are currently mainly oceanographers (fifteen operational oceanography centres around Europe already provide data for the Godiva2 portal) and the software is gaining traction in the meteorology and climate science communities.

b.1.2 The **LabBlog** (developed at the School of Chemistry at the University of Southampton) is a web-based blogging tool specifically designed for the practising scientist to record, disseminate and evaluate their research. The Blog can also be used as a collaboration tool that allows discussion between colleagues. For open science work the blog uses standard protocols (such as Really Simple Syndication, RSS) to publish its content to the public domain but also contains the necessary access control to keep any private work secure. The blog software is to be released with an open source license and community development will be encouraged.

b.1.3 The National Centre for Earth Observation (**NCEO**) and the National Centre for Atmospheric Science Climate Group (**NCAS-Climate**) are both high-profile interdisciplinary research centres involving numerous universities and institutes around the UK and many international collaborators. Both groups make use of the latest numerical models of the climate and earth system, validated by observations, to simulate the environment and its response to forcings such as an increase in greenhouse gas emissions. Scientists from many disciplines and institutes collaborate to understand the various aspects of these models and assess their strengths and weaknesses.

b.1.4 At the present time, these collaborations take place through face-to-face meetings, the scholarly literature and informal electronic exchanges of emails and documents. All of these methods suffer from serious deficiencies that hamper effective collaboration. For practical reasons, face-to-face meetings can be held only infrequently. The scholarly literature does not yet adequately link scientific results to the source data and thought processes that yielded them, and additionally suffers from a very slow turnaround time. Informal exchanges of electronic information commonly lose vital context; for example, scientists typically exchange static visualizations of data (as GIFs or PostScript plots for example), but the recipient cannot easily access the data behind the visualization, or modify the visualization in any way. Emails are rarely published or preserved adequately for future use. The recent adoption of “off the shelf” Wikis and basic blogs has addressed some of these issues, but does not usually address specific scientific needs or the interactive visualization of data.

b.2 The proposed solution

b.2.1 A Virtual Research Environment is the ideal solution to the above problems. In this project, a VRE will be created rapidly by combining the capabilities of Godiva2 and LabBlog to produce a system known as “SocialGodiva”. Scientists will be able to use Godiva2 to examine output from the latest cutting-edge climate and ocean models; upon finding a feature of interest (perhaps an extreme event, or a suspected problem with the model) the user will be able to create a new blog entry that is linked to the current visualization. Colleagues will be able to provide input through comments and by linking blog entries together. Through semantic and geospatial tagging, scientists will be able to discover colleagues working on similar scientific problems, thereby stimulating new collaborations. The system will be augmented by the addition of a geospatially-enabled database, based on the widely-used open-source PostgreSQL database with the PostGIS extensions. This database will associate blog entries with geographical areas and time periods and allow users to discover discussions that relate to particular areas of interest very efficiently. See Figure 1.

b.2.2 The blog engine will syndicate content using GeoRSS feeds (i.e. RSS feeds that are marked up with their geographical and temporal context, <http://www.georss.org/>) to which users can subscribe. Categories of feeds will initially include: (i) feeds of the discussions initiated by each user; (ii) feeds of the discussions that relate to each dataset; (iii) feeds of discussions that pertain to a set of geographical areas (e.g. ocean basins or continents); (iv) feeds of discussions with given tags (e.g. project names); and (v) a feed that

contains the discussions with the most activity in the previous week (“hot topics”). The textual content of these feeds will be readable by any RSS feed reader (including Google Reader, Firefox “live bookmarks” and FeedReader), but geo-enabled feed readers (including Google Earth and Google Maps) will be able to display discussions on a map, organized by the geographical areas under discussion.

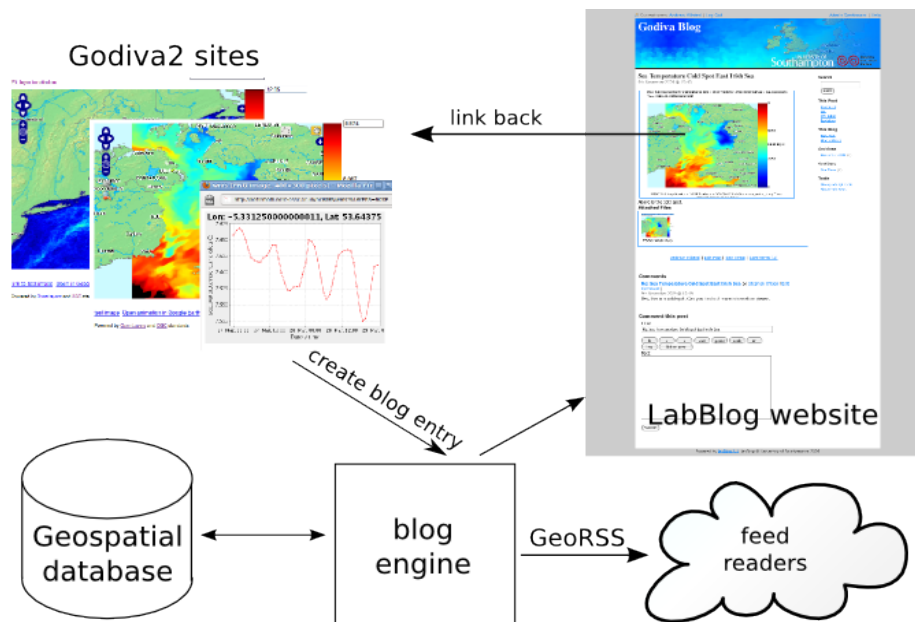


Figure 1: Sketch architecture, including a mock-up of the blog website

b.2.3 The novelty of this project lies in the new capabilities that result from the combination of a number of well-established technologies. The following is a brief illustration of a usage scenario that will be enabled by the new system:

1. A PhD student uses Godiva2 to visualize output from a high-resolution climate model and discovers an unexpected feature that may indicate a problem in the model. He creates a blog entry linked to the visualization and invites others to suggest possible explanations.
2. A colleague, working on a similar geographical region but in a different discipline and institution, discovers the blog entry (perhaps through a search or RSS feed), and realises that the feature can be explained by something known in her field but not in the field of the original researcher. They create an additional visualization of their own data, link it to the original visualization, and post a comment on both results.
3. The Principal Investigator (on a visit to Japan) wishes to monitor current activity in his group. He has subscribed to the RSS feed of “hot topics” and views this feed to discover the main areas of current discussion. He logs on to the blog system to post comments on these discussions to suggest new directions for investigations.

b.2.4 The system will start with basic capabilities, with users able to create blog posts based on map plots. As the project develops we shall add the capability for users to blog about other types of visualization, such as timeseries plots and vertical profiles. Other capabilities will be added based upon user consultation. Tagging and search capabilities could be made much more sophisticated, allowing users to defined their own custom GeoRSS feeds and making use of semantic relationships between terms. Links to relevant publications and presentations could be achieved using techniques developed in previous JISC VRE projects¹. Users of the blog system at the University of Reading could be linked (via open interfaces such as OpenSocial) with the Reading social network of researchers, which is being developed by the JISC VRE project Linksphere (<http://linksphere.org/>), and we have already begun discussions with the Linksphere team with this aim in mind.

b.3 User management and access control

b.3.1 Currently, Godiva2 is an open-access system that does not require a login. However, even for public data, users may wish to keep their blog posts and discussions private to themselves, their group or their institution. Therefore the blog system will implement access control, which will be based upon OpenID authentication. OpenID is an open, decentralized standard for authenticating users, allowing users to log on to different services with the same digital identity. It has been adopted by many relevant projects, including the NERC Data Grid and the JISC VRE project myExperiment (<http://www.myexperiment.org/>), and is being

¹ <http://www.jisc.ac.uk/whatwedo/programmes/vre/insscol.aspx.../vre2/crew.aspx>

evaluated by Linksphere. We shall consult with these projects in order to find the optimum level of security with which users feel comfortable, striking a balance between privacy and complexity.

b.3.2 We shall initially focus on datasets whose visualizations can be made public by agreement of the data provider (this is true for all the current datasets under Godiva2, even when the underlying data are not publicly available). We recognize, however, that many important datasets, particularly those under active research, are not public. Therefore as the project develops we shall link the system to the secure services of the NERC Data Grid, re-using technology from the NERC Technology Proof-of-Concept project MashMyData (coordinated by Dr Blower), which will run concurrently with this project. We have taken care to ensure that the techniques used in this project (e.g. WMS, OpenID) are compatible with those of the NERC data services: see the letter of support from Dr Bryan Lawrence.

c. Quality of Proposal and Robustness of Workplan

c.1 Project plan

c.1.1 The project will start as early as possible in January 2010 and run until October 31 2010. Staff effort will be spread throughout the project on a part-time basis: the extended timescale is to allow time for the user communities to adopt and learn how to use the new technology. The project will be managed by the PI at Reading (Blower). Communication with project partners at Southampton will be through emails and monthly face-to-face meetings. The majority of initial users will be at Reading, and so communication with these users will primarily be through informal face-to-face meetings with the Reading PDRA.

c.1.2 Development of the system will proceed iteratively. A first version of the system based on the HiGEM test case (section d.1.2) will be released as early as possible (by the end of February 2010) in order to promote the project and solicit feedback. Thereafter, there will be a monthly cycle of user consultation and modifications to the system (section b.2.4), with all progress being reported to the blog. New users will be engaged throughout the project as described in section d.1.3. The addition of access to secure data (section b.3.2) will commence at the half-way point of the project; this ensures that there is sufficient time to address this complex issue, but also ensures that there is time for the system to become established and produce results before then.

c.1.3 Initial development and testing will be performed on a development server and, once the system is sufficiently mature, the components will be transferred to existing operational servers at Reading and on the UK National Grid Service, which already hosts a version of the LabBlog engine (see the discussion on sustainability in section c.3.3).

c.2 Risk Analysis

Risk	Probability of occurrence	Impact upon project	Mitigations
Research staff unexpectedly become unavailable	Low/Medium	High	Effort will be redistributed between Reading and Southampton, or staff will be seconded from other projects.
Poor engagement of users	Low	High	Excellent existing contacts with user groups already established at Reading. Strong support from user group leaders. Iterative development and fast cycle time will keep users engaged.
Technical difficulties integrating the components	Low	High	Iterative development starting with simple problems will reveal problems early. Use of well-established and well-understood technologies. Synergies with other projects (e.g. MashMyData) share the burden of developing new technology.

c.3 Intellectual Property, recruitment and sustainability

c.3.1 Intellectual property arising from this project will be owned jointly by the Universities of Reading and Southampton. Resultant source code will be published under a liberal (modified BSD) licence in a public repository such as Sourceforge.

c.3.2 No recruitment of new staff will be necessary for this project. All staff members are already in post and are available for the duration of the project.

c.3.3 We anticipate that the live system will require little maintenance beyond that which is already routinely carried out to sustain the Godiva2 website and blog engine. (A small amount of extra effort will be required to maintain the geospatial database, but we do not expect this to be onerous, given the robust and mature nature of the technologies involved.) We will therefore guarantee to keep the live system running at least a year beyond the end of the project (i.e. until 31 October 2011), and fully intend to keep this running

much further into the future, for as long as it remains in active use. The project website will be updated and kept live for at least a year following project completion and will be archived when it can no longer be maintained. By ensuring compatibility with the NERC data services (see section b.3.2) we ensure that the system will be expandable to include other relevant datasets in future, thereby contributing to its longevity.

d. Engagement with the Community

d.1.1 The new capabilities developed in this project will be applied to scientific problems in NCEO and NCAS-Climate. Both of these centres are coordinated from Reading and involve the Director of the Reading e-Science Centre (Prof. Keith Haines), providing us with excellent links into the daily activities of these groups (see letters of support from the group leaders, Prof. Alan O'Neill and Prof. Rowan Sutton). In addition, the existing Godiva2 user community (consisting mainly of oceanographers in the UK, Europe, the US and Australia) will be engaged.

d.1.2 The first test case will concern NCAS-Climate's high-profile international activities in climate model intercomparison (the "CMIP-5" project, <http://cmip-pcmdi.llnl.gov/cmip5/>). We shall create a new Godiva2 site, linked to the LabBlog engine, for the visualization of climate models, focusing initially on the HiGEM model (<http://www.higem.nerc.ac.uk/>), which is the particular focus of the Reading group. HiGEM is one of the highest resolution climate models in the world, which will fully test the fast interactive visualization capabilities of Godiva2 and provide a great deal of detailed results for VRE interactions and the diagnosis of the model. These discussions will be visible to other climate modelling groups around the world, allowing these groups to comment about how HiGEM compares with other models.

d.1.3 This initial test case will be used to find out more about how users wish to interact with the VRE, and to act as a demonstrator to promote the VRE to other groups. Based on this test case we shall develop and apply the VRE in more areas, selected on the basis of scientific impact, the suitability of the SocialGodiva system to the particular problem and the number of users that will benefit. We shall consult with Prof. Dave de Roure (now the UK e-Social Science Strategic Director and an expert in user engagement) through our partnership in e-Research South in order to apply best practices in user engagement and retention.

d.1.4 Dissemination of project outcomes will occur routinely through the project website and blog, which will be updated weekly. Interim project results will be disseminated through a presentation at the European Geosciences Union conference (Vienna, May 2010), and the final project results will be presented at the GO-ESSP community workshop in October 2010 (venue in the US to be confirmed). A paper will be submitted at the end of the project to a suitable peer-reviewed journal (open access if possible, although such journals are unfortunately rare in environmental informatics). We shall work with JISC to disseminate project results through JISC's own mechanisms, including publication of the project's final report.

d.1.5 The success of the project will be evaluated initially by activity on the SocialGodiva system and by direct user feedback arising from regular consultation. In the longer term (beyond the end of the funded period) it will be possible to evaluate the project through publications that have arisen due to the new collaborations that will have been formed.

e. Budget

e.1.1 The modifications to Godiva2, the setup and maintenance of the geospatial database and the majority of user engagement and interactions will be carried out by a half-time PDRA at Reading. The setting-up and maintenance of the LabBlog will be carried out by a PDRA at Southampton under a sub-contract. In addition, we seek support to cover monthly face-to-face project meetings (10x£50), conference travel to EGU (£800) and GO-ESSP (£1000), travel to two JISC programme meetings for Blower and Frey (4x£100), a development server (£1200) and journal page charges (£400).

e.1.2 The benefits of this project to the lead institution and project partners are difficult to quantify, but will include increased communication between groups within the University of Reading and their many collaborators, an increase in the profile of these groups and an increase in the quality and interdisciplinarity of science that is performed. The system will be open to groups outside Reading and these groups will also be beneficiaries. The benefits to the University of Southampton will include recognition that their e-Science work in the Chemistry domain has great value in other domains also, perhaps creating opportunities for further applications of the LabBlog system.

	01/10 – 07/10	08/10 – 10/10	TOTAL 01/10 – 10/10

Non-Staff

