

Project Acronym:  
Version:  
Contact:  
Date:



## JISC Project Plan

### *Overview of Project*

#### 1. Background

Grasses form the cornerstone of human food production, and are an important focus of scientific research. Among their 11,000 species are most of the world's major food crops, including wheat, barley, oats, rice, maize, millets and sugarcane, and pasture species which feed livestock such as cattle and sheep. Grass species are ubiquitous on Earth, occurring in ecosystems on every continent and, when they form grasslands, they have a major influence on climate through the cycling of carbon and water between soils and the atmosphere. The influence of climate on grasses therefore lies at the heart of several important scientific problems, including the evolution of grasses and grasslands, the effects of climate change on these ecosystems over the coming decades, and the resulting future impacts on food security and biodiversity.

A NESCent (National Evolutionary Synthesis Center) expert working group of leading international scientists was established in 2009 to devise strategies for tackling some of these contemporary research problems. It recommended that the only way to achieve the necessary level of integration of existing knowledge is to develop an ecological informatics facility for all of the world's grasses. This e-Science facility must be built using a multi-disciplinary approach, taking primary information from existing digitised resources, linking and synthesising it within a centralised meta-database. This proposal stems directly from that recommendation and aims to create a world-leading ecological data portal for the grasses. The project has been developed in partnership with The Royal Botanic Gardens, Kew (RBG, Kew), representatives of user groups, and the software development company Knowledge Now Limited.

#### 2. Aims and Objectives

The aims of this project are:

- o To create GrassPortal, an ecological data facility with a diverse range of research and educational applications, by integrating geographical and ecological information for every grass species.
- o To stimulate the engagement of research and educational user-groups with this data facility by developing the functionality and visualisation capabilities of its web interface.

The objectives of this project are:

- A. To extract information from existing digitised data sources on the name and description, geographical distribution, ecological characteristics and evolutionary relationships for all the world's grass species;
- B. To integrate and link these data within GrassPortal, and develop robust mechanisms for data exchange between the various data resources;
- C. To design and implement the web interface of GrassPortal to meet the requirements of, and stimulate engagement with, research and educational users.

#### 3. Overall Approach

**Objective A: extract information from existing digitised content.** Information will be extracted from four existing strands of data:

- i. **Grass species nomenclature and descriptions.** A definitive list of grass species is important

because only 11,000 of an estimated 60,000 grass names are accepted as species, and most are synonyms. This information ('Synonyms'), together with a detailed diagnostic description for every species ('GrassBase'), has been compiled by grass experts in the Herbarium at RBG Kew and will form the 'backbone' of GrassPortal. As part of their IT strategy, Kew are integrating GrassBase with the World Checklist of Plant Families. Since these data are dynamic (updated), an important aspect of the partnership with Kew will involve working with the Herbarium and IT Applications groups to develop web protocols for GrassPortal to extract the necessary data from Synonyms and GrassBase.

ii. **Herbarium archives.** Many of the world's herbaria have digitised large parts of their collections by scanning herbarium sheets, and linking these to information on species and collection locations, thereby generating a large database of species occurrences. Data clumping has been achieved for many (but not all) of the world's digitised herbarium collections by the Global Biodiversity Information Facility (GBIF), which currently contains 3.2 million geo-referenced records for grasses. We will initially focus attention on GBIF and, since these data are dynamic (updated regularly), the GrassPortal system will acquire them through well-established GBIF web services. Building upon this basic structure, further data will be added through links developed with additional archives: so far these include the Kew herbarium, and we are in discussion with representatives of the Institute of Botany, Chinese Academy of Sciences (IBCAS) about their digital archive.

iii. **Global environmental data.** Information on the key determinants of plant geographical distributions, including average climatic conditions, fire regime, soil conditions, and vegetation properties. These data products are in the public domain, and have been compiled by a number of organisations. Since they are static (not updated, but superseded by new data products), we will acquire these data, process them into a standardized format, and store them locally for rapid access by GrassPortal. To satisfy IP requirements, written permission will be needed from the authors of some of these products, and dialogue with these people will be a crucial part of this work phase. We foresee no major issues in this, since all data providers do so for exactly the kind of application we are developing. However, we will aim to minimise risk in this area via three routes: a) beginning to contact authors about permissions early in the project (before Christmas or early in the New Year 2010, once agreement has been reached on the datasets we'll use); b) using our collaboration with Chris Still at UC Santa Barbara, who knows many of these authors personally (this is his research field); and c) identifying alternative data products for each major class of data – in many cases, these are developed from the same raw satellite / other observational datasets, so are basically equivalent.

iv. **Evolutionary relationships between grasses (phylogeny).** The evolutionary relationships between species are inferred by comparing DNA sequences in particular regions of the genome. As a condition of publication in scientific journals, these DNA data must be deposited in the public domain via the database 'Genbank'. Despite significant advances over the past decade, the evolutionary relationships of grasses remain uncertain in key areas, and will only be resolved with more DNA sequence data. In recognition of this fact, our project partner at the University of Lausanne is currently implementing an online system ('GrassWeb'), which regularly searches Genbank for new DNA data, and updates the grass phylogeny as required. Web protocols to exchange data between GrassPortal and GrassWeb will be a central aspect of this collaboration.

**Objective B: integrate data and link within GrassPortal.** These four independent strands of data will be integrated and linked within GrassPortal. The emphasis will be on defining the geographical distribution and ecological characteristics of each grass species, and integrating with evolutionary information. This will be achieved using four steps:

i. **Define geographical distribution.** Use digitised herbarium archives to extract geo-referenced information on the geographical occurrence of each recognised species;

ii. **Extract environmental conditions.** Map the occurrences of each species onto environmental datasets to extract information on the range of conditions that each species experiences;

iii. **Define ecological characteristics.** Model the ecological characteristics of each species by using widely used statistical approaches to summarise the environmental information;

iv. **Develop robust mechanisms for data exchange between the various data resources.** In the case of Synonyms, GrassBase and GrassWeb, this will be achieved through direct collaboration with developers and implementers in Kew and Switzerland. For the other data strands, it will be necessary to develop tools that facilitate interaction between the various data resources.

### **Objective C: make data accessible and enhance user engagement**

A public web portal will be developed for the user community to freely access the data. Investigators at the University of Sheffield are committed to hosting and sustaining this portal and the ecological datasets indefinitely. RBG Kew represents a natural home for GrassPortal, and its data management systems are currently undergoing significant development towards an integrated plant names system across the whole gardens. We will therefore work closely with Kew to ensure compatibility with this ongoing IT strategy, and will offer the portal to them in the longer term. GrassPortal will be designed in direct consultation with research user groups, placing emphasis upon maximising functionality, and enabling users to easily understand and explore the data using visualisation tools.

Design and implementation of the ecological data facility will proceed in an iterative fashion, beginning with a preliminary meeting of stakeholders and systems developers. The aim of this initial meeting will be to gain expert user input on the types of tasks that will be carried out, focusing on requirement building through scenarios and stories. The research users involved in this phase will be grass systematists and IT project managers from RBG Kew, our collaborator from the University of Lausanne, and project partners representing the NESCent expert working group of grass biologists.

After a period of rapid prototyping software development, the web portal and visualisation tools will be tested to evaluate functionality in collaboration with research users from Kew, Lausanne and NESCent, and a wider community of users with interests in climate change, agronomy (crops and biofuels) and conservation. Victoria Uren will lead the evaluation process through studies of user interaction with the tools, and users will be encouraged to apply the data to research case studies, as direct and robust tests of GrassPortal functionality in the real world. A final phase of refinement will then take place, after which the portal will be publicly launched.

### **Critical factors**

- i. Acquisition of global environmental data, and IP consent for its local storage and re-use.
- ii. Development and successful implementation of web services for data acquisition from Synonyms, GrassBase, GrassWeb and GBIF.
- iii. Forging links with a broader community of herbarium managers, and successful linkage to their datasets via web protocols.
- iv. Engagement of research end-user groups for beta testing and system evaluation.

## **4. Project Outputs**

### **4.1. Deliverables**

#### **4.1.1. Content**

*Species-Based Material:* A public web-based system will contain links to species names, taxonomic and plant trait data for every grass species described in the RBG Kew GrassBase and Synonyms databases. For those species included in the GrassWeb system, species names will also be linked to evolutionary relationships dynamically updated by the University of Lausanne.

*Geographical Material:* Each species name will be linked to data describing geographical species occurrences from GBIF and georeferenced parts of the digitized collections belonging to RBG Kew. These georeferenced data will be mapped onto a suite of environmental data (including climate, soils and vegetation) stored locally on the GrassPortal server.

All data will be available to download in standard formats (e.g. text files).

#### **4.1.2. Tools**

The GrassPortal System will provide added value for users by providing tools for: i) searching species

Project Acronym: GrassPortal

Version: 2

Contact: Colin Osborne: email, [c.p.osborne@sheffield.ac.uk](mailto:c.p.osborne@sheffield.ac.uk); tel, 0114 222 0146

Date: 20-01-2010

using taxonomic, geographical, ecological, environmental and phylogenetic criteria; ii) visualizing species distributions and environmental preferences via mapping applications; iii) calculating, summarising, displaying and comparing ecological preferences, range sizes, and biological characteristics for each species; iv) evaluating the quality of herbarium record-based (e.g. GBIF) species data.

#### **4.1.3. User interface**

GrassPortal will feature a contemporary interface that is well designed, both in terms of functionality (ease of use, possession of useful tools, etc.) and aesthetics. The portal must marry the ability to deliver large quantities of data with a visually pleasing and engaging interface.

We will also provide an open web API for each of the new data resources produced by the data portal. The emphasis will be on novel content, especially the data summaries for each species (i.e. 4.1.2-iii and 4.1.2-iv, above). To manage IPR issues, we will avoid "relaying" datasets gathered from other sources and which remain essentially unmodified.

#### **4.1.4. User manuals**

Online manuals will be made available to guide users through the use and capabilities of GrassPortal. The manuals will also provide users with a brief description, and pertinent information related to the relative benefits and disadvantages of, each dataset.

#### **4.1.5. Case Studies**

Online documentation will contain a number of simple case studies, illustrating the breadth of applications covered by GrassPortal, and focusing on research end users. The exact format for these has yet to be decided. At the least, they will be text-and-pictures illustrations of usage, but we will also investigate the possibility of multi-media case studies, using animations and / or podcasts / videos.

#### **4.1.6. Training packages or tutorials**

Resources suitable for "afternoon lab classes", and developed and documented for use in HE teaching in biology, ecology or geography. We envisage these learning exercises and sample data sets being a starting point for the development of further resources in collaboration with end-user groups.

#### **4.1.7. Technical manual**

Instructions for system administration, maintenance and updating will be compiled by K-Now. The manual will contain sufficient detail to allow a computer literate, but non-expert, user to perform the operations needed to allow the portal to be sustained beyond the life of the project.

#### **4.1.8. Events**

Two workshops will be held during the 18 month project period:

i) *Requirements gathering*: The management team, international project partners and Kew IT managers will convene to establish the priorities for functionalities in the portal, constraints on system architecture, data handling and mapping needs. The end result will be a detailed system specification for the design of the portal.

ii) *Evaluation and end-user engagement meeting*: The same group as for the requirements gathering (and possibly other end users) will meet to plan and initiate end-user testing and engagement with the portal. Suitable case studies and scenarios will be developed to evaluate the portal's performance.

#### **4.1.9. Software**

A server based application allowing indexing, storage and faceted retrieval of grass species information using complex query and interaction modalities via a web portal. The software will also provide means to add and normalise grass species data relating to plant traits and ecological preferences for inclusion into the server based system.

#### 4.1.10. License agreements

Exclusive single user license for the server application will be granted to the host of the portal (Sheffield or Kew). This will cover the software only, since the digital content will be completely in the public domain.

#### 4.2. Knowledge and Other Outputs

- i. We will publish one paper in the primary scientific literature describing the basic structure of GrassPortal, the data available, and the benefits of using this system.
- ii. We anticipate that a second paper in the primary scientific literature will evaluate the quality of species occurrence data by comparing GBIF records with the expert range maps for species provided by GrassBase. This paper will highlight geographical biases in GBIF, emphasising areas of the world in which further digitized species records are required.
- iii. We anticipate that a third paper in the primary scientific literature will evaluate the geographical range and ecological preference data for every grass species, examining variation among species, and investigating phylogenetic patterns.
- iv. Beta testing of the system will involve research applications by interested user groups, each of which will potentially lead to research publications. The system evaluation process will also proceed in a scientifically rigorous fashion, and will also be publishable as a piece of research.
- v. The GrassPortal website will have a blog providing users with up-to-date information on the progress and latest happenings in the development of the system. For example, announcements of new data sets or tools will be made in this way. This approach will also be used for engaging with GrassPortal users e.g. it could be a way of advertising for contributors of new data.
- vi. Knowledge about research and development of GrassPortal will be disseminated via presentations at botanical and ecological academic conferences.

### 5. Project Outcomes

GrassPortal will:

- i. provide an integrated set of tools for conducting research on grasses
- ii. help generate new interest in research related to grasses.
- iii. make it easier to conduct ecological, biogeographical and phylogenetic research related to the grasses.
- iv. provide access to enormous quantities of data to researchers all over the world. This will be of particular benefit to researchers with small budgets (especially in the developing world) as GrassPortal will provide them with the opportunity to engage in cutting-edge research.
- v. will be a test case for expanding this sort of platform to other biodiversity data.

### 6. Stakeholder Analysis

**Table 6.1.** Potential stakeholders in GrassPortal, their interest in the project (Getting the project to succeed = Success; or acting as users of the portal = Use), their level of interest in the project (Low to High), and their importance to ensuring the project's success (Low to High).

Stakeholder	Interest	Interest level	Importance
JISC	Success	High	High
K-Now	Success	High	High
Kew	Success & Use	High	High
University of Lausanne	Success & Use	High	High
Other project partners	Success & Use	High	High
University of Sheffield	Success & Use	Intermediate	Intermediate
End users (HE research)	Use	High	High
End users (Industry)	Use	Intermediate	Low
End users (HE education)	Use	Intermediate	Low

Special interest groups e.g. botanical societies	Use	Intermediate	Low
General public	Use	Intermediate	Low

## 7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing	1	5	5	Project manager in place, and committed to the project; Kew recruitment underway for February. Expect no problems in recruiting IT expert at this prestigious institution.
Organisational	1	5	5	Partners fail to deliver. Close working partnerships are being developed through regular meetings and communication. All partners have high interest in and motivation for project's success.
Technical	2	5	10	Data exchange between systems. Web service already available for GBIF, and close partnership with Kew and Lausanne should minimise risk of data exchange issues with the other systems.
Legal	1	5	5	IPR. We will maintain dialogue with JISC to ensure that IPR issues are handled appropriately. Data contributors all have open access policies, so we foresee no major issues.
User engagement	2	3	6	Users may fail to engage with GrassPortal. We are taking steps to manage this risk by targeting high profile end-user groups now via direct contact and publicity materials.

## 8. Standards

Type of project information	Name of standard or specification	Version	IPR	Notes	URL
Earth observations and other environmental data	GeoTIFF	Revision 1.0	Open	Most of the environmental data is provided in GeoTIFF format. However, where this is not the case, we will use the standard the data is supplied in.	<a href="http://trac.osgeo.org/geotiff/">http://trac.osgeo.org/geotiff/</a>
Species' occurrence data	Darwin Core	9 Oct 2009	Open	Darwin Core is a body of standards for the dissemination of information about biological diversity.	<a href="http://www.tdwg.org/standards/450/">http://www.tdwg.org/standards/450/</a>

## 9. Technical Development

Our project partner involved in application development, K-Now, will be following the Rapid Application Development (RAD) methodology for this project. This is a hybrid of prototyping (for early user involvement, increasing chances of acceptability) and incremental development (division of work by criticality and priority with selected capabilities released before any other facilities). This process will lead to the production of an early prototype, which can then be easily steered to focus effort on the most needed areas. The RAD method for development of GrassPortal was chosen because it will allow for the best system in the timeframe and with the resources available.

## 10. Intellectual Property Rights

### 10.1 Data use agreement

We do not seek ownership over any of the existing digital material made available through GrassPortal. For this information, we are likely to use an IPR agreement similar to that of the Global Biodiversity Information Facility (GBIF, <http://data.gbif.org/tutorial/datauseagreement>), which allows for open access to all of the data it provides, given that users acknowledge the owners of the data (third parties) and cite their use of GrassPortal in accessing the data. The data use agreement of GBIF is provided in Appendix C for information.

### 10.2 Third parties

All of the data owned by third parties that is provided by GrassPortal, is freely accessible for non-commercial purposes, provided that proper acknowledgement is made. However, we will seek written permission to re-disseminate the data through GrassPortal. Below is a preliminary outline of data providers:

**Table 10.1.**

Data	Data provider	IPR	URL
<b>Species' occurrence records</b>			
Herbarium specimens	GBIF	Open-access. Proper acknowledgement required.	<a href="http://www.gbif.org/">http://www.gbif.org/</a>
Herbarium specimens	RBG Kew digitised herbarium collection	?	<a href="http://apps.kew.org/herbcat/gotoHomePage.do">http://apps.kew.org/herbcat/gotoHomePage.do</a>
<b>Environmental</b>			
Various climatic factors (e.g. temperature, humidity) & altitude	Climatic Research Unit. University of East Anglia	Open-access. Proper acknowledgement required.	<a href="http://www.cru.uea.ac.uk/cru/data/hrg.htm">http://www.cru.uea.ac.uk/cru/data/hrg.htm</a>
Various climatic factors & altitude	WORLDCLIM	Open-access. Proper acknowledgement required.	<a href="http://www.worldclim.org/current.htm">http://www.worldclim.org/current.htm</a>
Altitude, slope & aspect	International Institute for Applied Systems Analysis (IIASA)	Written permission required.	<a href="http://www.iiasa.ac.at/Research/LUC/luc07/Products-Datasets/global-terrain-slope.html">http://www.iiasa.ac.at/Research/LUC/luc07/Products-Datasets/global-terrain-slope.html</a>

Soil properties (e.g. N concentration)	Global Soil Data Task Group	Open-access. Proper acknowledgement required.	<a href="http://daac.ornl.gov/SOILS/guides/igbp-surfaces.html">http://daac.ornl.gov/SOILS/guides/igbp-surfaces.html</a>
Soil properties	ISRIC-WISE	Open-access. Proper acknowledgement required.	<a href="http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/WISE+spatial+ver.+3.0.htm">http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/WISE+spatial+ver.+3.0.htm</a>
Soil properties	IIASA Harmonized World Soil Database	Written permission required.	<a href="http://www.iiasa.ac.at/Research/LUC/luc07/External-World-soil-database/HTML/index.html?sb=1">http://www.iiasa.ac.at/Research/LUC/luc07/External-World-soil-database/HTML/index.html?sb=1</a>
Tree cover	University of Maryland Land Cover Facility	Open-access. Proper acknowledgement required.	<a href="http://www.landcover.org/data/treecover/">http://www.landcover.org/data/treecover/</a>
Land cover type (e.g. grassland, desert)	University of Maryland Land Cover Facility	Open-access. Proper acknowledgement required.	<a href="http://glcfapp.umiacs.umd.edu:8080/esdi/index.jsp">http://glcfapp.umiacs.umd.edu:8080/esdi/index.jsp</a>
Summed Annual Global Net Primary Production	University of Maryland Land Cover Facility	Open-access. Proper acknowledgement required.	<a href="http://glcfapp.umiacs.umd.edu:8080/esdi/index.jsp">http://glcfapp.umiacs.umd.edu:8080/esdi/index.jsp</a>
Normalised Difference Vegetation Index (NDVI)	European Commission: Global Environment Monitoring	Open-access. Proper acknowledgement required.	<a href="http://bioval.jrc.ec.europa.eu/products/vgt_mosaic_global/download_global_spot_vgt.php">http://bioval.jrc.ec.europa.eu/products/vgt_mosaic_global/download_global_spot_vgt.php</a>
Burnt area - Global VGT burnt area product 2000 -2007 (L3JRC)	European Commission: Global Environment Monitoring	Open-access. Proper acknowledgement required.	<a href="http://bioval.jrc.ec.europa.eu/products/burnt_areas_L3JRC/GlobalBurntAreas2000-2007.php">http://bioval.jrc.ec.europa.eu/products/burnt_areas_L3JRC/GlobalBurntAreas2000-2007.php</a>
Human influence on the natural environment	World Conservation Society and CIESIN	Written permission required.	<a href="http://ciesin.columbia.edu/wild_areas/#">http://ciesin.columbia.edu/wild_areas/#</a>
Density of livestock (cattle, goats & sheep)	UN Food and Agriculture Organisation	Open-access. Proper acknowledgement required.	<a href="http://www.fao.org/ag/againfo/resources/en/glw/GLW_dens.html">http://www.fao.org/ag/againfo/resources/en/glw/GLW_dens.html</a>
IUCN/UNEP World Database On Protected Areas	IUCN and UNEP	Written permission required.	<a href="http://www.wdpa.org/Download.aspx">http://www.wdpa.org/Download.aspx</a>

### 10.3 New content owned by GrassPortal

New digital content generated by GrassPortal, such as geographical range size and ecological preference data, will be freely accessible for non-commercial purposes, provided that proper acknowledgement is made. We will devise an appropriate IPR statement / data use agreement to cover this. We would like help from JISC in formulating this agreement, which will cover both the new digital content generated, and the documentation that will accompany it.

Project Acronym: GrassPortal

Version: 2

Contact: Colin Osborne: email, [c.p.osborne@sheffield.ac.uk](mailto:c.p.osborne@sheffield.ac.uk); tel, 0114 222 0146

Date: 20-01-2010

## ***Project Resources***

### **11. Project Partners**

#### **11.1 University of Sheffield; Animal and Plant Sciences**

Main contact 1: Colin Osborne (Royal Society University Research Fellow)

Contact details: [c.p.osborne@shef.ac.uk](mailto:c.p.osborne@shef.ac.uk)

Role: Project Director

Main contact 2: Vernon Visser (PDRA)

Contact details: [v.visser@shef.ac.uk](mailto:v.visser@shef.ac.uk)

Role: Project Manager

Main contact 3: Rob Freckleton (Professor of Population Biology)

Contact details: [r.freckleton@shef.ac.uk](mailto:r.freckleton@shef.ac.uk)

Role: Geographical range and ecological preference modelling

#### **11.2 University of Sheffield; Computer Science**

Main Contact: Victoria Uren (Senior Research Fellow)

Contact details: [V.Uren@dcs.shef.ac.uk](mailto:V.Uren@dcs.shef.ac.uk)

Role: System functionality and visualization tools. System evaluation with end users.

#### **11.3 Royal Botanical Gardens Kew (RBG Kew)**

Main Contact 1: David Simpson (Assistant Keeper, Systematics, in the Herbarium, Library, Art & Archives Department)

Contact details: [d.simpson@kew.org](mailto:d.simpson@kew.org)

Role: GrassBase, Synonyms, herbarium archives.

Main Contact 2: A.N. Other (RA)

Contact details: [a.n.other@kew.org](mailto:a.n.other@kew.org)

Role: GrassPortal set-up, testing and evaluation.

Main contact 3: Abigail Barker (Science and Horticulture Applications Manager, IT Department)

Contact details: [A.Barker@kew.org](mailto:A.Barker@kew.org)

Role: Ensuring GrassPortal compatibility with Kew IT strategy and system.

#### **11.4 Knowledge Now (K-Now)**

Main Contact: Sam Chapman (Technical Director)

Contact details: [sam.chapman@k-now.co.uk](mailto:sam.chapman@k-now.co.uk)

Role: Applications development and testing.

Page 9 of 21

Document title: JISC Project Plan

Last updated: April 2007

Project Acronym: GrassPortal

Version: 2

Contact: Colin Osborne: email, [c.p.osborne@sheffield.ac.uk](mailto:c.p.osborne@sheffield.ac.uk); tel, 0114 222 0146

Date: 20-01-2010

### **11.5 University of Lausanne**

Main Contact: Nicolas Salamin

Contact details: [Nicolas.Salamin@unil.ch](mailto:Nicolas.Salamin@unil.ch)

Role: GrassWeb evolutionary relationships database.

### **11.6 Brown University**

Main Contact: Erika Edwards (Assistant Professor)

Contact details: [Erika\\_Edwards@brown.edu](mailto:Erika_Edwards@brown.edu)

Role: End-user, representative of NESCent expert working group.

### **11.7 University of Washington**

Main Contact: Caroline Strömberg (Assistant Professor)

Contact details: [caestrom@u.washington.edu](mailto:caestrom@u.washington.edu)

Role: End-user, representative of NESCent expert working group.

## **12. Project Management**

Project Director: Colin Osborne

Project Manager: Vernon Visser (100%)

Management Team: Project Director, Project Manager, Rob Freckleton, Victoria Uren, Sam Chapman, David Simpson, Abigail Barker, in consultation (remotely) with Nicolas Salamin.

Colin Osborne will lead the project and take overall responsibility for its management. All matters will be decided by agreement of the management team, which will meet every two months, with further teleconferencing and regular email as necessary between meetings.

## **13. Programme Support**

The IPR statement, and keeping up-to-date with the latest technological / web advances would seem to be the most useful areas for support at present.

## **14. Budget**

See Appendix A.

## ***Detailed Project Planning***

## **15. Workpackages**

Project start date: 1<sup>st</sup> September 2009

Project completion date: 28<sup>th</sup> February 2011

Duration: 18 months

Phase I. Requirements gathering and initial data sample. The project will start with a requirements gathering phase of three months in which the stakeholders (Sheffield, Kew, research users) will establish the priorities for functionalities in the portal, constraints on the architecture, data handling and mapping needs, etc. A two-day meeting of investigators, project partners, representatives of

Page 10 of 21

Document title: JISC Project Plan

Last updated: April 2007

research user groups, and implementers has been scheduled during this phase. During the same period, the Sheffield project manager will begin to gather an initial data sample for development purposes. Outcomes of this phase will be a detailed system specification and an initial sample of data merged from key resources. These will be used for the initial system implementation.

Phase II. Implementation of portal and data gathering / processing. There will follow a three month phase of implementation by K-Now to produce an initial prototype with basic facilities for search and visualisation. In parallel, the project manager will complete data gathering and begin data processing (environmental niche modelling and assemblage data). These parallel strands must be sufficiently advanced by month 6 to allow the project RA (to be a graduate with experience in informatics based at RBG, Kew) to start populating the portal with the data, and linking to GrassBase. A period of close work between the implementers, the project manager and the RA will follow, in which more sophisticated functions are added to the portal to prepare for a release of the second prototype in month 12. Recruitment of end-users will also begin during this phase, using a dual strategy of targeting high profile organizations (e.g. Energy Biosciences Institute in the USA – biofuels from grasses, National Institute of Agricultural Botany in the UK – crop breeders, Sports Turf Research Institute in the UK, Microsoft Research in the UK – ecological research) through direct personal contacts, and broadcasting advertisements via mailing lists (e.g. evoldir, ecolog-L). The aim will be to build up a group of end-users who can firstly provide feedback on their requirements, and secondly to act as beta testers of the system, and to participate in system evaluation. Victoria Uren will lead a formal evaluation with the aim of producing a publishable piece of academic work (see Appendix D for questionnaire on which this will be based). Ongoing interaction with end-users will ensure the portal meets their requirements. It may be helpful to organize a workshop during Phase II to facilitate this activity.

Phase III. Beta testing and data population. The second prototype will be opened up to a small group of beta testers including end users and project members. End users will be encouraged to develop 'real world' research with the portal, and a three-day meeting has been scheduled during this phase to facilitate engagement with the new data products. A formal evaluation process and feedback from end users will be used to identify bugs in the system. During this phase the portal will continue to be populated. By month 16 a robust, tested portal will be ready to be released for public use.

A full breakdown of these phases into specific work packages is provided in Appendix B.

## 16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Feb 2010	Achievement against technical aims and objectives	Does the prototype meet the initial requirements?  What lessons have we learned?  Are design changes required?	System testing - systematic tests of functionality by developer team and Sheffield RA.	Functionality is robust.
April 2010	System requirements from end users	Does the system meet end user	Formal requirements gathering study	Feedback on system to developers, and update to

		requirements?  What additional / alternative functionality is required?	with end-users.	requirements.
June 2010	Effectiveness of project management	Is progress satisfactory?  What lessons have we learned?	Internal review of requirements vs implemented functionality.	Necessary functionality is in place.
Jul-Aug 2010	Achievement against technical aims and objectives	Does the first version meet the requirements?  Can the system perform the tasks described in the scenarios?  Is the system usable by the expected client base?  What lessons have we learned?  Are design changes required?	Alpha testing - internal acceptance testing of the software by project biologists.	Production of integrated datasets and visualisations of data.  Qualitative feedback on suitability as a research/educational tool and usability.
Aug 2010	Effectiveness of the project management	Have development milestones been met?  Is the representative sample of data in place?  What lessons have we learned?  Is our dissemination effective?	Internal review of requirements vs implemented functionality.  Internal review of the data available in the portal.	Necessary functionality is in place.  Sufficient data is in place
Sept-Nov 2010	Benefits	Does the	Lab based	Satisfactory scores

	Stakeholder engagement	system support students in achieving learning objectives?  Do students engage with the system?  Can the system be used as an educational tool?	learning and assessment with biology students at Sheffield University.  Satisfaction surveys.	in assessed with system.  Student satisfaction.  Educator satisfaction
Dec 2010-Feb 2011	Stakeholder engagement  Robustness of system	Are stakeholders on board?  Is our dissemination effective?  Is the system robust under public use?	Beta-testing - first public use of the system with mechanisms for feedback to the project team and usage monitoring.  Online satisfaction survey.	Usage statistics.  Bug report statistics.  Public user satisfaction.  Feedback
Feb 2011	Benefits  Impact	Have Objectives been met?  Does the system have benefits?  What are the key findings?	Collation of key evaluation outcomes	User satisfaction  Volume / types of data available  Functionality achieved  Educational suitability  Research suitability  Stakeholder engagement

## 18. Dissemination Plan

### 18.1 Project Dissemination Plan

Our overall strategy is to engage with potential end-users from an early stage of the project. This will involve the dissemination of information relating to GrassPortal to potential end-users through various means (discussed in detail below). Additionally, during this process we hope to recruit new data providers and beta testers of the portal. In particular, we hope to attract high profile organisations that

would be willing to allow us to use their organisation's use of the portal for advertising purposes and to elevate the public profile of the project.

Timing	Dissemination Activity	Audience	Purpose	Key Message
10/12/09 - 11/12/09	Workshop 1	<ul style="list-style-type: none"> <li>•Management team</li> <li>•International project partners</li> <li>•Kew IT managers</li> </ul>	The main purpose of this workshop is to decide on the GrassPortal system specification. This is not really intended as an advertising activity, but it will help to raise awareness among the project partners of the portal's capabilities and they will be encouraged to promote the portal within their own institutions and their work contacts.	Establish the priorities for functionalities in the portal, constraints on the architecture, data handling and mapping needs, etc.
1/10/09 -	Fliers / brochures	<ul style="list-style-type: none"> <li>•Potential end-users</li> <li>•HE research and education sectors</li> <li>•Businesses with an interest in the portal's capabilities</li> <li>•Special interest groups (e.g. natural history societies etc.)</li> </ul>	We have designed two advertising brochures: one for a scientific audience and another for the general public. These brochures will be made available on the project website, at conferences and will be included as part of our targeted email campaign (see below). The brochures are meant to stimulate interest in the portal and highlight its capabilities and potential uses. They are also being used to recruit new data providers and beta testers of the portal before its public launch.	<ul style="list-style-type: none"> <li>•Stimulate general interest in the portal</li> <li>•Highlight the portal's capabilities</li> <li>•Recruit new data providers and beta testers</li> </ul>
1/10/09 -	Targeted emails	Potential end-users	We are in the process of identifying potential end-users of the portal and have been and will be emailing these people / organisations to: <ul style="list-style-type: none"> <li>•encourage their interest in the portal</li> <li>•request new data for the portal</li> <li>•recruit beta testers of the portal before its public launch</li> </ul>	<ul style="list-style-type: none"> <li>•Stimulate general interest in the portal</li> <li>•Highlight the portal's capabilities</li> <li>•Recruit new data providers and beta testers</li> </ul>

1/11/09 -	Portal website	<ul style="list-style-type: none"> <li>•Potential end-users</li> <li>•HE research and education sectors</li> <li>•Businesses with an interest in the portal's capabilities</li> <li>•Special interest groups (e.g. natural history societies etc.)</li> </ul>	The website will be used both for advertising purposes and for disseminating information related to the progress of the portal and any other news that may be of potential interest to the eventual end-users	<ul style="list-style-type: none"> <li>•Advertise the utility and capabilities of the portal</li> <li>•Provide detailed information we cannot fit on a brochure</li> <li>•Inform potential users about the progress of the portal</li> </ul>
31/01/10 - 31/03/10	End-user requirement gathering	<ul style="list-style-type: none"> <li>•Research end-users</li> </ul>	A formal end-user requirements gathering study will be carried out. This is designed to evaluate the match of our proposed system to requirements of the research community. However, it will also be a perfect opportunity for disseminating information about the system.	<ul style="list-style-type: none"> <li>•Advertise the utility and capabilities of the portal</li> </ul>
1/1/10 -	Journal articles	HE research community	<ol style="list-style-type: none"> <li>i. The first paper will describe the basic structure of GrassPortal, the data available, and the benefits of using this system.</li> <li>ii. A second paper will evaluate the quality of species occurrence data by comparing GBIF records with the expert range maps for species provided by GrassBase. This paper will highlight geographical biases in GBIF, emphasising areas of the world in which further digitized species records are required.</li> <li>iii. A third paper will evaluate the geographical range and ecological preference data for every grass species, examining variation among species, and investigating phylogenetic patterns.</li> </ol>	<ul style="list-style-type: none"> <li>•Advertise the utility and capabilities of the portal to the HE research community</li> </ul>

1/8/10 – 31/10/10	Evaluation and End-user Engagement meeting	<ul style="list-style-type: none"> <li>•Management team</li> <li>•International project partners</li> <li>•Kew IT managers</li> <li>•Invited end-users</li> </ul>	A second workshop is planned where we will plan and initiate end-user testing and engagement with the portal. Suitable case studies and scenarios will be developed to evaluate the portal's performance. Once again the main purpose of this meeting will not be to advertise the portal, but it will provide an opportunity for all attendees to test the capabilities of the portal and thereby allow them to disseminate information to a wider network of potential end-users.	<ul style="list-style-type: none"> <li>•Initiate portal testing and develop recommendations for changes to the portal</li> <li>•Educate participants as to the capabilities of the portal</li> </ul>
1/7/10 -	Conferences	HE research community	We aim to attend a number of key botanical and ecological conferences in order to promote GrassPortal. We envisage conference advertising will take one of three forms: <ul style="list-style-type: none"> <li>i. Talk</li> <li>ii. Poster</li> <li>iii. Demonstration</li> </ul> Conference demonstrations will involve presenting the capabilities and utility of the GrassPortal in conducting research.	<ul style="list-style-type: none"> <li>•Stimulate general interest in the portal</li> <li>•Highlight the portal's capabilities</li> <li>•Recruit new data providers and beta testers</li> </ul>
1/1/11 – 28/2/11	Press Release	<ul style="list-style-type: none"> <li>•General public</li> <li>•Potential end-users</li> <li>•HE research and education sectors</li> <li>•Businesses with an interest in the portal's capabilities</li> <li>•Special interest groups (e.g. natural history societies etc.)</li> </ul>	Within a few weeks of the public launch of the portal we will make a press release advertising the date of the launch, the portal's capabilities and the significance of this project. By this time we should have a publication in the scientific literature, which will highlight the utility of the portal. In addition, we hope to have a number of high profile end-users who have already used the portal so that we can advertise this fact to the general public.	<ul style="list-style-type: none"> <li>•Inform the public of the date of the portal's launch</li> <li>•Advertise the capabilities of the portal</li> </ul>

1/1/11 – 28/2/11	Adverts to mailing lists and online discussion groups	<ul style="list-style-type: none"> <li>•General public</li> <li>•Potential end-users</li> <li>•HE research and education sectors</li> <li>•Businesses with an interest in the portal's capabilities</li> <li>•Special interest groups (e.g. natural history societies etc.)</li> </ul>	Posting online adverts and sending adverts to mailing lists will help to target more select audiences and increase the advertising exposure of the portal.	<ul style="list-style-type: none"> <li>•Inform the public of the date of the portal's launch</li> <li>•Advertise the capabilities of the portal</li> </ul>
1/1/11 -	Invitations to small groups for demonstrations in Sheffield	<ul style="list-style-type: none"> <li>•Potential end-users</li> <li>•HE research and education sectors</li> <li>•Businesses with an interest in the portal's capabilities</li> </ul>	Small group demonstrations give us the opportunity to give select audiences an in depth look at the use and capabilities of the portal. This will enable us to target select groups of potential end-users as well as act as an advertising tool for potential data providers.	<ul style="list-style-type: none"> <li>•Advertise the capabilities of the portal</li> <li>•Recruit data providers</li> </ul>
1/2/11 -	Launch event	<ul style="list-style-type: none"> <li>•Partner &amp; JISC representatives, including senior management.</li> <li>•HE end user representatives</li> <li>•Businesses with an interest in the portal's capabilities</li> </ul>	Public event at Kew to officially launch GrassPortal.	<ul style="list-style-type: none"> <li>•Advertise the capabilities of the portal</li> <li>•Advertise the project to the general media</li> <li>•Recruit end users</li> </ul>
?	JISC Programme meetings	<ul style="list-style-type: none"> <li>•JISC Programme managers</li> <li>•Other JISC-funded projects</li> </ul>	JISC Programme meetings will allow us to inform JISC and other JISC-funded projects of GrassPortal's progress. It will also be an opportunity to receive advice and test out ideas for the portal. It may also perhaps help create a network of people who are able to spread information about the portal.	<ul style="list-style-type: none"> <li>•Report on the portal's progress</li> <li>•Receive advice</li> <li>•Create a network of advertisers</li> </ul>

## 19. Exit and Sustainability Plans

Essentially the portal is the main output of this project and it is our intention that the portal continues to exist after the lifespan of this project.

Project Output	Action for Take-up & Embedding	Action for Exit
GrassPortal	In order to make GrassPortal a success and ensure it becomes "embedded" within the research	1) GrassPortal will either be hosted by Kew or Colin Osborne after the lifespan

	<p>community we plan to:</p> <ol style="list-style-type: none"> <li>1) Consult all project partners and select end-users about what functionalities they would like from the portal (The requirements gathering meeting will form a major part of this).</li> <li>2) Assemble a large database of grass species geographical occurrences and also a large collection of environmental datasets which can be downloaded from the portal. In addition, we aim to link the portal to a number of external databases that provide information relating to, among others, the evolutionary history of grass species, synonymy of grass species names, grass physical traits, etc.</li> <li>3) Provide tested tools for environmental niche modelling and grass species assemblage analysis that are both easy to use and effective in their purpose.</li> <li>4) Thoroughly test the model through a well-planned evaluation process involving a number of beta testers, case studies and real world research.</li> <li>5) Advertise the portal to a wide audience of potential end-users, not only within the scientific community, but also to the general public, industry and specialist societies.</li> </ol>	<p>of this project. As all the necessary infrastructure will be in place by the time this project ends, we do not envisage there being much difficulty in keeping the portal online.</p> <ol style="list-style-type: none"> <li>2) Copies of all project documentation will be provided to JISC.</li> <li>3) A detailed technical manual describing instructions for system administration, maintenance and updating will be compiled (as described section 4.1). This document will ensure the continued existence of the portal should there be any transfer in the location of where it is hosted.</li> </ol>
--	---	--

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
GrassPortal	<ol style="list-style-type: none"> <li>1) GrassPortal will have very limited hardware requirements so there will be little expense in maintaining the portal.</li> <li>2) We envisage a high usage of the portal by the scientific community, which in itself will justify</li> </ol>	<ol style="list-style-type: none"> <li>1) Kew agrees to host the portal. This will ensure its continued existence due to the high profile of this organisation as well as their interest in the continued existence of the portal.</li> <li>2) The portal is hosted by Colin Osborne at the University of Sheffield. The portal's utility, its success in the research community</li> </ol>	<ol style="list-style-type: none"> <li>1) Ensure compatibility with Kew's IT strategy.</li> <li>2) Kew will decide at a later date whether they will be able to host the portal or not. Regardless of their decision, the portal has the commitment of all parties that it should continue to</li> </ol>

	the continued existence of the portal. 3) GrassPortal has many prospects for expansion (e.g. into community-level data and with user-generated content), which offer potential routes for securing future funding for the portal.	and the partnership with Kew ensures its continued existence 3) Further funding is secured to expand the capabilities of the portal (community-level data of grass species, user-generated content...)	exist and there will be facilities in Sheffield to host the portal should that be necessary. 3) If the portal's capabilities, and the type of data it offers, are to be increased we would need to seek additional funding.
--	--	---	--

## Appendices

### Appendix A. Project Budget

### Appendix B. Workpackages

See file "GrassPortal Appendix B.doc"

### Appendix C. GBIF data use agreement

#### 10.1 GBIF Data Use Agreement

##### Background

The goals and principles of making biodiversity data openly and universally available have been defined in the Memorandum of Understanding on GBIF, paragraph 8 (see the relevant excerpts in the attached Annex).

The Participants who have signed the MoU have expressed their willingness to make biodiversity data available through their nodes to foster scientific research development internationally and to support the public use of these data.

GBIF data sharing should take place within a framework of due attribution.

Therefore, using data available through the GBIF network requires agreeing with the following:

##### 1. Data Use Agreements

- i. The quality and completeness of data cannot be guaranteed. Users employ these data at their own risk.
- ii. Users shall respect restrictions of access to sensitive data.
- iii. In order to make attribution of use for owners of the data possible, the identifier of ownership of data must be retained with every data record.
- iv. Users must publicly acknowledge, in conjunction with the use of the data, the data providers whose biodiversity data they have used. Data providers may require additional attribution of specific collections within their institution.
- v. Users must comply with additional terms and conditions of use set by the data provider. Where these exist they will be available through the metadata associated with the data.

##### 2. Citing Data

Use the following format to cite data retrieved from the GBIF network:

*Biodiversity occurrence data provided by: (Accessed through GBIF Data Portal, [www.gbif.net](http://www.gbif.net), YYYY-MM-DD)*

Project Acronym: GrassPortal

Version: 2

Contact: Colin Osborne: email, [c.p.osborne@sheffield.ac.uk](mailto:c.p.osborne@sheffield.ac.uk); tel, 0114 222 0146

Date: 20-01-2010

For example:

*Biodiversity occurrence data provided by: Field Museum of Natural History, Museum of Vertebrate Zoology, University of Washington Burke Museum, and University of Turku (Accessed through GBIF Data Portal, [www.gbif.net](http://www.gbif.net), 2007-02-22)*

### 3. Definitions

- GBIF Participant: Signatory of the GBIF-establishing Memorandum of Understanding (MoU).
- GBIF Secretariat: Legal entity empowered by the GBIF Participants to enter into contracts, execute the Work Programme, and maintain the central services for the GBIF network.
- GBIF network: The infrastructure consisting of the central services of the GBIF Secretariat, Participant Nodes and data providers. Making data available through GBIF network means registering and advertising the pertinent services via the GBIF central services.
- Node: A data provider designated by a GBIF Participant that maintains a stable computer gateway that makes data available through the GBIF network.
- Participant node: An organisational unit designated by the GBIF Participant to coordinate activities in its domain. It may also provide data.
- Biodiversity data: Primary data on specimens, observations, names, taxonomic concepts, and sites, and other related data on biological diversity.
- Metadata: Data describing the attributes and combinations of biodiversity data.
- Data: Biodiversity data and metadata.
- Data provider: A custodian of data making it technically available. This may or may not be the data owner. If not they will have declared to GBIF that they have permission to make the data available.
- Data sharing: The process of and agreements for making data freely and universally available on the Internet.
- User: Anyone who uses the Internet to access data through the GBIF network.
- Owner of data: The legal entity possessing the right resulting from the act of creating a digital record. The record may be a product derived from another, possibly non-digital product, which may affect the right.
- Sensitive data: Any data that the Node does not want to make available, e.g. precise localities of endangered species.

## Appendix D. Questionnaire for formal study of end-user requirements

- 1) What is your task/job for which you think GrassPortal could assist?
- 2) What methods/resources would you use currently to address these tasks/jobs?
- 3) Given the above methods/resources how would you currently access/query the knowledge within them?
- 4) If looking online for knowledge how do you perform query/exploration?
- 5) Which tools do you currently use alongside existing resources to process, visualise or publish found knowledge?
- 6) Give three specific questions you can foresee wanting answers to, with which you hope GrassPortal could assist?
- 7) Can you prioritise the data sets you'd need to solve your task/jobs? (first most important)
- 8) What is lacking in the ability of existing approaches to see and understand query results?

Project Acronym: GrassPortal

Version: 2

Contact: Colin Osborne: email, [c.p.osborne@sheffield.ac.uk](mailto:c.p.osborne@sheffield.ac.uk); tel, 0114 222 0146

Date: 20-01-2010

9) What devices would you use to access GrassPortal? (Mobile/Desktop/Laptop/PDA/Other)?

10) Can you relate the importance of the following features (1=low, 5=high)

a) textual list of results

b) geographical display (e.g. a map)

c) temporal display (e.g. a timeline)

d) phylogenetic display

e) feature summary of findings

f) comparison of features between findings

g) graphs of results

h) adding notes/comments/tags to findings

i) tag cloud navigation to explore popular subjects