A modelling portal for the UK plant systems biology community

Principal Investigators:

Dr Helen Ougham, Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth, Ceredigion, SY23 3EB. helen.ougham@bbsrc.ac.uk; tel. (+44) 01970 823094

Professor Andrew Millar, Professor of Systems Biology, University of Edinburgh, Biological Sciences (CSBE), Mayfield Road, Edinburgh, EH9 3JH. Andrew.Millar@ed.ac.uk; tel. +44 (0)131 651 3325

Subcontractor:

Dr Robert Muetzelfeldt, Simulistics, 11 Tantallon Place, Edinburgh EH9 1NZ.
r.muetzelfeldt@ed.ac.uk; tel. (+44) 0131 667 8938

Case for support

The objectives of the proposal, which has been produced in direct response to an need identified by UK plant systems biology researchers, are:

- To identify, bring together and develop those crop and other plant models which will be of particular use to the UK plant systems biology community, but which are currently in legacy formats or lacking in user-friendly interfaces

- To represent these models in an existing open-standard declarative XML-based format

- To test the behaviour of the models represented in this format, using existing tools, validating the re-implementation of the models against the output of the original models

- To produce a portal to disseminate the models, and examples of their output, via the Centre for Systems Biology in Edinburgh, in XML format so that they can be used in the programming language and on the computational platform of a scientist's choice

- To enable researchers to share reviews, annotations and enhancements of models

The longer-term objectives are:

- To provide a resource which can be integrated with models covering the whole range of biological scales, from the molecular to the ecological and environmental

- To capitalise on the computational resources of the UK systems biology centres and Grid technology to facilitate accurate modelling of complex data-rich plant-based systems

Description of the resource

The portal will make available, in open-source format, those plant growth models which are identified as most relevant to the needs of the UK plant systems biology and crop science communities. The first priority will be to implement ten such existing models, representing at least one each from the following categories: a temperate monocot, a temperate dicot crop species; Arabidopsis; a bioenergy crop; a plant-soil interaction model; a nutrient uptake and usage model; a highly complex model (50+ parameters)

The models, which are currently available in a range of formats, will be represented in an open-standard declarative XML format. Use of a declarative language makes it possible to display the models in a variety of ways, depending on the needs of user communities, and facilitates model transformation (e.g.

simplifying a complex model). It also provides an efficient mechanism for making much of the model metadata available automatically through its internal structure.

To validate the reimplementation of the models in the declarative XML format, the modelling language Simile will be used. This modelling environment (Muetzelfeldt & Massheder 2003) has been developed to overcome the problems involved in implementing plant and ecological simulation models using conventional programming languages: problems such as the effort and skill needed to program the models, the lack of transparency in models implemented as programs, and the lack of reuseability of models and submodels. It allows both spatial modelling and individual-based modelling. Its visual modelling interface makes it accessible to non-programmers, at the same time allowing models to be largely self-documenting. Simile has been used in a number of international research programmes (see attached CV, R. Muetzelfeldt). Models are saved as a text file in a structured format, making them amenable to sharing with other modelling environments. Simile already provides the ideal environment for representing many crop models and their outputs visually, generating efficient underlying code. The current project will therefore be able to focus primarily on model implementation and dissemination rather than de novo development of methodologies, which would not be cost-effective given the existence of an appropriate tool. However, some enhancements to Simile will be required to accommodate the whole range of models to be studied, and the subcontractor Simulistics under the leadership of Robert Muetzelfeldt will carry out such enhancements, and provide advice and training to the BBSRC-funded scientists employed on this project.

It is not the aim of the proposal to make the models available to the whole community in Simile format, though some end-users may choose to adopt Simile in their own labs and acquire the models in this form. Rather, the models will be provided as XML, which can be run through one of several code-generators, allowing users flexibility in the language they use for implementation and the computational platforms (desk-top PCs, clusters or high-performance computing facilities) on which they run. Annotated examples of typical output will be provided for each model through the portal, together with other relevant links such as publications and contact details for the originators of the models.

It is intended that the plant models available through this portal should form part of the continuum of systems biology modelling capability in the UK. They will therefore be integrated into CSBE's Systems Biology Software Infrastructure (SBSI). Some modification of this infrastructure will be required to integrate the models. Comprehensive support can then be delivered with limited additional funding (one RA). Current Systems Biology Markup Language (SBML) will not express the models curated by this pilot project, so they will not yet be deposited in the international BioModels repository. However, the success of SBML, as a standards-based declarative language for representing models, strengthens the case for doing the same thing for models such as crop growth models that are not expressible in biochemical-pathway terms. CSBE is linked to the SBML and SBGN communities, both directly and via the other UK CISBs and EBI. These links will help to keep the portal's development abreast of international standards development, including plant and trait ontologies, and allow the project staff to provide input to the development of later model exchange standards. CSBE's involvement will ensure that the models are kept abreast of evolving international standards. Within this project, analytical tools available in SBSI will support several methods for parameter optimisation and model analysis, including massively parallel methods using high-performance computers at Edinburgh.

Demand for the resource and support from potential users

This proposal is a direct response to a need identified as a priority by the UK plant systems biology community, particularly Arabidopsis researchers, during a workshop on "Succeeding in Plant Systems Biology" organised by BBSRC and GARNet, the Genomic Arabidopsis Resource Network, in July 2005 (http://www.bbsrc.ac.uk/about/gov/panels/isb/docs/PSB_workshop_report_sept05.pdf). In the course of this meeting, concern was expressed that existing types of models used by systems biologists, which are suitable for representing metabolic pathways and genetic regulatory networks at the

subcellular and cellular level, lacked the properties which would be necessary for modelling the growth of whole plants, and plant responses to the environment. To address this concern, a further meeting ("Interfacing Systems Biology with Crop and Ecosystem Modelling"; Swindon, March 2006) was organised by GARNet Administrator Ruth Bastow and Coordinator Andrew Millar, together with IGER crop scientists Helen Ougham, Sid Thomas and Alan Gay. The objective of this meeting was to bring together researchers from the systems biology, bioinformatics and agro-ecological modelling communities to exchange ideas, enhance awareness of each others' fields, explore synergisms and make recommendations on fruitful future directions. The outcome was a strong consensus that existing crop models, some produced as long ago as the 1960s and 70s, together with more recent approaches such as L-systems modelling of plant morphogenesis, had enormous potential to contribute to plant systems biology. However, it was acknowledged that much UK and international expertise in crop modelling has suffered attrition in the last two decades, and that many of the models developed by these experts were in danger of being lost. Some are in obsolete programming languages; some are available only in printed form; many are restricted to a narrowly-defined range of conditions; and most lack user-friendly interfaces and adequate documentation to allow other users to adapt them. A key recommendation from the workshop was therefore the development of: "a Plant Specific Model Repository - a central resource [of] current plant based models for others to use and test. These models could be open to community review and annotation either via the web or at an annual event such as summer school where models would be tested and validated as part of a training exercise." (http://garnet.arabidopsis.info/Modellers_Meeting_Report_final_v2.pdf)

For such a resource to be of maximum and long-lasting use, it needs to be more than a passive repository. It must be capable of incorporating other emerging plant models; it must adopt appropriate international standards so that the plant and crop-level models are compatible with those at other scales - for example, the biochemical and the ecological - and it must ensure that the models are available through a resource with long-term support, in a "future-proof" non-proprietary format to guarantee their usability long term. The present application addresses all these requirements, and it brings together PIs from leading UK centres of crop science and systems biology, together with a subcontractor specialising in implementing crop, agricultural and ecological models, to develop the resource demanded by the community.

The primary beneficiaries will be BBSRC-funded UK plant scientists wishing to apply systems biology approaches to the study of plant growth and development. Crop scientists and the plant breeding and agricultural industries will also benefit. There is particular potential for research into plant responses to climate change, and to modelling the growth of crops relatively new to the UK economy, such as bioenergy species. Finally, the portal will raise the UK's profile as a centre of excellence in plant and crop modelling, and as an open-access resource it will benefit those in other countries, particularly developing countries, where crop modelling can contribute to sustainable agricultural development.

Strategic relevance to BBSRC

Within the remit of the Bioinformatics and Biological Resources initiative, this application addresses the aims of "Establishment and maintenance of a new resource that would be applicable to a broad BBSRC user community", and "Interfacing and integrating resources to better meet user need" - the latter is considered particularly important by the potential users who have called for the establishment of this resource. It is relevant to the Cross-Committee Priority on Bioinformatics and e-Science; the PMS priority of Integrating Plant Physiology; and the GDB priorities for Post/Functional Genomics which include Access to Facilities and Data. It will support BBSRC's Data Sharing Policy by providing the infrastructure through which crop models and their outputs can be disseminated.

Uniqueness and international context

The best-known international initiative to make crop and ecological models more widely accessible is the Register of Ecological Models (REM; http://dino.wiz.uni-kassel.de/ecobas.html), which despite its name includes many crop models. However REM is a meta-database; it describes the models, their input and output parameters, and references the originators, but does not make the source code available or provide examples of model runs. Nor does it address the current need for crop models to be integrated with other activities in the plant systems biology field. The proposed portal will be developed in collaboration with the maintainers of REM, and will use it as one source of information about existing crop and other relevant models.

Management of the resource

The framework for management of the proposed model portal resource will be established in two stages: 1. *Portal design and construction phase*; 2. *Implementation, dissemination and user interaction phase*. Continuity of advisory and project management structures is desirable and we propose here a common approach from stage 1 through to stage 2.

Strategy

Objectives presented in the present document represent a response to needs identified in community consultation exercises carried out via GARNet, including workshops at Edinburgh and Swindon 2005-2006. The *Advisory Group* (AG) will be drawn from the user community to support the *Project Management Board* (PMB) by assessing progress towards objectives, reviewing/revising direction and goals and advising on project resources. AG will also interface with related national and international initiatives and communities, including GARNet, MONOGRAM, Systems Biology Centres (particularly Nottingham), the UK Crop Modellers Group, and REM.

Resources

The Project Management Board will comprise the PIs, the project leader for the subcontractor, and an IS/IT representative from CSBE. PMB will be responsible for managing the following project resources:

- Financial – budget management, timesheets, reporting for audit
- Human – recruitment, training, terms and conditions, health and safety, compliance with employment policies and legislation (diversity, equal opportunity, flexible working etc)
- Equipment and facilities – procurement, asset management, access charging, accommodation
- IT/IS/IM – hardware, software, security, curation, integration, connection, accessibility, use of licensed software

Operational management

PIs will line-manage project RAs according to established practices of participating institutions.

- IGER: The PI, Dr Helen Ougham, is coordinator of the Informatics and Statistics Platform, with which the project will be associated, and which includes statisticians, multivariate analysts and crop modellers. She has over 20 years' experience of crop science research, including modelling of plant growth, and over 30 years' experience of computational methods. With two IGER colleagues, she was invited by GARNet to organise the workshop on "Interfacing Systems Biology with Crop and Ecosystem Modelling" in which this proposal originated. Staff performance and development is managed through the formal BBSRC probation and PPDR system. Projects within IGER are fully supported by the established Institute financial, HR, training, technical, IS/IT, QA and business office infrastructure.
- Edinburgh: PI Professor Andrew Millar is the co-director of the Centre for Systems Biology Edinburgh, and the coordinator of GARNet. He is the principal author of the recent Systems Biology report, and holds a Research Development Fellowship with the role of coordinating systems biology activities in the UK. His lab is one of the UK's leading centres for systems biology research,including both bench scientists and modellers, and one of its priorities is to develop methodologies which will operate across scales from the molecule to the whole organism and beyond. Staff are managed, and their career development facilitated and monitored, through the formal processes of Edinburgh University.

Quality management

The applicant institutions are compliant with the Joint Code of Practice for research and have robust procedures for internal quality management. The quality management system employed by IGER gained formal ISO 9001:2000 approval in 2006, and is subject to regular external scrutiny by Lloyds Register Quality Assurance – a UKAS accredited organisation.

Milestones, review

- Operational management to achieve milestones will be the responsibility of the PMB.
- Regular review meetings will be held, routinely by video or telecom link but with in-person meetings at least once every six months, in rotation around the sites of the participating institutions. The proceedings of all meetings will be recorded and made available online through the project website.
- Meeting reports will be provided to the AB, who will monitor progress against project expectations and the evolving needs of the user community and make recommendations about modifications or improvements to the workplan.

Communication and outreach

- Online resources: Phase 1 (*Portal design and construction*) project website will be a password-restricted management tool accessible to PMB and AB. It will hold project documentation, including the application, participant details, protocols etc, and outputs such as meeting reports and publications.
- Online resources: Phase 2 (*Implementation*) launches the portal, structured as proposed in the present application. All models acquired from third parties will be made available with full acknowledgement of the originators, including links to relevant publications and to other relevant Web resources. The models will be integrated into CSBE's Systems Biology Software Infrastructure (SBSI). SBSI development is funded by the CISB award to support intracellular models and associated simulation, visualisation and analysis. This infrastructure will ensure long-term accessibility of the outputs of this project.
- Publications. Papers arising from the tools developed in this project and applications of them for integrating models will be submitted for publication to high-impact journals in accordance with the publication policies of the participating institutions and will make full acknowledgement of BBSRC funding support.
- Conferences and workshops. Progress during Phase 1 will be reported at the annual GARNet meeting, to the crop community via the UK Crop Modellers' Group and the cross-Institute programme MONOGRAM, and through the activities of the relevant Systems Biology Centres, particularly Edinburgh and Nottingham. An initial presentation of this project will be given at the September 2007 New Phytologist Symposium on Systems Biology and the Biology of Systems, which follows on from the original GARNet meetings in Edinburgh and Swindon in which the project was conceived.
- Feedback from users and stakeholders will be mediated by AG, interactive online communication, conferences and workshops and specific collaboration with groups and individual modellers with defined applications for the tools developed in this project.
- Science in Society. The project has its origin in a defined community need. AG will have an important role in continuing upstream engagement with the objectives and implementation of the resource. The partner institutions have well developed SiS structures, as has BBSRC and stakeholder groups such as GARNet. Given the degree of public interest in systems and modelling aspects of biology (eg climate change, energy crops, land use and so on), we see many opportunities for promoting the work carried out in this project through a range of media outlets.