Integrating Large Data into Plant Science:
From Big Data to Discovery

Hosted by GARNet and the Exeter Centre for the
Study of the Life Sciences (Egenis)

April 21st-22nd, 2016

Dartington Hall, Totnes, Devon, UK

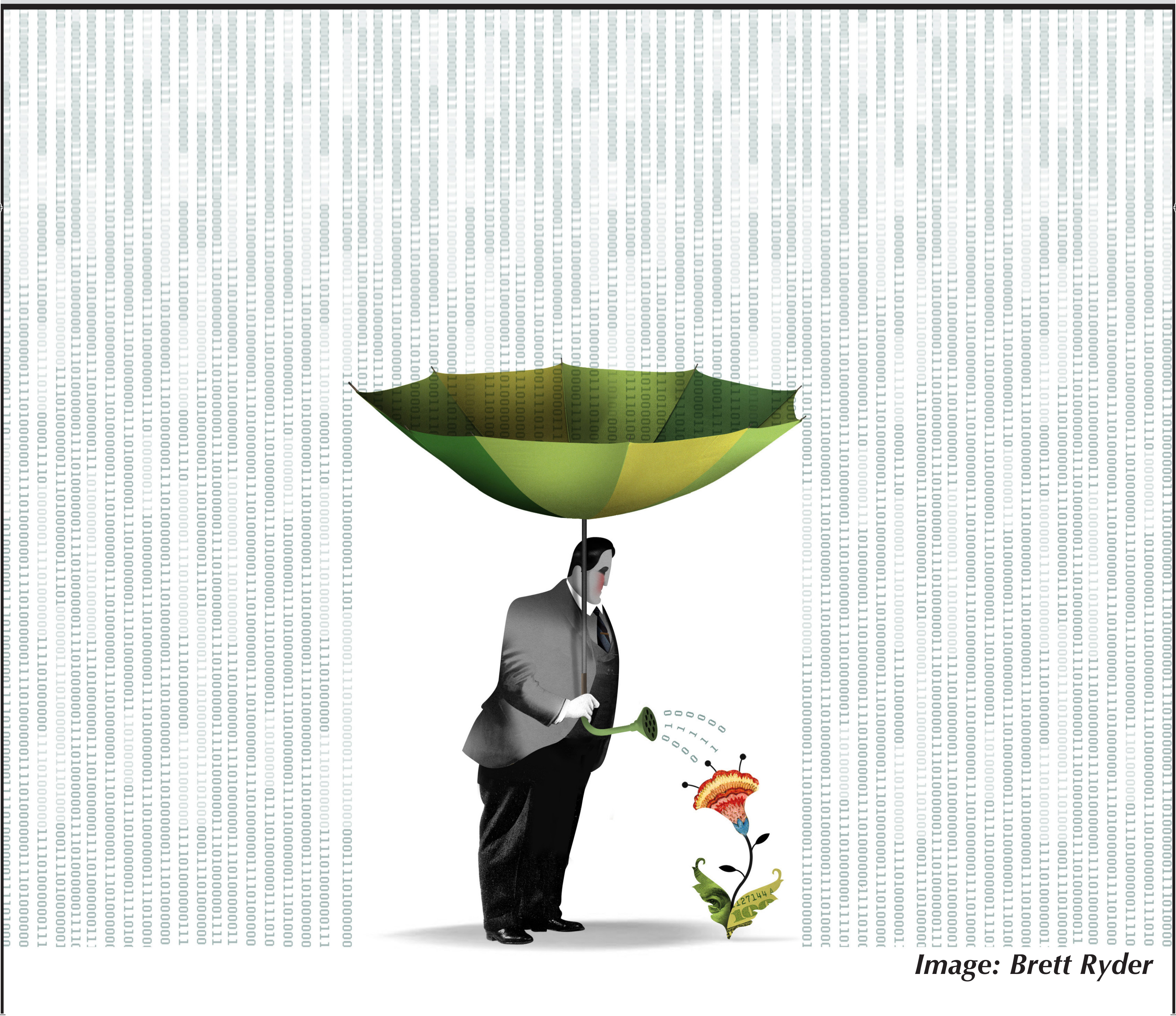


Image: Brett Ryder

Workshop Overview

Integrating Large Data into Plant Science: From Big Data to Discovery

April 21st-22nd 2016

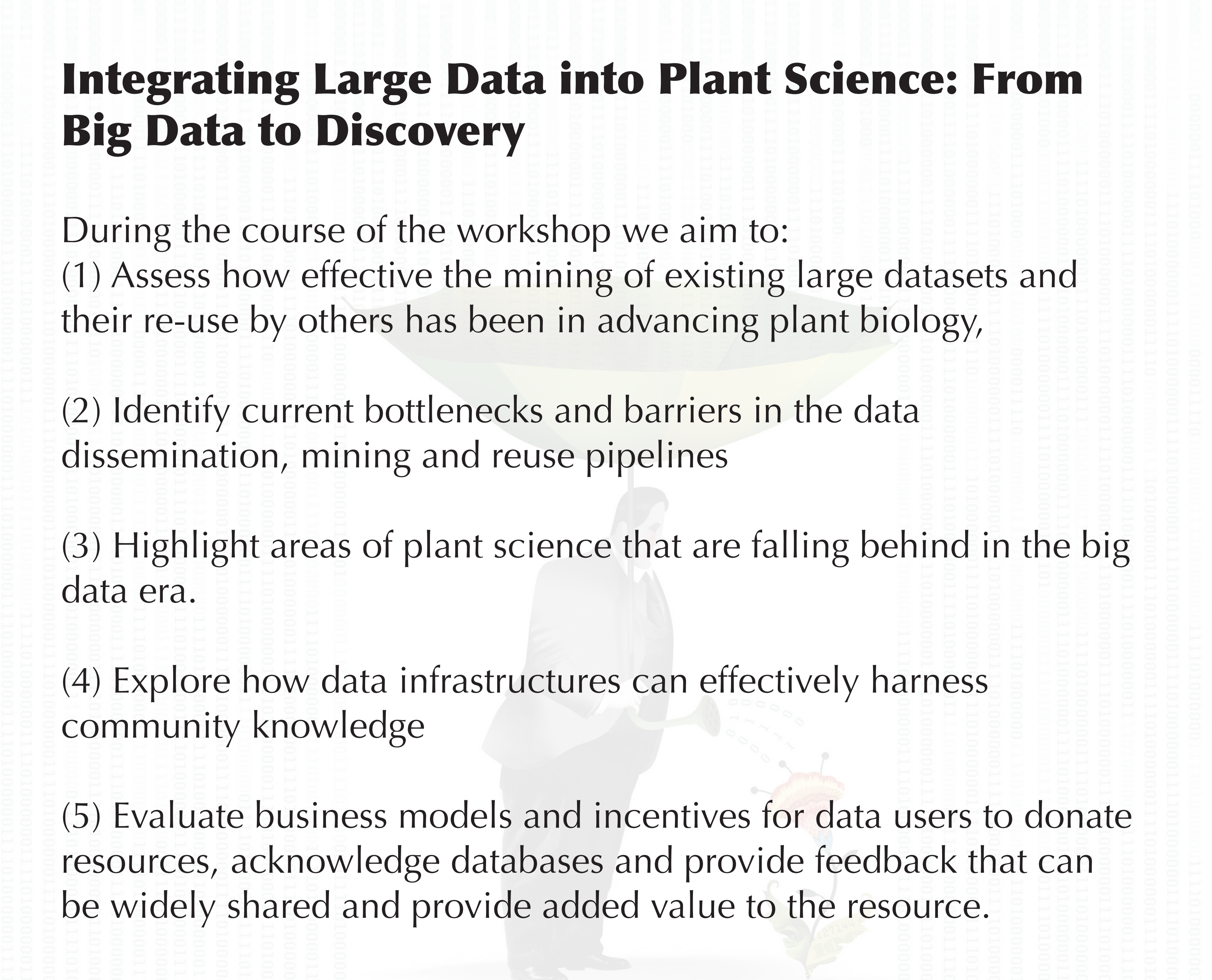
Organisers: Sabina Leonelli, Ruth Bastow, Geraint Parry and David Salt

This workshop brings together prominent biologists, data scientists, database leads, publishers, representatives of learned societies and funding bodies to discuss ways of harnessing and integrating large plant data to foster discovery.

Over the last decade, data infrastructures such as cloud, grids and repositories have garnered attention and funding as crucial tools to facilitate the re-use of existing datasets. This is a complex task, and within plant science a variety of strategies have been developed to collect, combine and mine research data for new purposes.

This workshop aims to review these strategies, identify examples of best practices and successful re-use both within and beyond plant science, and discuss both technical and institutional conditions for effective data mining.

Workshop Overview



Integrating Large Data into Plant Science: From Big Data to Discovery

During the course of the workshop we aim to:
(1) Assess how effective the mining of existing large datasets and their re-use by others has been in advancing plant biology,

(2) Identify current bottlenecks and barriers in the data dissemination, mining and reuse pipelines

(3) Highlight areas of plant science that are falling behind in the big data era.

(4) Explore how data infrastructures can effectively harness community knowledge

(5) Evaluate business models and incentives for data users to donate resources, acknowledge databases and provide feedback that can be widely shared and provide added value to the resource.

The workshop is jointly organised by the Exeter Centre for the Study of Life Sciences (Egenis) and GARNet, with funding from BBSRC and the the European Research Council via Starting Grant "The Epistemology of Data-Intensive Science" (*www.datastudies.eu*).
This is a follow-up from the GARNet-Egenis data sharing workshop exploring how successful data sharing projects work, how data are being integrated and how the process should be improved.
A report from this meeting can be found at:
http://www.garnetcommunity.org.uk/reports

Brief Schedule

THURSDAY 21st April

12:00-12:45:	Lunch
12:45-1:15:	**Introduction: Ruth Bastow & Sabina Leonelli.**
1:15-3:00:	**Session 1: Cases of Data Re-Use.** **Chair: David Salt.**
3:00-3:30:	Coffee break
3:30-5:35:	**Session 2: Data Infrastructures to Foster Re-Use.** **Chair: David Studholme.**
5:45-6:45:	**Panel Discussion 1: Challenges of Re-Use.** **Chair: Geraint Parry**
7:30:	Conference dinner

FRIDAY 22nd April

9:00-11:05:	**Session 3: Integrating Community Knowledge.** **Chair: George LittleJohn**
11:05-11:30:	Coffee break
11:30-12:30:	**Panel Discussion 2: Strategies for Community Involvement.** **Chair: Ruth Bastow.**
12:30-13:30:	Lunch
13:30-15:00:	**Panel Discussion 3: Business models for Data Infrastructures.** **Chair: Sabina Leonelli.**
15:00-15:30:	Concluding Session and Steps Forward.

Detailed Schedule

THURSDAY 21st April

12:00-12:45: Lunch

12:45-1:15: **Introduction: Ruth Bastow & Sabina Leonelli.**

1:15-3:00: **Session 1: Cases of Data Re-Use.**
Chair: David Salt.

Angela Hancock (MFPL, Vienna): *Data integration for evolutionary analysis*

Gordon Simpson (University of Dundee): *What do genomes really encode? Analysing Arabidopsis transcriptomes and epitranscriptomes.*

George Bassel (University of Birmingham): *3D digital single cell analysis*

Dan Bebber (University of Exeter): *Big Data and the global food security debate*

3:00-3:30: Coffee break

3:30-5:35: **Session 2: Data Infrastructures to Foster Re-Use.**
Chair: David Studholme.

Carole Goble (University of Manchester): *FAIRDOM - FAIR asset management and sharing experiences in Systems Biology*

David Salt iHub (University of Aberdeen): *iHUB – An information and collaborative management platform for ionomic research*

Nick Provart (BAR, University of Toronto): *Raising the BAR for hypothesis generation in plant biology using large data sets*

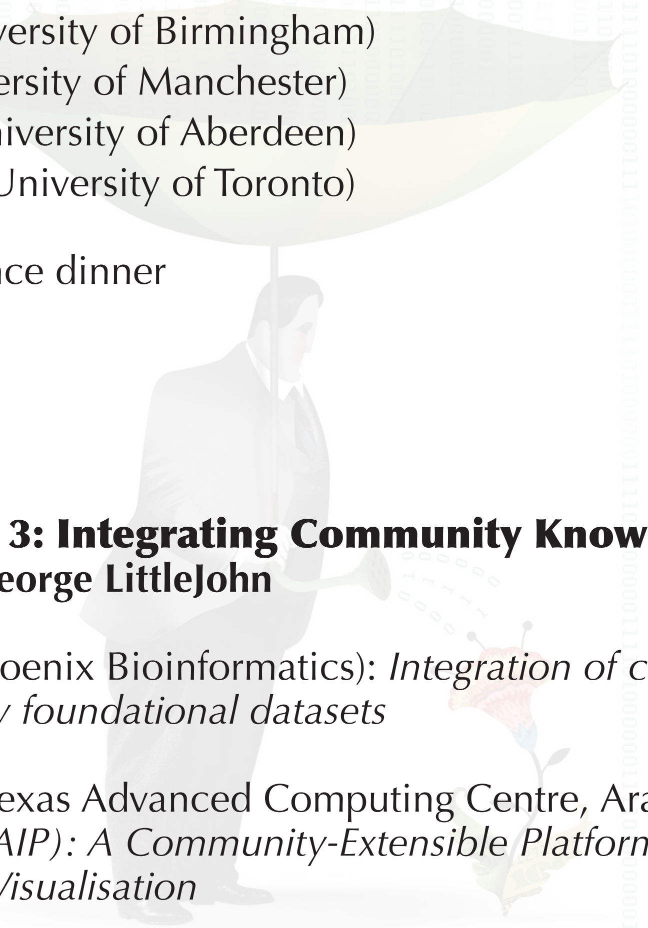
Tomasz Zielinki (BioDare, University of Edinburgh): *Tipping the balance – introducing data management on a centre-wide level*

David Johnson (OERC, University of Oxford): *Data Infrastructures to foster data re-use*

Detailed Schedule

THURSDAY 21st April

5:45-6:45: **Panel Discussion 1: Challenges of Re-Use.**
Chair: Geraint Parry



Panel Members:
Angela Hancock (MFPL, Vienna)
George Bassel (University of Birmingham)
Carole Goble (University of Manchester)
David Salt iHub (University of Aberdeen)
Nick Provart (BAR, University of Toronto)

7:30: Conference dinner

FRIDAY 22nd April

9:00-11:05: **Session 3: Integrating Community Knowledge.**
Chair: George LittleJohn

Eva Huala (TAIR/ Phoenix Bioinformatics): *Integration of community data to produce high quality foundational datasets*

Matthew Vaughn, (Texas Advanced Computing Centre, Araport): *Arabidopsis Information Portal (AIP): A Community-Extensible Platform for Open, Reusable Data and Visualisation*

Elizabeth Arnauld, (CropOntology): *Crop Ontology: traits and variables harmonising field data for meta-analysis*

Georgios Gkoutos (Birmingham University) *Exploring the phenome*

Dan MacLean, (The Sainsbury Lab, Norwich): *Crowdsourcing from scientists: how much bang for very little buck?*

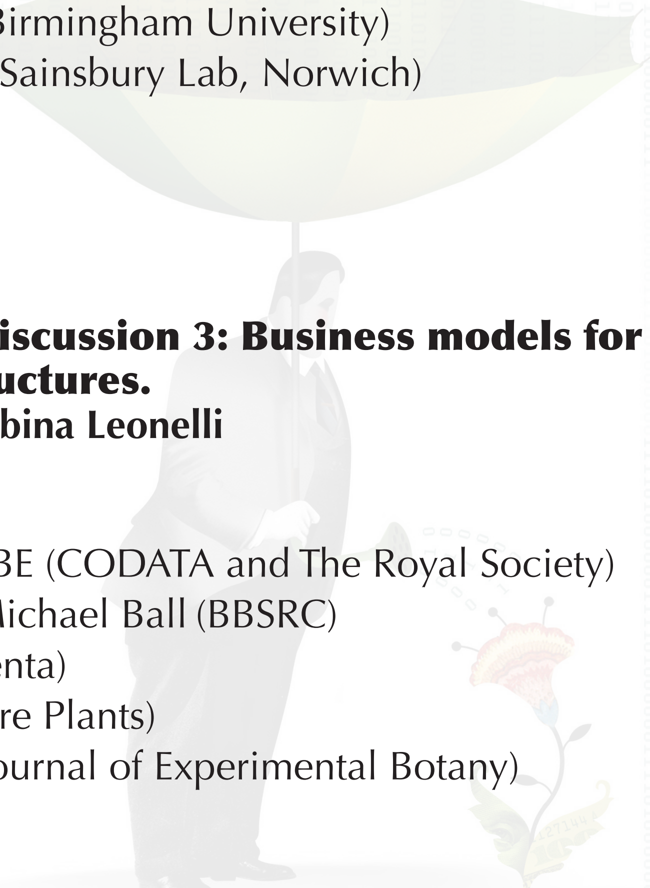
11:05-11:30: Coffee break

Detailed Schedule

11:30-12:30: **Panel Discussion 2: Strategies for Community Involvement.**
Chair: Ruth Bastow.

Panel Members:
Eva Huala (TAIR/ Phoenix Bioinformatics)
Matthew Vaughn, (Texas Advanced Computing Centre, Araport)
Georgios Gkoutos (Birmingham University)
Dan MacLean, (The Sainsbury Lab, Norwich)

12:30-13:30: Lunch



13:30-15:00: **Panel Discussion 3: Business models for Data Infrastructures.**
Chair: Sabina Leonelli

Panel Members:
Geoffrey Boulton OBE (CODATA and The Royal Society)
Rowan McKibbin/ Michael Ball (BBSRC)
Derek Scuffel (Syngenta)
Chris Surridge (Nature Plants)
Rebecca Cunning (Journal of Experimental Botany)

15:00-15:30: Concluding Session and Steps Forward.

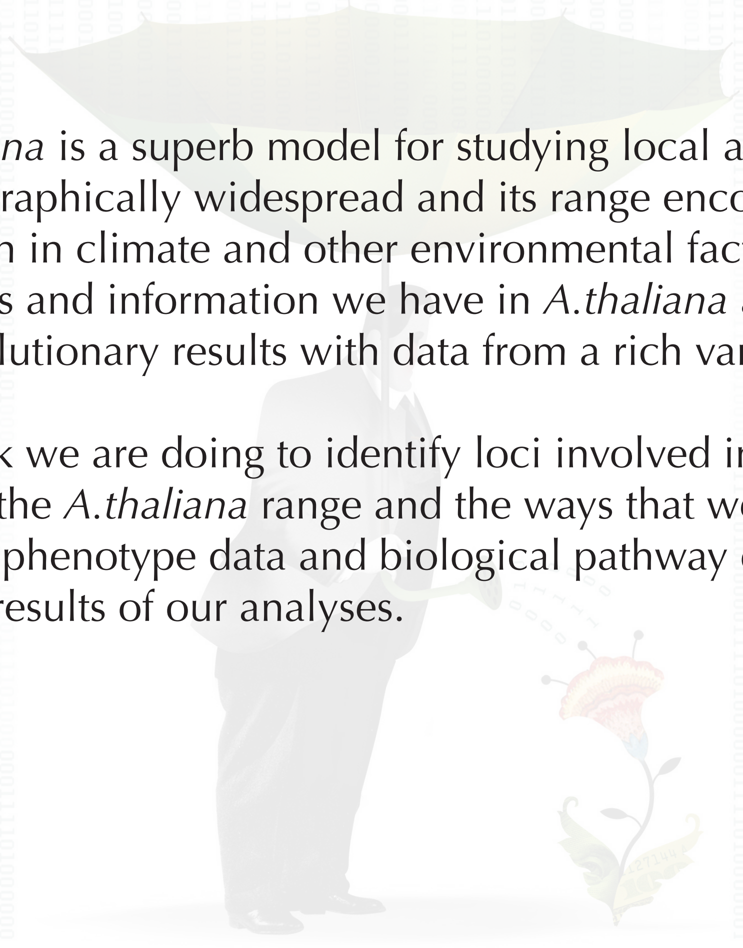
Session 1: Cases of Data Re-Use.

Speaker: Angela Hancock (University of Vienna)

Title: Data integration for evolutionary analysis

Arabidopsis thaliana is a superb model for studying local adaptation because it is geographically widespread and its range encompasses extensive variation in climate and other environmental factors. Moreover the wealth of tools and information we have in *A.thaliana* allows for integration of evolutionary results with data from a rich variety of sources.

I will discuss work we are doing to identify loci involved in local adaptation to climate across the *A.thaliana* range and the ways that we have integrated these results with phenotype data and biological pathway data to validate and interpret the results of our analyses.



Session 1: Cases of Data Re-Use.

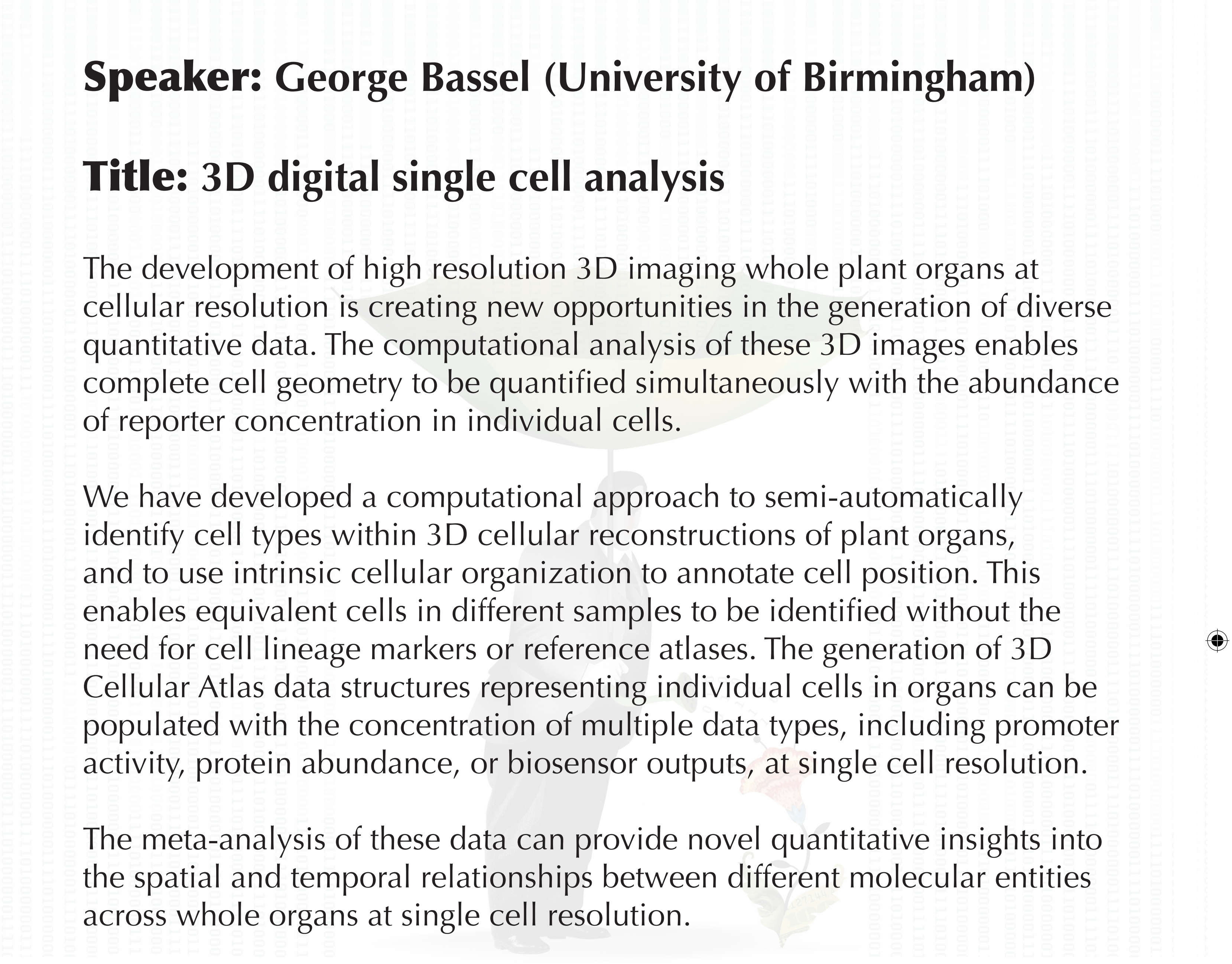
Speaker: Gordon Simpson (James Hutton Institute,
University of Dundee)

Title: What do genomes really encode? Analysing Arabidopsis transcriptomes and epitranscriptomes

Through conventional mutant screens, we identified RNA binding proteins required for normal Arabidopsis development. In order to understand how these RNA binding proteins might function, we used different RNA-Sequencing technologies and a range of approaches to analyse the resulting data obtained from the corresponding mutants. Consequently, we have developed experience of the limitations of the design and analysis of such experiments, and simultaneously provided data valuable for reannotating the Arabidopsis genome. We have used single molecule direct RNA sequencing to map polyadenylated RNA 3′ ends transcriptome wide. These data obviate disadvantages of using reverse transcriptase that include internal priming and strand switching, thus giving authentic stranded data for the first time. We shared these data not only via ENA, but also through our own database "polyAdb" and worked with Araport to include such data as tracks in the Araport 11 genome browser.

In order to understand how to best use RNA-Seq to analyse differential gene expression we examined the performance of different statistical tools to study 48 biological replicates of first yeast RNA-Seq data and more recently Arabidopsis data. We have extended this approach to consider the differential analysis of RNA processing, since alternative transcripts processed from the same gene can have crucially important functions. I will describe our understanding of the difficulties of answering this important question. Finally I will discuss some of our most recent findings working with RNA decay and epitranscriptome mutants that have important implications for how we develop the annotation of the Arabidopsis genome.

Session 1: Cases of Data Re-Use.



Speaker: George Bassel (University of Birmingham)

Title: 3D digital single cell analysis

The development of high resolution 3D imaging whole plant organs at cellular resolution is creating new opportunities in the generation of diverse quantitative data. The computational analysis of these 3D images enables complete cell geometry to be quantified simultaneously with the abundance of reporter concentration in individual cells.

We have developed a computational approach to semi-automatically identify cell types within 3D cellular reconstructions of plant organs, and to use intrinsic cellular organization to annotate cell position. This enables equivalent cells in different samples to be identified without the need for cell lineage markers or reference atlases. The generation of 3D Cellular Atlas data structures representing individual cells in organs can be populated with the concentration of multiple data types, including promoter activity, protein abundance, or biosensor outputs, at single cell resolution.

The meta-analysis of these data can provide novel quantitative insights into the spatial and temporal relationships between different molecular entities across whole organs at single cell resolution.

Session 1: Cases of Data Re-Use.

Speaker: Dan Bebber (University of Exeter)

Title: Big data and the global food security debate

The dimensions of global food security are increasingly measurable. National production and consumption data maintained by FAOSTAT, the statistical division of the UN FAO, form the foundation of numerous academic analyses of changing patterns of the food availability.

Since the launch of Landsat in 1978, remote sensing of climatic and plant physiological parameters have allowed researchers to model the relationships between the Earth's physical state and the socioeconomic responses of agriculture and the food system.

I will provide two examples of how Big Data, in this case meaning comprehensive, regional-to-global scale datasets, can be used to increase understanding of one component of global food security, namely, the impacts of crop pests and pathogens.

Firstly, I will discuss the use of historical observational datasets in understanding the changing distributions of crop pests and pathogens.

Secondly, I will show has climatic reanalysis data can be used to model the epidemiology of crop pathogens. Importantly, I will highlight the limitations of such datasets and suggest pathways to improving data accuracy and precision.

Session 2: Data Infrastructures to Foster Re-Use.

Speaker: Carole Goble CBE (University of Manchester)

Title: FAIRDOM - FAIR asset management and sharing experiences in systems biology

Over the past 5 years we have seen a change in expectations for the management of all the outcomes of research – that is the "assets" of data, models, codes and SOPs. Don't stop reading. Yes, data management isn't likely to win anyone a Nobel prize. But publications should be supported and accompanied by data, methods, procedures, etc. to assure reproducibility of results. Funding agencies expect data (and increasingly software) management retention and access plans as part of the proposal process for projects to be funded. Journals are raising their expectations of the availability of data and codes for pre- and post- publication. The multi-component, multi-disciplinary nature of Systems Biology demands the interlinking and exchange of assets and the systematic recording of metadata for their interpretation.

The FAIRDOM (Findable, Accessible, Interoperable, Reusable Data, Operations and Models) Initiative has 8 years of experience of asset sharing and data infrastructure ranging across European programmes (SysMO and EraSysAPP ERANets), national initiatives (de.NBI, German Virtual Liver Network, UK SynBio centres) and PI's labs . It aims to support Systems Biology researchers with data and model management, with an emphasis on standards smuggled in by stealth and sensitivity to asset sharing and credit anxiety.

This talk will use the FAIRDOM Initiative to discuss the FAIR management of data, SOPs, and models for Sys Bio, highlighting the challenges of and approaches to sharing, credit, citation and asset infrastructures in practice. I will also highlight recent experiments in affecting sharing using behavioural interventions.

http://www.fair-dom.org

http://www.fairdomhub.org

http://www.seek4science.org

Session 2: Data Infrastructures to Foster Re-Use.

Speaker: David Salt (University of Aberdeen)

Title: iHUB – An information and collaborative management platform for ionomic research

The iHUB (*www.ionomicshub.org*) is a collaborative workspace with components that integrate both e-research infrastructure and human interactions to maximise both community access to ionomic resources, and knowledge extraction from these resources. The iHUB facilitates open communication and exchange of information, and lowers the barrier for entry into the field of ionomics (the study of the elemental composition of an organism), allowing a diverse set of researchers from developed and developing nations to participate in advancing the field. The information obtained from the discoveries facilitated by the iHUB will be valuable from the point of view of improving crop yields, enhancing agricultural sustainability, and improving food quality for human nutrition.

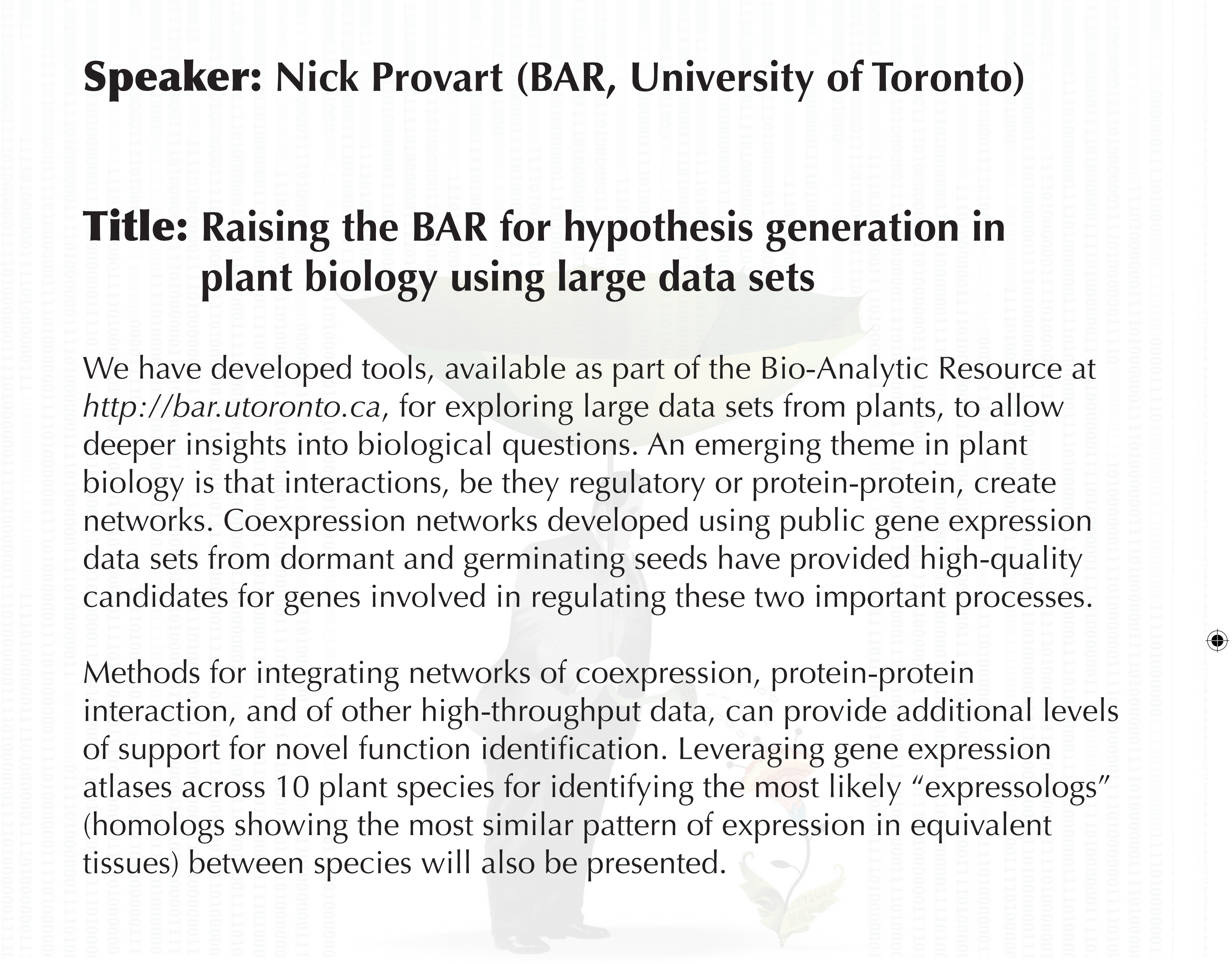
The iHUB has been running since 2007 and has had 73,882 session with 15,050 users from 2,128 cities in 122 countries world-wide, and currently contains ionomic information available to the public on over 280,000 samples of Arabidopsis thaliana, rice and yeast.

The iHUB provides an integrated laboratory information management system (LIMS) supporting the formalised collection of ionomic data from multiple laboratories. The iHUB data infrastructure is built on a relational database and underpinned by a controlled vocabulary. The iHUB contains data discovery, data display, and data export features. Data from different experiments can be aggregated into ‘datasets’ which can be shared with collaborators and associated with Digital Object Identifiers (DOI) for citation in publications. Further, the iHUB Blog, Publications and Datastore features facilitate discussion and the sharing of publications and data files. The iHUB also contains the Ionomics Atlas, a tool for mapping ionomic phenotypes and associated genotypes to the landscape. International iHUB workshops have been held where users discuss future developments of the iHUB and communicate new research findings.

Session 2: Data Infrastructures to Foster Re-Use.

Speaker: Nick Provart (BAR, University of Toronto)

Title: Raising the BAR for hypothesis generation in plant biology using large data sets



We have developed tools, available as part of the Bio-Analytic Resource at *http://bar.utoronto.ca*, for exploring large data sets from plants, to allow deeper insights into biological questions. An emerging theme in plant biology is that interactions, be they regulatory or protein-protein, create networks. Coexpression networks developed using public gene expression data sets from dormant and germinating seeds have provided high-quality candidates for genes involved in regulating these two important processes.

Methods for integrating networks of coexpression, protein-protein interaction, and of other high-throughput data, can provide additional levels of support for novel function identification. Leveraging gene expression atlases across 10 plant species for identifying the most likely "expressologs" (homologs showing the most similar pattern of expression in equivalent tissues) between species will also be presented.

Speaker: Tomasz Zielinki (BioDare, University of Edinburgh)

Title: Tipping the balance – introducing data management on a centre-wide level

21st century science is governed by Data-Intensive Scientific Discovery in which data allows integration and reinforcement between theory, experimentation and simulation. The importance of data on its own, apart from its role as the support for research findings, has been also acknowledged by funding bodies. All data generated from publicly funded research are required to be widely accessible to allow their further dissemination and their re-use outside the initial scope of the studies.

The value of data depends strongly on its comprehensive description: the existence of metadata. However, the annotation process is a laborious one, it is an extra burden for scientists and the main limiting factor which prevents wide-spread data deposition in suitable repositories. We address this problem by adding extra value to data management systems (for example data visualization or pre-processing) in order to provide benefits for the data producers and not only for the data consumers.

In this talk, I will present our experience with adoption within the centre of data management systems for biological data: BioDare, SEEK and OpenBIS. The alternative platforms are based on different philosophies and have complementary strengths, which could cater for the diverse needs of the SynthSyS centre.

Session 2: Data Infrastructures to Foster Re-Use.

Speaker: David Johnson (OERC, University of Oxford)

Title: Data infrastructures to foster data re-use

Enabling data re-use is highly desirable to make best of resources in gathering and analysing data, but is now essential in carrying out research to satisfy recent changes in grant requirements. To support these new demands, a growing number of efforts have been set up through national and international infrastructures for data archiving: for example the UK Data Archive for social science data and specialised public databases for biological data such as the European Nucleotide Archive at the EMBL-EBI. Publishers too have shown commitment, enabling data re-use by accepting data publications, for example NPG's Scientific Data and BMC's GigaScience journals.

At Oxford, we have been leading efforts to foster data re-use in the life sciences by developing a metadata tracking framework, ISA, which supports standards-compliant descriptions of functional multi-omics experiments and is being extended to support high-throughput plant phenotyping, and BioSharing, a data standards, policies and databases registry.

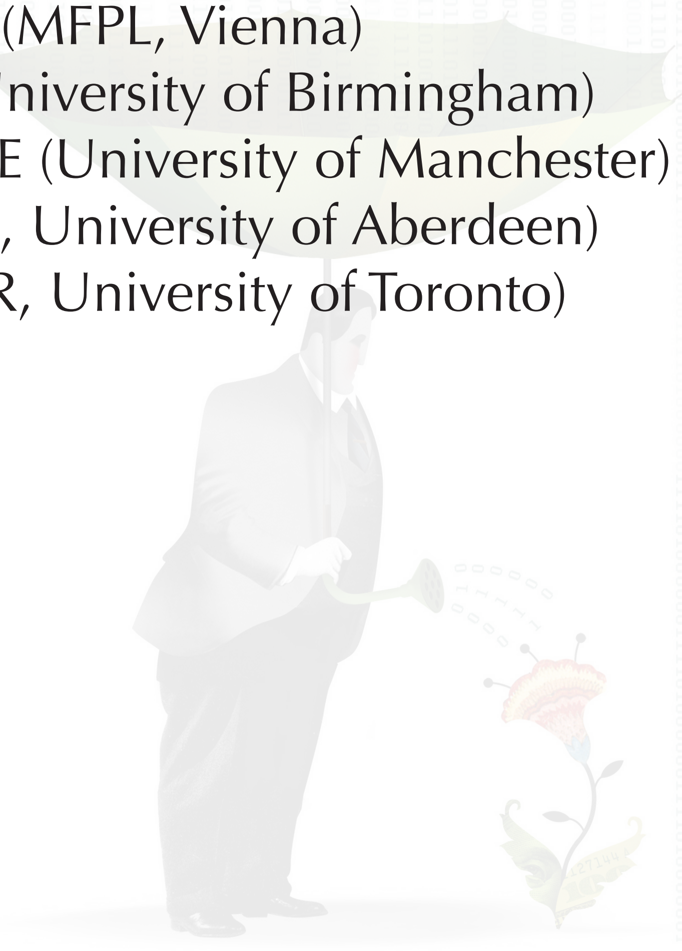
In this talk, I will provide an overview of ISA and BioSharing, and recent contributions to the BBSRC-funded open 'omics data platform for the plant sciences, COPO.

Panel Discussion 1: Challenges of Re-Use

Chair: Gerant Parry

Panel Members:

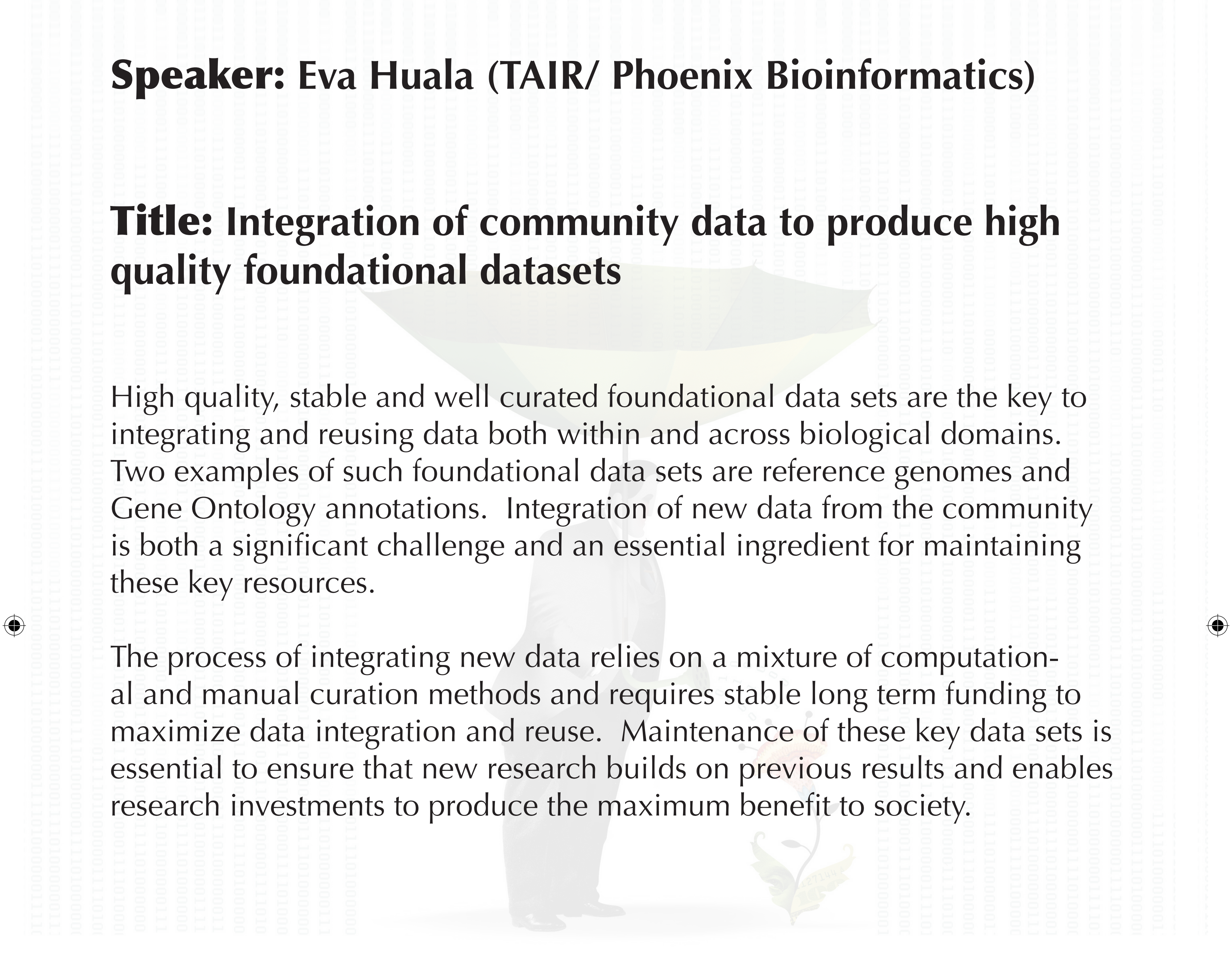
Angela Hancock (MFPL, Vienna)
George Bassel (University of Birmingham)
Carole Goble CBE (University of Manchester)
David Salt (iHUB, University of Aberdeen)
Nick Provart (BAR, University of Toronto)



Session 3: Integrating Community Knowledge.

Speaker: Eva Huala (TAIR/ Phoenix Bioinformatics)

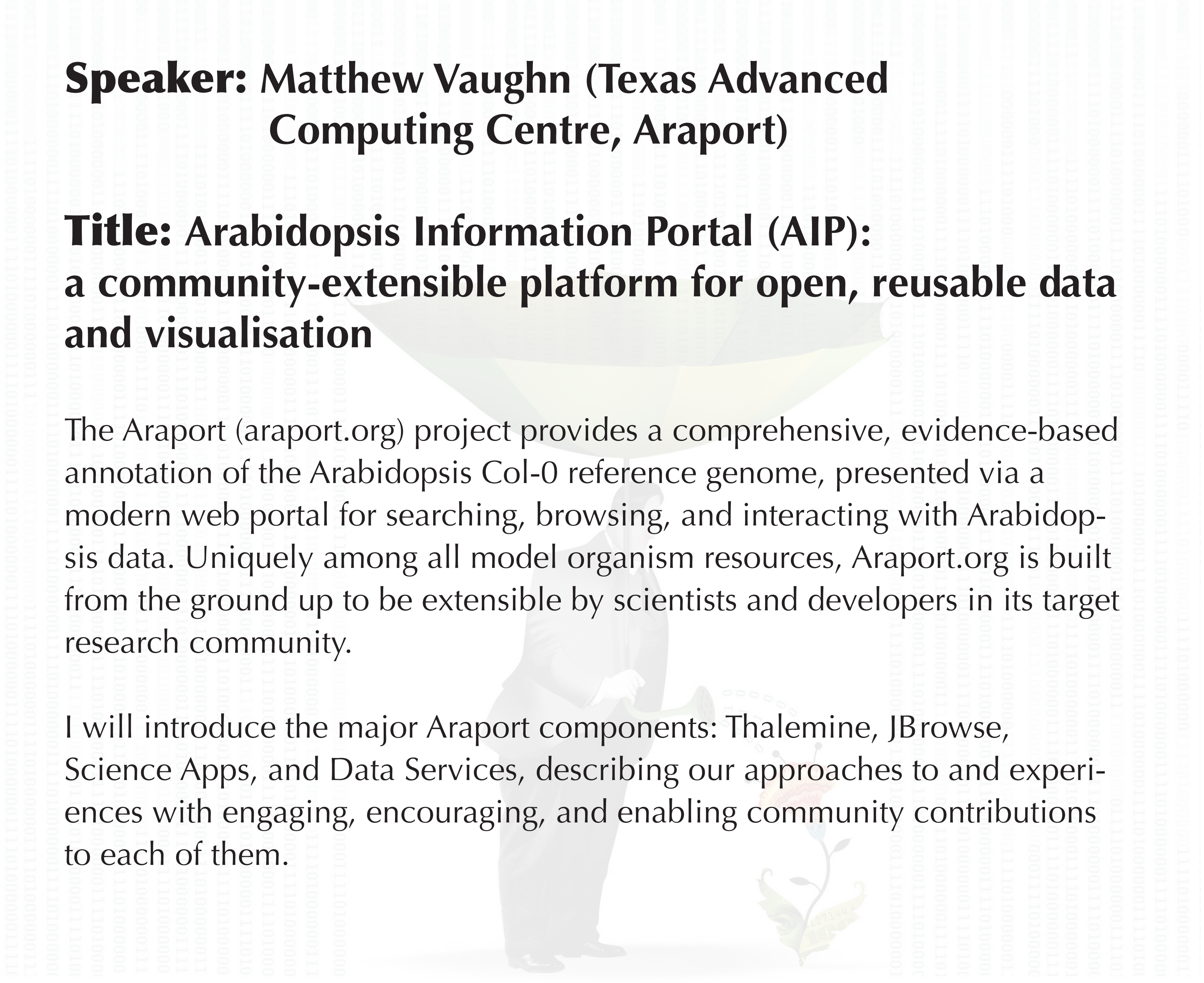
Title: Integration of community data to produce high quality foundational datasets



High quality, stable and well curated foundational data sets are the key to integrating and reusing data both within and across biological domains. Two examples of such foundational data sets are reference genomes and Gene Ontology annotations. Integration of new data from the community is both a significant challenge and an essential ingredient for maintaining these key resources.

The process of integrating new data relies on a mixture of computational and manual curation methods and requires stable long term funding to maximize data integration and reuse. Maintenance of these key data sets is essential to ensure that new research builds on previous results and enables research investments to produce the maximum benefit to society.

Session 3: Integrating Community Knowledge.



Speaker: Matthew Vaughn (Texas Advanced
Computing Centre, Araport)

Title: Arabidopsis Information Portal (AIP): a community-extensible platform for open, reusable data and visualisation

The Araport (araport.org) project provides a comprehensive, evidence-based annotation of the Arabidopsis Col-0 reference genome, presented via a modern web portal for searching, browsing, and interacting with Arabidopsis data. Uniquely among all model organism resources, Araport.org is built from the ground up to be extensible by scientists and developers in its target research community.

I will introduce the major Araport components: Thalemine, JBrowse, Science Apps, and Data Services, describing our approaches to and experiences with engaging, encouraging, and enabling community contributions to each of them.

Session 3: Integrating Community Knowledge.

Speaker: Elizabeth Arnauld (CropOntology)

Title: Crop Ontology: traits and variables harmonising field data for meta-analysis

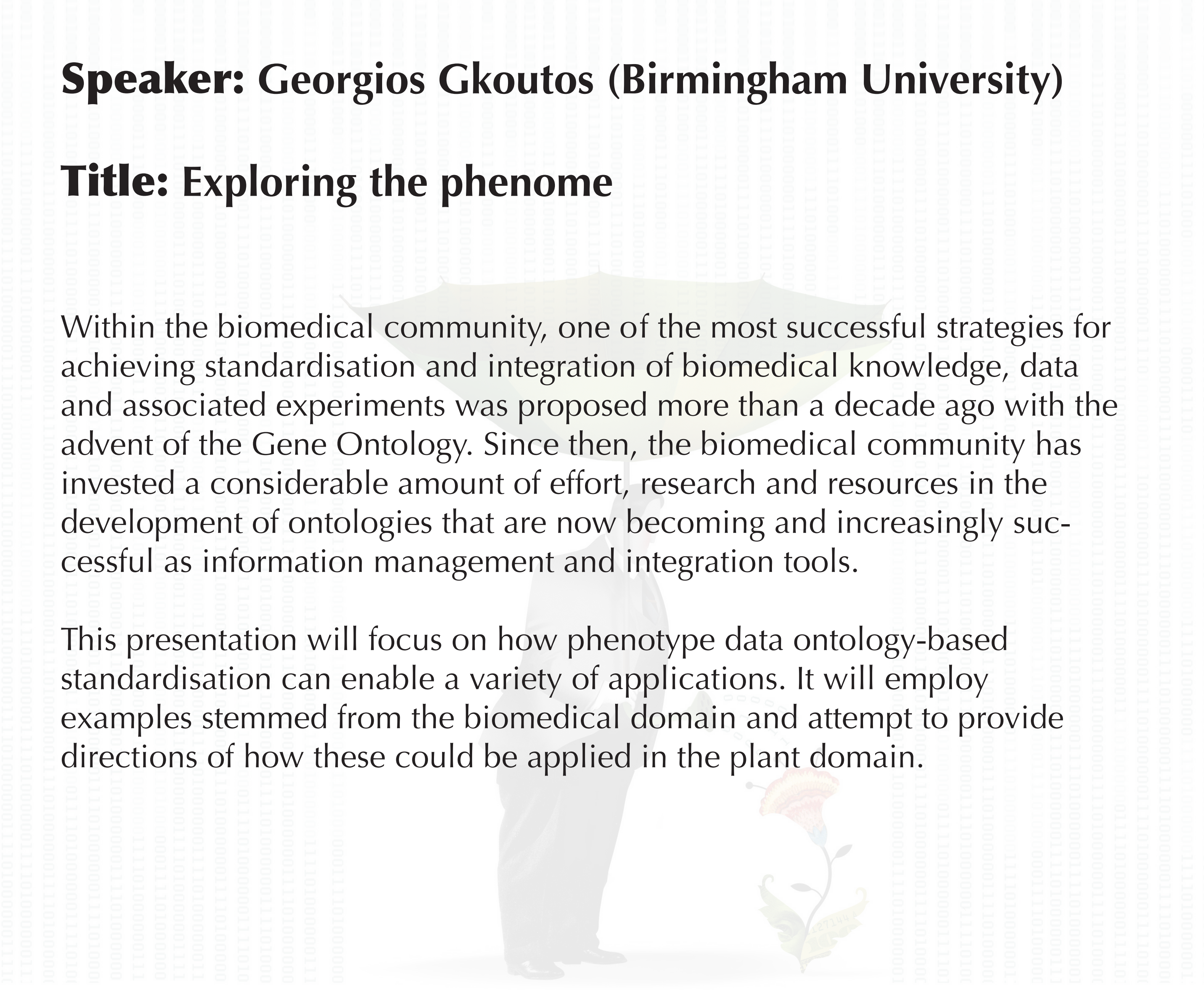
The Crop Ontology is a service of the Integrated Breeding Platform (*www.integratedbreeding.net*) in collaboration with the CGIAR and partners and under the leadership of Bioversity international. The Crop Ontology (*www.cropontology.org*) provides harmonised and validated breeders trait names, measurement methods, scales for currently 18 crops that are used by the Breeding Management System (BMS). Standard Trait Dictionaries (TDs) are a required for the Integrated Breeding Platform (IBP) to facilitate data exchange across teams and enable meta-data analysis. IBP Trait Dictionary now includes the Œstandard variable. A standard variable is equal to ‘one trait+one method+one variable’ and a trait can be measured through different variables, according to the method or the scale used. These variables will accurately annotate the measurements stored in the BMS databases and also will support the creation of standard manual or electronic fieldbooks.

The Crop Ontology contributes to the content of the reference ontologies of the Planteome project (*http://www.planteome.org/*). The Next Generation Breeding Databases developed by Boyce Thompson Institute (US) also embed the Crop Ontology traits. Using similar methodology, an Agronomy Ontology is being developed to support combining results of field management practices with crop traits which is important to fully understand the dynamic of varying factors within any cropping system. Curation is currently performed to secure the compliance between the Agronomy ontology and the variables of the International Consortium for Agricultural Systems Applications (ICASA).

Session 3: Integrating Community Knowledge.

Speaker: Georgios Gkoutos (Birmingham University)

Title: Exploring the phenome

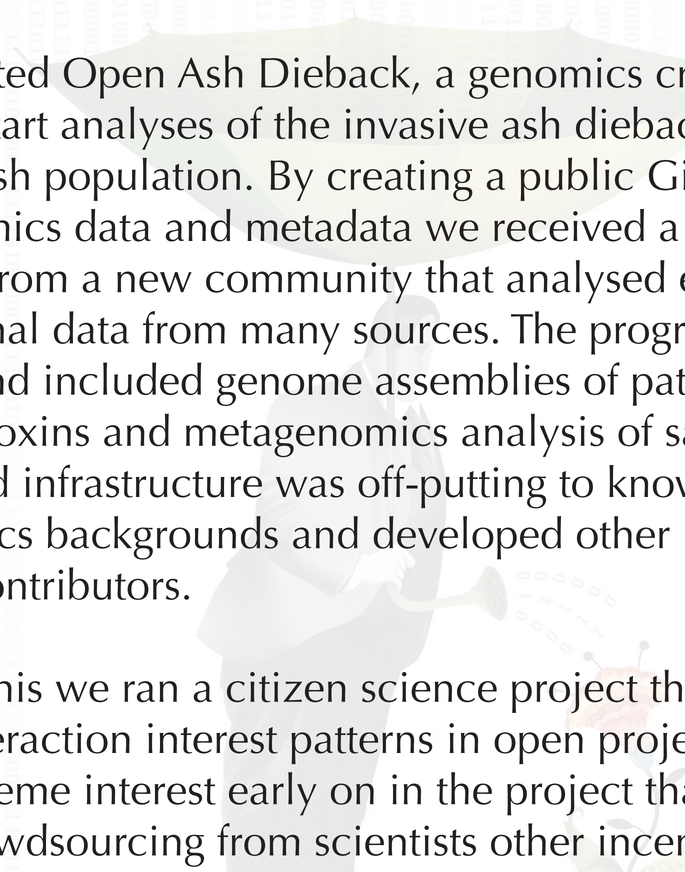


Within the biomedical community, one of the most successful strategies for achieving standardisation and integration of biomedical knowledge, data and associated experiments was proposed more than a decade ago with the advent of the Gene Ontology. Since then, the biomedical community has invested a considerable amount of effort, research and resources in the development of ontologies that are now becoming and increasingly successful as information management and integration tools.

This presentation will focus on how phenotype data ontology-based standardisation can enable a variety of applications. It will employ examples stemmed from the biomedical domain and attempt to provide directions of how these could be applied in the plant domain.

Speaker: Dan MacLean (The Sainsbury Lab, Norwich)

Title: Crowdsourcing from scientists: how much bang for very little buck?



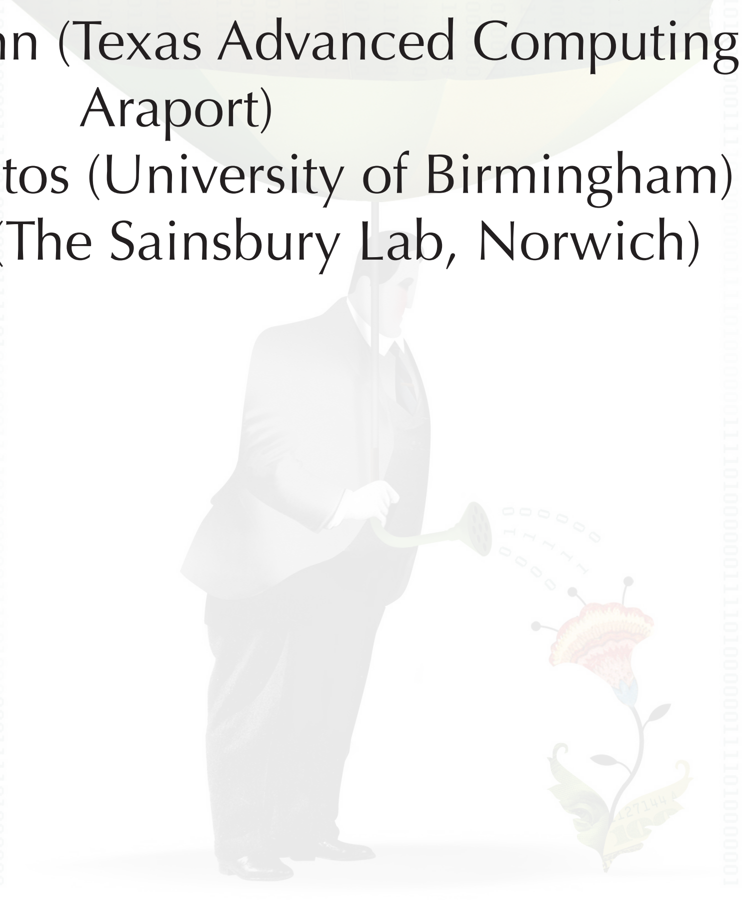
In 2013 we initiated Open Ash Dieback, a genomics crowdsourcing initiative to kickstart analyses of the invasive ash dieback fungus that was killing the UK's ash population. By creating a public Git repository containing genomics data and metadata we received a wide variety of new scientific results from a new community that analysed existing data and provided additional data from many sources. The progress after one year was significant and included genome assemblies of pathogen and host, identification of toxins and metagenomics analysis of samples. We observed that our Git based infrastructure was off-putting to knowledgeable scientists with non-genomics backgrounds and developed other interfaces to include these potential contributors.

Concurrently to this we ran a citizen science project that revealed much about human interaction interest patterns in open projects. In both projects we observed extreme interest early on in the project that dwindled later, though when crowdsourcing from scientists other incentives including the deposition of new data and imminent publications could restimulate activity. We conclude that a crucial factor in developing productive communities is timing, ensuring that data and access is provided earlyis essential to maximise output.

Panel Discussion 2: Strategies for Community Involvement

Chair: Ruth Bastow

Panel Members



Eva Huala (TAIR/ Phoenix Bioinformatics)
Matthew Vaughn (Texas Advanced Computing Centre, Araport)
Georgios Gkoutos (University of Birmingham)
Dan MacLean (The Sainsbury Lab, Norwich)

Panel Discussion 3: Business models for Data Infrastructures

Chair: Sabina Leonelli

Panel Members:

Geoffrey Boulton OBE (CODATA and The Royal Society)

President of the ICSU Committee on Data for Science and Technology (CODATA) and currently involved in CODATA strategies to create an open data platform in Africa, to advance the frontiers of data science, to seek widespread endorsement of the International Accord on Open Data and to develop and promote enabling processes and practices of open data in national research systems.

Rowan McKibbin (BBSRC)

The Head of Science Strategy for the Exploiting New Ways of Working sector, leads on the synthetic biology, systems biology, data intensive bioscience and technology development portfolios.

Michael Ball (BBSRC)

Strategy and Policy Manager

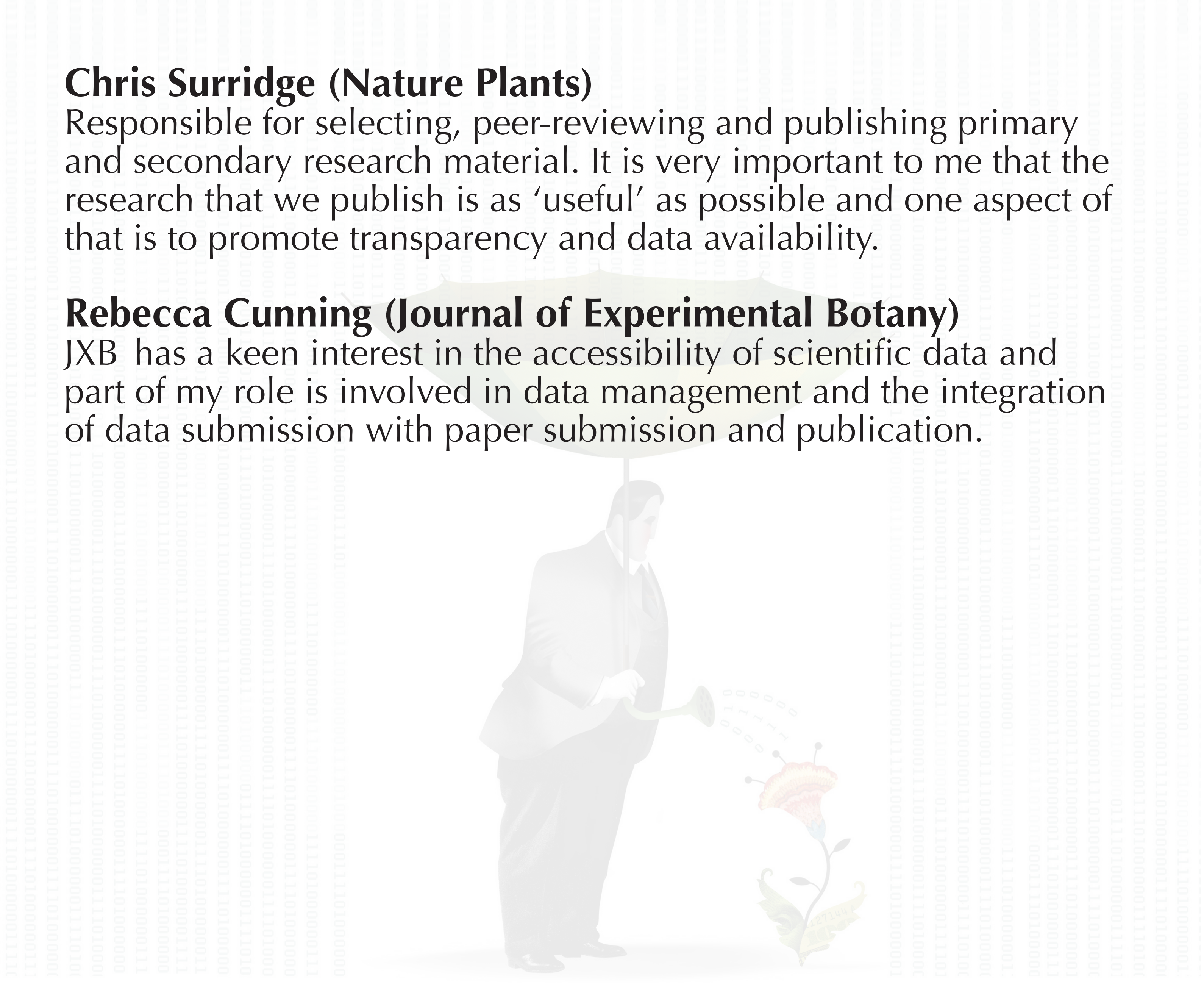
Derek Scuffell (Syngenta)

Derek is the Data Strategist for R&D in Syngenta, responsible for delivering the technical and cultural approaches that are needed to ensure that Syngenta data assets are able to be reused and exploited to bring the most value to Syngenta's research. This requires sustainable frameworks and approaches for Data Integration, Quality, Management and Trust. Derek is also a technical lead for digital disciplines in Syngenta R&D projects.

Panel Discussion 3: Business models for Data Infrastructures

Chris Surridge (Nature Plants)
Responsible for selecting, peer-reviewing and publishing primary and secondary research material. It is very important to me that the research that we publish is as 'useful' as possible and one aspect of that is to promote transparency and data availability.

Rebecca Cunning (Journal of Experimental Botany)
JXB has a keen interest in the accessibility of scientific data and part of my role is involved in data management and the integration of data submission with paper submission and publication.



Attendee List

Angela Hancock
Max F. Perutz Laboratories, Vienna
angela.hancock@univie.ac.at

Annemarie Eckes
TGAC
Annemarie.Eckes@tgac.ac.uk

Boas Pucker
Bielefeld University
boas.pucker@uni-bielefeld.de

Carole Goble CBE
Manchester University
carole.goble@manchester.ac.uk

Colin Sauze
Aberystwyth Phenomics Centre
cos@aber.ac.uk

Dan Maclean
The Sainsbury Laboratory, Norwich
dan.maclean@sainsbury-laboratory.ac.uk

David Johnson
OERC, Oxford University
david.johnson@oerc.ox.ac.uk

David Studholme
Exeter University
D.J.Studholme@exeter.ac.uk

Dominique Arnaud
Exeter University
D.E.Arnaud@exeter.ac.uk

Anne Plessis
Plymouth University
anne.plessis262@gmail.com

Anyela Camargo-Rodriguez
Aberystwyth Phenomics Centre
avc1@aber.ac.uk

Brian Rappert
Exeter University
b.rappert@exeter.ac.uk

Chris Surridge
Nature Plants
c.surridge@nature.com

Dan Bebber
Exeter University
d.bebber@exeter.ac.uk

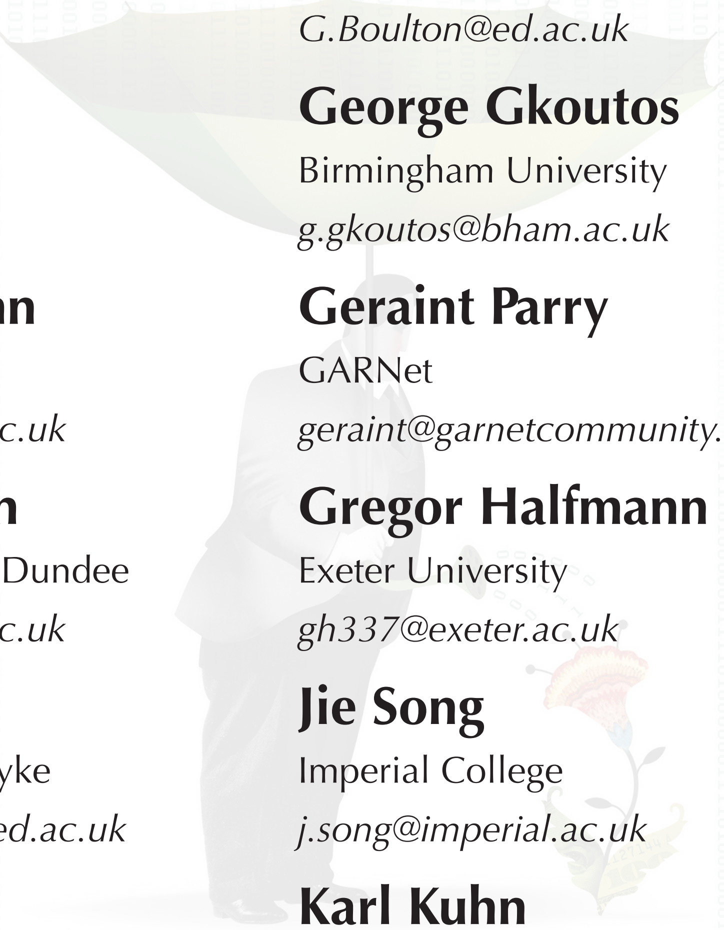
Daniel Seaton
Edinburgh SynSys
dseaton@staffmail.ed.ac.uk

David Salt
University of Aberdeen
david.salt@abdn.ac.uk

Derek Scuffell
Syngenta
derek.scuffell@syngenta.com

Elisabeth Georgii
Helmholtz Zentrum Mônchen
elisabeth.georgii@helmholtz-muenchen.de

Attendee List



Elizabeth Arnaud
Crop Ontology
e.arnaud@cgiar.org

Eva Huala
TAIR, Phoenix Bioinformatics
evahuala@gmail.com

George Bassel
Birmingham University
gbassel@gmail.com

George Littlejohn
Exeter University
G.R.Littlejohn@exeter.ac.uk

Gordon Simpson
James Hutton Institute, Dundee
g.g.simpson@dundee.ac.uk

Hadewij Sint
Rothamstead, North Wyke
hadewij.sint@rothamsted.ac.uk

John Hancock
TGAC
John.Hancock@tgac.ac.uk

Lizzy Donkin
Aberystwyth Phenomics Centre
eld38@aber.ac.uk

Michael Ball
BBSRC
Michael.Ball@bbsrc.ac.uk

Erik Alexandersson
Swedish University of Agricultural Sciences
Erik.Alexandersson@slu.se

Geoffrey Boulton OBE
Edinburgh University
G.Boulton@ed.ac.uk

George Gkoutos
Birmingham University
g.gkoutos@bham.ac.uk

Geraint Parry
GARNet
geraint@garnetcommunity.org.uk

Gregor Halfmann
Exeter University
gh337@exeter.ac.uk

Jie Song
Imperial College
j.song@imperial.ac.uk

Karl Kuhn
Delta-T Devices
Karl.Kuehn@delta-t.co.uk

Matt Vaughn
Texas Advanced Computer Centre
vaughn@tacc.utexas.edu

Attendee List

Michele Dominy
Bard University
mdominy@bard.edu

Nicholas Provart
Bio-Analytic Resource for Plant Biology
nicholas.provart@utoronto.ca

Pawel Krajewski
Institute of Plant Genetics PAS, Warsaw
pkra@igr.poznan.pl

Robert Davey
TGAC
robert.davey@tgac.ac.uk

Rowan McKibbin
BBSRC
rowan.mckibbin@bbsrc.ac.uk

Sabina Leonelli
Exeter University
S.Leonelli@exeter.ac.uk

Staffan Mueller-Wille
Exeter University
S.E.W.Mueller-Wille@exeter.ac.uk

Tomasz Zielinski
BioDare, Edinburgh University
tzielins@staffmail.ed.ac.uk

Niccolo Tempini
Exeter University
N.Tempini@exeter.ac.uk

Nick Smirnoff
Exeter University
N.Smirnoff@exeter.ac.uk

Rebecca Cunning
Journal of Experimental Botany
r.a.cunning@lancaster.ac.uk

Roger Boyle
Aberystwyth Phenomics Centre
rob21@aber.ac.uk

Ruth Bastow
GARNet
ruth@garnetcommunity.org.uk

Sean May
NASC
Sean.May@nottingham.ac.uk

Thibault Racovski
Exeter University
tr282@exeter.ac.uk

Xu Jian
Singapore University
dbsxj@nus.edu.sg

Zaki Ahmad
Royal Holloway University of London
Zaki.Ahmad.2015@live.rhul.ac.uk

GARNet NatVar16:

Natural Variation as a tool for Gene Discovery and Crop Improvement

Sessions and Keynote Speakers:

- **Epigenetic Variation**
  - **Detlef Weigel (Tuebingen)**
- **Ecology and Population Studies**
  - **Joy Bergelson (Chicago)**
- **GWAS**
  - **Magnus Nordberg (GMI)**
- **NatVar and QTL Analysis**
  - **Carlos Alonso-Blanco (CNB)**
- **Genetics of Adaptation**
  - **Jon Agren (Uppsala)**
- **Breakthrough Technologies**
  - **Robin Allaby (Warwick)**
- **Translational Studies**
  - **Ian Bancroft (York)**

Image: Detlef Weigel

Cambridge: Dec 12th-13th 2016

www.GARNetNatVar2016.weebly.com

Early-Bird Registration Opens July 1st.

£180 academics, £130 PDRA/ PhD students

Notes

Notes

GARNet2016:
Innovation in the Plant Sciences

Sessions and Keynote Speakers include:

- Fronters in Imaging
  - Ben Scheres (Wageningen)
- Cell Signalling
  - Niko Geldner (Lausanne)
- Plant Synthetic Biology
  - Cathie Martin (JIC)
- Genomics tools for Gene Discovery
  - Katherine Denby (Warwick)

- Workshops: *'Finding your Arabidopsis Gene in Wheat'*
  'Exploring *Araport*', *'Introduction to CRISPR-CAS'*

- Flash Presentations:
  'Two Slides, Two Minutes, No Waiting!'

Image: George Bassel

Cardiff: Sept 6-7th 2016

www.GARNet2016.weebly.com

Early-Bird Registration Now Open

£160 academics, £110 PDRA/ PhD students