

# **Report of the ERASysBio Summer School Data Management for Systems Biology II**

26<sup>th</sup> – 30<sup>th</sup> July 2009

Universidad de Verano de Adeje, Tenerife, Spain



Towards European standards in training and education in systems biology

## Report of the ERASysBio Summer School Data Management for Systems Biology II

26<sup>th</sup> – 30<sup>th</sup> July 2009, Universidad de Verano de Adeje, Tenerife, Spain

### Introduction

The partners of ERASysBio are committed to encourage the adoption of data standards, data management and sharing best practices in systems biology across the European Research Area.

One of the grand challenges in systems biology is the coupling of mathematical models and experimental data. Models are created from observing and extracting parameters from experimental data and are validated by testing model predictions against further experimental data. Modellers and experimentalists, however, often have different scientific backgrounds. Modellers are predominantly mathematicians, physicists and engineers, whereas experimentalists are biochemists, molecular biologists, or chemists. As a consequence, learning to work together and understand one another's work can be difficult. The second summer school Data Management for Systems Biology II aims to bridge the gap between the two communities, giving modellers experience in experimental data interpretation and experimentalists experience in constructing and testing models from data.

### Aims

- to achieve effective communication between all participants;
- to raise awareness of the urgent need to adopt data standards, data management and sharing best practices in the ERA;
- to provide knowledge on the current tools, methods, practices, models, challenges and solutions in the management of data in systems biology;
- to equip participants with tools for decision-making.

### Location

The summer school took place on 26<sup>th</sup>-30<sup>st</sup> July 2008 at the University of La Laguna Summer Campus, situated in Adeje, on the south-west coast of the island of Tenerife, only 10km away from Reina Sofia International Airport. The University of La Laguna is the oldest university in the Canary Islands, with its origin dating back to 1701 when Augustine monks established a centre for higher learning in the city of La Laguna. In 1992 the summer campus was founded in Adeje to promote culture, research, politics and philosophy among other disciplines, on the island.



## Organisers and sponsors

The summer school was organised by the Biotechnology and Biological Sciences Research Council (BBSRC), in collaboration with the Spanish Ministry of Science and Innovation (MICINN), the German Ministry of Education and Research (BMBF), Project Management Juelich (PtJ), ERASysBio and the Borough Council of Adeje. BBSRC, BMBF, MICINN, PtJ, ERASysBio and the Borough Council of Adeje co-funded this event.

## Participants

Twenty-four postdoctoral researchers from academic institutions in countries of the European Research Area attended the summer school, following a few last minute cancellations. There was approximately an equal number of experimentalists and modellers, i.e. bioinformaticians, software engineers, computer modellers. The countries represented in the summer school were Austria, Finland, Germany, Ireland, Italy, Luxembourg, The Netherlands, Slovenia, Spain, Sweden, and United Kingdom.

## Experts

The following experts contributed to the summer school:

### Thomas Höfer

Thomas is head of the Research Group Modeling of Biological Systems German Cancer Research Center and Bio Quant Center Heidelberg, Germany.

‘To rationalize the functioning of gene-regulatory networks, we need to understand their dynamics at multiple levels of molecular organisation. I will discuss studies iterating between experiment and mathematical analysis on (1) the differentiation of T cells and (2) the organisation of chromatin-associated multi-protein complexes in DNA repair. These studies make use of a variety of analytic and computational tools as well as molecular biology and live-cell imaging for kinetic quantification. The resulting experimentally-based mathematical models aid the discovery of new functional interactions in gene networks and reveal functional properties that emerge at the systems level. ‘

### Olga Krebs

Olga is working in the Scientific Databases and Visualization Group led by Isabel Rojas, at EML-Research, Heidelberg, Germany. She is one of developers of SabioRK - System for the Analysis of Biochemical Pathways - Reaction Kinetics (<http://sabio.villa-bosch.de/SABIORK>). Olga has a PhD in molecular genetics, 17 years professional experience in the "wet" lab and 9 years in the area of design and development of databases for bioinformatics and systems biology. She is a member of SysMO DB team, leading on the metadata design of SysMO-SEEK, SysMO-JERM and the database solutions.



## **Yuri Lazebnik**

Yuri Lazebnik is a Professor at Cold Spring Harbor Laboratory. His main interest is cancer biology, which took him to study cell cycle, apoptosis, and more recently cell fusion. While doing his experiments, Yuri became frustrated with an apparent disparity between the effort of biological research and its meaningful outcome. This frustration made him to ask the following question: can a biologist fix a radio?

By trying to answer this question Yuri argues, perhaps to the converted, that this disparity can be explained by a similar disparity between the complexities of the processes that we study and the tools with which most of biologists communicate. This argument will be intended to support the view that creating commonly accepted tools of communication that are sufficiently sophisticated to adequately reflect biological processes and, importantly, are useful to an average biologist not only can be of significant practical benefit for a particular field, but might change the very foundation of how biological research is organized. Yuri hopes that one of young scientists attending this meeting will finally develop such a tool.

## **Stuart Moodie**

Stuart Moodie is currently a Research Fellow in the group of Professor Igor Goryanin at the School of Informatics, Edinburgh University, UK. His research is focused on the computational representation and analysis of data related to Systems Biology. He is the main developer of the Edinburgh Pathway Editor and an editor of the Systems Biology Graphical Notation, a standard for the visualisation of biological pathways. He is currently the lead scientist in the BioMedCAD project, which aims to build integrated databases of metabolic and signalling pathways.

## **Sven Sahle**

Sven Sahle, Ph. D. in Theoretical Chemistry, is a junior group leader in the Department for Modeling of Biological Processes at the University of Heidelberg, Germany. He's one of the lead software engineers of COPASI[1], a software tool for modeling of biochemical reaction networks. His interests include the dynamic properties of biochemical models as well as simulation and analysis methods for those models. Sven is also an SBML Editor, elected by the SBML community.

SBML – The Systems Biology Markup Language SBML[2] is a computer-readable format for representing models of biochemical reaction networks in software. It is applicable to models of metabolism, cell-signaling, and many others. It has been developed since the year 2000 by an international community to enable the exchange of models between different software tools. By now numerous tools support SBML and it is also used to store models in model databases like biomodels.net[3].

The models SBML deals with usually consist of biochemical reactions, i.e. they describe processes which consume and produce different chemical substances, e.g. metabolites. The reactions are associated with a quantitative description of how fast they



happen under different circumstances (kinetic law), enabling numerical simulations and analysis of the model. In recent versions of SBML the parts of the model (like reactions and reactants) can also be linked to specific pieces of biological knowledge in databases like UniProt, KEGG, etc.

[1] [www.copasi.org](http://www.copasi.org)

[2] [www.sbml.org](http://www.sbml.org)

[3] [www.biomodels.net](http://www.biomodels.net)

### **Susanna Sansone**

Susanna-Assunta Sansone is a Coordinator at the European Bioinformatics Institute, Hinxton, UK, leading a team of software engineers, biocurators and trainees.

Her activities lie in the area of data curation, ontology and software development to assist data management (<http://isatab.sf.net>). She has developed a significant interest in the area of standardization for the purpose of enabling reporting, sharing and meta-analysis of biological, biomedical and environmental studies, commonly ranging from simple one assay-based to complex multi-assay studies. She is a founding member of several grass-root standards initiatives, contributing to the development of several minimum requirement checklists, ontologies and formats.

### **Ralf Steuer**

Ralf Steuer is a post-doctoral researcher at the Humboldt-University Berlin, Germany, and also holds a part-time position at the University of Manchester, UK. Key research interests include mathematical modelling of cellular metabolism, cellular information processing and metabolomics data analysis. Ralf is active in the research program FORSYS-Partner funded by the German BMBF, as well as in the research initiative SysMO (SulfoSys, SysMO-LAB and MOSES). His expertise includes Monte-Carlo sampling of metabolic networks, dealing with incomplete information and unknown parameters, the interpretation of metabolomics correlations, stochastic processes and the computational characterization of stability and regulation in metabolic networks.

### **Katy Wolstencroft**

Katy Wolstencroft is a post-doctoral researcher on the myGrid project. She runs the myGrid outreach program, providing regular training courses and consultancy for workflow design. Her background is in Bioinformatics and Bio-ontologies and her research interests involve workflows, e-Science and ontologies and the application of cutting-edge computer science technologies to biological problems.



## The programme

The summer school programme is at **Annex 1**.

The event commenced with the welcoming remarks by Julio Barbas, as ERASysBio Spanish partner, from the Ministry of Science and Innovation in Madrid, Spain.

A networking exercise, led by Gabriela Pastori from the Biotechnology and Biological Sciences Research Council in the UK, introduced participants in a structured environment. Participants interacted in a speed-dating exercise comprising brief introductions and rapid discussion.

### Session 1 – Data and standards

This session started with the keynote lecture 'Managing systems biology data: the SysMO-DB approach', by Katy Wolstencroft, University of Manchester. The session was followed by an introduction to the Data hands-on tutorial sessions:

- Conceptual Modelling and Standardisation of wet lab data, by Olga Krebs, EML-Research, Heidelberg, Germany.
- Systems Biology Markup Language (SBML) by Sven Sahle, University of Heidelberg, Germany.
- Standard Operating Procedures, by Katy Wolstencroft, University of Manchester, UK

The parallel, hands-on afternoon sessions covered the following topics:

- Data Modelling / Standardisation
- An introduction to SBML
- Standard Operation Procedures (SOPs):
  - Definition and storage of SOPs (experimentation)
  - Implementing SOPs in multi-partner projects

The session ended with the short presentation 'Systems Biology Requirements for Experimental Design', by Olga Krebs, EML-Research, Heidelberg, Germany.

### Session 2 - Tools for systems biology

This session was initiated with the keynote lecture 'Dealing with models when most parameter values are unknown', by Sven Sahle, University of Heidelberg, Germany. The session was followed by an introduction to the Tools hands-on tutorial sessions, by Katy Wolstencroft, Olga Krebs, Sven Sahle, and Stuart Moodie.

The parallel, hands-on tutorial sessions covered the following topics:

- Taverna, by Katy Wolstencroft, University of Manchester, UK
- Copasi/Sycamore, by Sven Sahle, University of Heidelberg, Germany
- SabioRK, by Olga Krebs, EML-Research, Heidelberg, Germany
- Edinburgh Pathway Editor, Stuart Moodie, Centre for Systems Biology at Edinburgh, UK



This session ended with the keynote lecture 'Maximising interaction between scientific communities through data exchange and software interoperability', by Susanna Sansone, European Bioinformatics Institute, UK

### **Session 3 - No barriers - bringing experimentalists and modellers together**

Yuri Lazebnik, Cold Spring Harbor, US, opened session with the keynote lecture 'Living in one world – common tools for communication between disciplines'. This was followed by an introduction to the No Barriers hands-on tutorial sessions, by Ralf Steuer, Humboldt University, Germany.

The parallel, hands-on tutorial sessions were run by Olga Krebs, Katy Wolstencroft, Sven Sahle, Yuri Lazebnik, Susanna Sansone, Stuart Moodie and Ralf Steuer. This final session was designed around participants. Four different topics were chosen by participants and groups were assembled according to interest. Participants presented their work in the morning on the following day, and discussed issues and outcomes as well as lessons learned.

This session ended with the keynote lecture 'Two waves of T-bet: Mechanistic models of gene-regulatory networks', by Thomas Höfer, German Cancer Research Centre, Heidelberg, Germany.

### **Final Session – Feedback**

This session was led by Gabriela Pastori. The Final Session provided participants an opportunity to sit back and reflect on what they learned, and also to think of the future. Participants were presented with a few questions to be discussed in breakout groups.

1) Is there anything that you are going to do differently as a result of this training? What do you take home?

Participants responded positively to these questions and expressed their strong will to continue the process started at the summer school. The main points of the discussion were:

- Awareness of tools available.
- New approaches and tools to experiment with: Copasi, workflows, databases to annotate data.
- Opportunity to meet people with different backgrounds, working in different areas.
- Opportunity to new lines of communication, cooperation through grants.
- Better understanding, more confidence on what each side needs.
- Meeting potential collaboration partners, making new contacts, video conference as an alternative.
- Understanding different ways and different view points to solve problems.
- Visual representation as a tool to help communication between experimentalists and modellers.



- Preparation previous to a meeting between experimentalists and modellers; each side should present what expectations they have and make clear what they are trying to achieve.

The group shared their future plans for using tools (Copasi, SBML) in SysMO Seek. The experience at the summer school served to stimulate participants to take a closer look at problems, programmes, work and explore how to implement the new knowledge.

2) Are there any barriers that you would have to overcome to implement your plans back home?

#### *Institutional, geographical barriers*

- The current institutional structure does not favour communication between experimentalists and modellers.
- There is a language barrier that still needs to be overcome.
- Lines of communication need to be established between experimentalists and modellers.
- There are limited opportunities for contact; people work in different projects, and there are topic-oriented as well as geographical barriers to face. Video conference offers an alternative; there still seems to be a need to meet in person though.
- The way grant proposals are written determines the way people are encouraged to work. Both grant applicants and advisory boards assessing systems biology grants should be aware of the impact of a project's structure on the relationship between experimentalists and modellers.

#### *Information flow*

- At present, there is a lack of overview of what information is available, i.e. catalogues, tutorials, and how to choose the best tool for each case. Tutorials, workshops, other resources are seen as essential as much as receiving feedback on its usage and application.
- Establishing where the sources of information are seems an important barrier to overcome.
- A great degree of patience and open mind is required from both experimentalists and modellers. Both need to be prepared to invest time in trying to understand each other's position. This process may take longer than expected but it's thought to be necessary, and critical to success.

3) Are there any wider issues that you would like to comment on and/or be included in future summer schools?

- Introduction to modelling, model building.
- Introduction to experiments, statistics.
- More presentations of new tools.
- Experts' talks should avoid being research-specific.
- Successive, not parallel sessions as far as possible.
- Online tutorials on Copasi and Taverna.



- If possible, work out participants' preference for sessions in advance; or make short taster session of each tool so people can choose their preferred session.
- Sharing participants' profiles in advance of summer school.

This session ended with the presentation of the Data Management Summer School Huddle – ShareSysBio – by Gabriela Pastori. ShareSysBio is supported by the Biotechnology and Biological Sciences Research Council (BBSRC), under the auspices of ERASysBio.

ShareSysBio aims to provide a forum for scientific discussion and a tool for sharing knowledge and best practices in the management of data and models generated by publicly-funded systems biology projects. ShareSysBio offers a space for early-career systems biologists having attended ERASysBio Data Management Summer Schools to continue networking, exchanging documents, tips, ideas, and most of all, best practices in the management and exchange of data and models.

Participants attending the summer schools in 2008 and 2009 have become members of ShareSysBio.

### **Acknowledgements**

We wish to thank Gracia Rodríguez and Adolfo Gómez, at Universidad de La Laguna, Jose Miguel Rodríguez Fraga, Epifanio Díaz, Moisés Albertos and Israel Dorta, at Ayuntamiento and Universidad de Verano de Adeje, for their kind hospitality and effective support at the Centro Cultural de Adeje. We thank Julio Barbas and María Jesús Sillero, at Ministerio de Ciencia e Innovación, Spain, for their generous help, and Julio for his enthusiasm and commitment to this summer school. We also thank our colleagues Gisela Miczka, Veronika Simons, Bernhard Gillessen and Gezahegn Gebreselassie, at Project Management Juelich, Germany, for their valuable support to this project, and Bernhard for his kind and generous help during the event. Special thanks are due to Gabriela Pastori, at the Biotechnology and Biological Sciences Research Council, for the overall organisation of the summer school and for all her dedicated work during the summer school itself. Finally, we thank all experts and participants who attended the summer school, shared their experience and lively interest in this topic, and contributed to its success.

END

*BBSRC, 2009*





## ANNEX 1



**ERASysBio Summer School**  
**Data Management for Systems Biology II**  
Universidad de Verano de Adeje, Tenerife, Spain  
26<sup>th</sup> – 30<sup>th</sup> July 2009

### Sunday, 26<sup>th</sup> July 2009

Delegates to Arrive  
17:00 – 19:00 Registration  
20:00 Dinner

---

### Monday, 27<sup>th</sup> July 2009

07.15 – 08.15 Breakfast  
08.20 Bus departing to summer school venue  
09:00 – 09:05 Welcome by ERASysBio  
Julio Barbas, MICINN, Spain  
09:05 – 10:30 Networking exercise  
Led by Gabriela Pastori, BBSRC, UK

Aim: to introduce participants in a structured environment. Participants to interact in a speed-dating exercise comprising brief introductions and rapid discussion

Structure: half of participants will remain seated; the other half will rotate every 3 / 4 minutes



10:30 – 10:45 Coffee Break and Refreshments

**Session 1 Data and standards**

10:45 – 11:30 Keynote lecture: 'Managing systems biology data: the SysMO-DB approach', **Katy Wolstencroft**, University of Manchester

11:30 – 12:25 Introduction to parallel sessions including:

- Conceptual Modelling and Standardisation of wet lab data, by **Olga Krebs**, EML-Research, Heidelberg, Germany.
- Systems Biology Markup Language (SBML) by **Sven Sahle**, University of Heidelberg, Germany.
- Standard Operating Procedures, by **Katy Wolstencroft**, University of Manchester, UK

12:25 – 12:30 Formation of Breakout Groups and General Introduction to the Hands-on Sessions

12:30 – 13:30 Lunch

13:30 – 15:45 Parallel Hands-on Tutorial Sessions

- Data Modelling / Standardisation
- An introduction to SBML
- Standard Operation Procedures (SOPs):
  - Definition and storage of SOPs (experimentation)
  - Implementing SOPs in multi-partner projects

15:45-16:00 Coffee Break and Refreshments

16:00-16:30 Feedback from Parallel Sessions

16:30 – 17:00 Short lecture: Systems Biology Requirements for Experimental Design, **Olga Krebs**, EML-Research, Heidelberg, Germany.

17:15 Bus departing to hotel

20:00 Dinner

---



## Tuesday, 28<sup>th</sup> July 2009

- 07.15 – 08.15 Breakfast  
08.20 Bus departing to summer school venue  
09:00 – 09:15 Recap of day 1

### Session 2 Tools for systems biology

- 09:15 - 10:00 Keynote lecture: 'Dealing with models when most parameter values are unknown', **Sven Sahle**, University of Heidelberg, Germany.
- 10: 00 – 10:20 Introduction to Tools Hands-on Tutorial Sessions  
Katy Wolstencroft, Olga Krebs, Sven Sahle, Stuart Moodie
- 10:20 – 10:35 Coffee Break and Refreshments
- 10:30 – 10:40 Formation of break-out groups
- 10:45 – 13:30 Parallel Hands-on Tutorial Sessions  
Taverna, by **Katy Wolstencroft**, University of Manchester, UK  
Copasi/Sycamore, by **Sven Sahle**, University of Heidelberg, Germany
- 13:30 – 14:00 Lunch
- 14:00 – 16:00 Parallel Hands-on Tutorial Sessions  
SabioRK, by **Olga Krebs**, EML-Research, Heidelberg, Germany  
Edinburgh Pathway Editor, **Stuart Moodie**, Centre for Systems Biology at Edinburgh, UK
- 16:00 16:15 Coffee Break and Refreshments
- 16:15 -16:45 Presentation of other Tools by participants
- 16:45 – 17:30 Feedback from Parallel Sessions
- 17:30 – 18:15 Keynote lecture: 'Maximising interaction between scientific communities through data exchange and software interoperability', **Susanna Sansone**, European Bioinformatics Institute, UK
- 18:30 Bus departing to hotel
- 



**Wednesday, 29<sup>th</sup> July 2009**

**Session 3      No barriers - bringing experimentalists and modellers together**

- 07.15 – 08.15    Breakfast
- 08.20            Bus departing to summer school venue
- 09:00 – 09:15    Recap of day 2
- 09:15 – 10:15    Keynote lecture: ‘Living in one world – common tools for communication between disciplines’, Yuri Lazebnik, Cold Spring Harbor, US
- 10:15 – 10:30    Coffee Break and Refreshments
- 10:30 – 10:45    Introduction to the No Barriers Hands-on Tutorial Sessions, **Ralf Steuer**, Humboldt University, Germany
- 10:45 – 11:00    Formation of break-out groups
- 11:00 – 13:00    Parallel Hands-on Tutorial Sessions  
Olga Krebs, Katy Wolstencroft, Sven Sahle, Yuri Lazebnik, Susanna Sansone, Stuart Moodie and Ralf Steuer
- In Systems Biology, the relationship between modellers and laboratory scientists is crucial. Learning to work together and understand each other can be difficult. In this session, participants will learn to bridge the gap between the two communities: modellers will be given experience in experimental data interpretation and experimentalists will learn to construct and validate models using experimental data. More details to follow.
- 13:00 – 14:00    Lunch
- 14:00 – 16:00    Parallel Hands-on Tutorial Sessions (cont.)
- (15:00 - 15:15    Coffee Break and Refreshments)
- 16:00-16:45    Feedback/Reporting  
This final session will be designed around participants. We will discuss issues and problems that have arisen throughout the day: what have they learned.
- 17:00 -18:00    Keynote lecture: ‘Two waves of T-bet: Mechanistic models of gene-regulatory networks’, **Thomas Höfer**, German Cancer Research Centre, Heidelberg
- 20:00            Dinner
- 



**Thursday, 30<sup>th</sup> July 2009**

- 07.15 – 08.15 Breakfast
- 08.20 Bus departing to summer school venue
- 09:00 – 09:15 Recap of day 3
- 09:15 – 11:00 **No barriers session - presentations**  
**Gabriela Pastori, BBSRC, UK**
- 11:00 – 11:30 Coffee Break and Refreshments
- 11:30 – 12:00 **Final Session**, including Overall Feedback  
**Gabriela Pastori, BBSRC, UK**
- 12:00 – 12:15 Closing remarks  
Gabriela Pastori, BBSRC, UK
- 12:30 – 13:30 Lunch & End
- 13:30 Bus departing to hotel

END



Organised and Sponsored by



Under the auspices of

