

Deliverable 5.1, January 2014

Survey of the existing systems biology community

The evolving landscape of
the European systems biology community

Work package 5 Community Building and Synergies

ISBE WP5 Coordinators:
James Sharpe and Jutta Steinkoetter

Main responsible authors:
Joaquim Calbó, James Sharpe
Centre for Genomic Regulation (CRG)

Contact:
joaquim.calbo@crg.eu

Project ref. no.	INFRA-2012-2.2.4: 312455
Project title	ISBE – Infrastructure for Systems Biology Europe
Nature of Deliverable	R= Report
Contractual date of delivery	Month 18
Actual date of delivery	Month 18
Deliverable number	D5.1
Deliverable title	Survey of the existing systems biology community
Dissemination Level	PU
Status & version	Version 2
Number of pages	20
WP relevant to deliverable	WP5
Lead Participant	CRG, James Sharpe
Author(s)	Joaquim Calbó, James Sharpe, Jutta Steinkoetter, Eadaoin McKiernan, William Fitzmaurice, Katy Wolstencroft , Angela Krueger, Garry Corthals, Eeva Rainio, ...
Project coordinator	Richard Kitney
EC Project Officer	Andreas Holtel

Dissemination level: PU = Public, RE = Restricted to a group specified by the Consortium (including Commission services), PP = Restricted to other programme participants (including Commission Services), CO= Confidential, only for members of the Consortium (including the Commission Services)

Nature of Deliverable: P= Prototype, R= Report, D=Demonstrator, O = Other.

Table of Contents

Introduction	4
The evolving landscape of European systems biology	4
Methodology and results	5
1. Dataset 1: Internet-search-based database of Systems Biology groups and centres in Europe	5
2. Dataset 2: Pubmed-based database of Systems Biology researchers and institutions in Europe	6
3. Overlap of the two datasets (coverage)	7
4. Survey based on Datasets 1 and 2	8
5. Questionnaire-based survey	12
6. Community website: main tool to define and understand the landscape of Systems Biology in Europe	17
7. Participation to the ICSB2013 in Copenhagen: opening ISBE to the international community and identifying international state-of-the-art institutions	18
Conclusions	19

Introduction

The evolving landscape of European systems biology

Although systems biology (SB) has multiple historical roots (dynamical systems theory, molecular biology, physiology and high-throughput “omics” technologies), it is rather new as a field recognized on its own. Its unique interdisciplinary character involves collaborations between molecular biologists, geneticists, computer scientists, physicists, as well as mathematicians. These researchers often consider themselves as experts in tightly-defined disciplines, rather than as part of the SB community.

Thus, the SB community is highly inter-disciplinary, creating both opportunities and challenges. While it enables innovative collaborations and new approaches to previously intractable problems, this dispersal and the different disciplinary backgrounds can become barriers to fostering a sense of a genuine community. In addition, it becomes a challenge for the mission of ISBE to create an integrated infrastructure providing the community with access to state-of-the-art expertise and technologies.

In order to reach its objectives, ISBE needs to generate an understanding of the landscape of SB in Europe: who is involved in SB projects, what are the main research areas addressed, what are the technologies needed now and in the future, how the systems approaches can be valuable for the life science community in general. This document summarizes the work towards the definition of the European SB landscape that the WP5 – Community building and synergies– has developed in the first period of the ISBE project, and it is primarily focused on the community of systems biologists, including but not limited to the researchers that already use systems approaches in their research or that participate in SB projects. Using web searches, data mining, questionnaires, assistance to meetings, and the recently launched community website, we have collected relevant information that should facilitate the understanding of the community and its needs.

The resulting portrait, however, is not complete because the landscape is constantly evolving, and because the area of influence of SB can be extended to virtually all areas of research in life sciences, to all activity sectors, and to all geographical areas. Importantly, ISBE has established an important tool for the continuous monitoring of the evolution of the European Systems Biology community: the community web-portal (see Deliverable 5.2).

Methodology and results

1. Dataset 1: Internet-search-based database of Systems Biology groups and centres in Europe

The first activity towards the identification of the European community of systems biologists was based on the internet-based search of research groups, departments and institutions in Europe that self-identify themselves as performing Systems Biology. The methodology was established as follows:

- i. List the institutions involved in main SB initiatives in Europe (i.e. ISBE, ERASysBio, SystemsX).
- ii. Complement the list with direct search of institutions and departments with the term “Systems Biology” in their name, for the underrepresented countries.
- iii. Visit the website of each of the listed institutions. Search for Departments and for Groups with the term “Systems Biology” in their name.
- iv. For each identified research group, the following information was manually identified from the website:
 - Name of the group leader
 - e-mail address¹
 - Group/lab
 - Programme/Department/Division
 - Area of expertise
 - Topics of research
 - Group website (url)
 - Personal website (url)
 - Institution (name, acronym, url, country)

This process generated a list of >500 researchers leading SB-groups in >100 European institutions, and permitted to run a preliminary analysis of the state-of-the-art in the field. The main findings of this analysis are discussed in the following sections.

¹ e-mail addresses were carefully kept confidential and only used to run the ISBE questionnaire (see section 5)

2. Dataset 2: Pubmed-based database of Systems Biology researchers and institutions in Europe

A more unbiased approach to the identification of researchers working on SB projects is to screen the publications in the field. We have exploited this option in order to obtain a dataset complementing the one described in section 1, and also to populate the European SB community website with seed data.

For this endeavour WP5 established a collaboration with WP2 (data management, main collaborator: Katy Wolstencroft -University of Manchester-), which had been agreed on during the WP5-WP2 joint meeting in Barcelona (April 2013). The following criteria were established in the search of articles listed in Pubmed to constrain the dataset:

- Main affiliation located to one of the European countries (EU-28 + EEA)
- Published in one of the main SB journals (Molecular Systems Biology, BMC Systems Biology,...) OR title or keywords contain the term "Systems Biology"
- 2008-2012

A set of approx. 2,000 articles were selected according to these criteria, and the metadata downloaded and used to extract the following information:

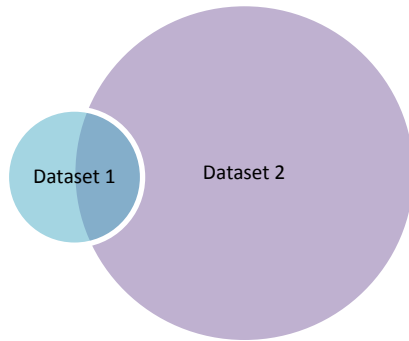
- Authors (>7,000)
- Affiliation (Institution, City, Country) associated to the main author of the paper (>700)
- Keywords as provided by Pubmed (>4,000)
- Publication details: date, journal, title

Since the metadata retrieved from Pubmed contains information of only one affiliation, without directly attributing this affiliation to one of the co-authors, the decision was taken to link the provided affiliation with the last-listed author of each article. We estimated that this assumption would be correct in more than 80% of the cases. However, this approach implies that most of the researchers in the dataset are not linked to any affiliation.

In spite of the above-mentioned limitations to the analysis, this approach allowed the mapping of >1,300 researchers, which is the base for the analysis presented here.

The data obtained was interpreted in a similar way as the web-based dataset, and the results obtained are presented in the following sections.

3. Overlap of the two datasets (coverage)



In order to assess the coverage of the dataset obtained by internet-based search, the overlap of the two datasets has been calculated and represented in the figure:

About 260 out of the 560 SB group leaders listed in the Dataset 1 were also present in Dataset 2, meaning that these researchers fulfilled the restrictive inclusion criteria used in the unbiased approach. The 300 group leaders listed in Dataset 1 that are missing in Dataset 2 are likely the sum of

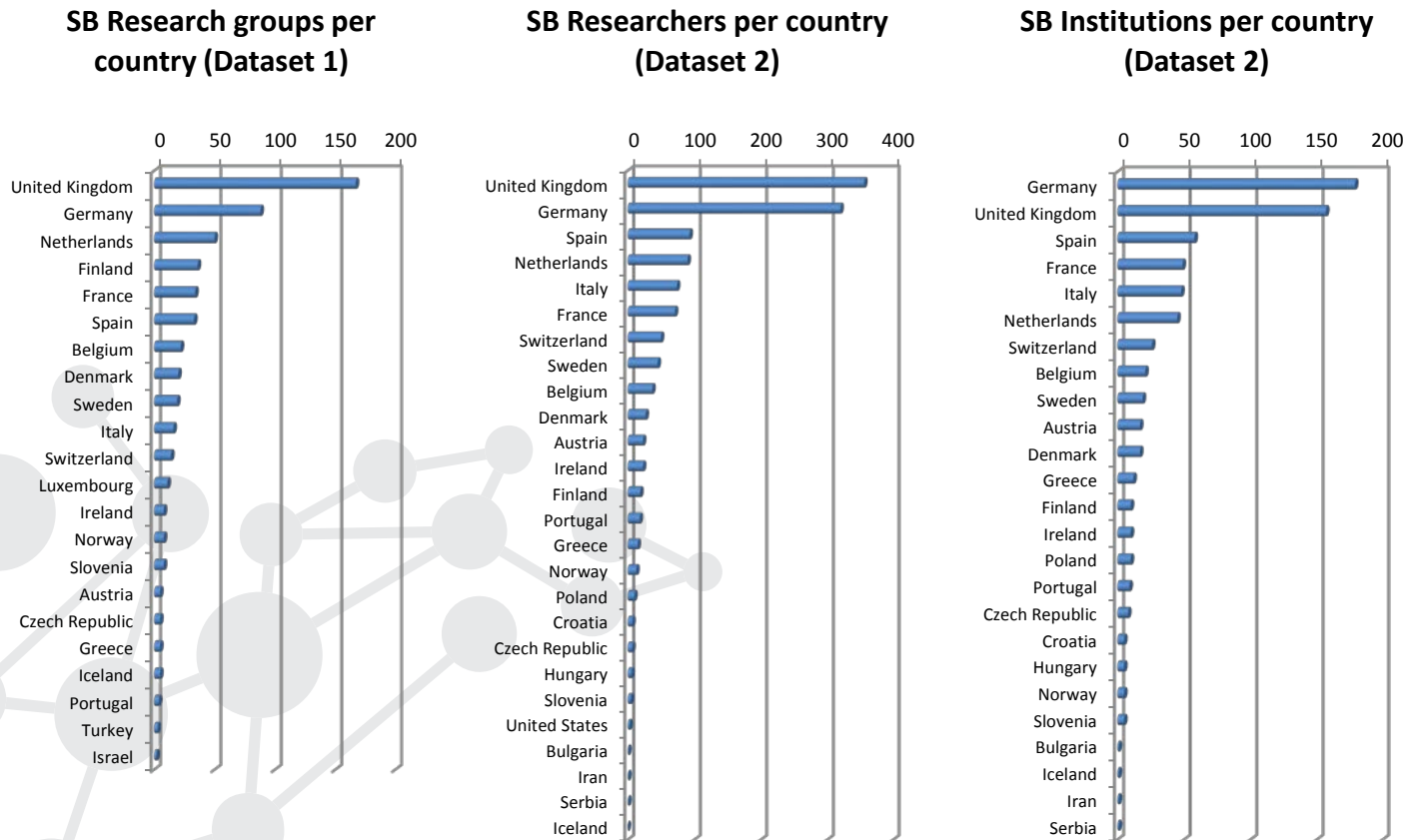
the following groups: genuine SB group leaders that did not publish according to the inclusion criteria (false negatives of Dataset 2), researchers mistakenly listed as SB group leaders that do not (any more) perform SB research (false positives of Dataset 1).

Obviously, most of the 7000 researchers listed in Dataset 2 are not present in Dataset 1, due to the unbiased approach that allows a broad screening of the SB community, but also to the inclusion in the second study of: 1) researchers that are not group leaders, 2) non-European researchers co-authoring Europe-based articles.

We thus consider that the two approaches and datasets are strongly complementary and that together provide a relevant coverage of the European SB community.

4. Survey based on Datasets 1 and 2

4.1. SB groups/researchers/institutions per country



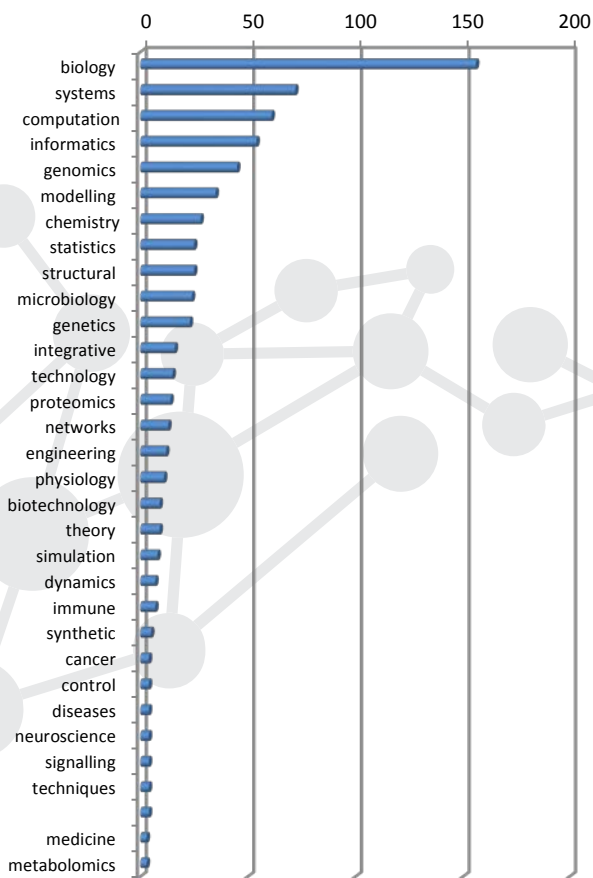
Dataset 1 (website search) covered 22 countries (17 of them are EU-28 member states). A large majority of the listed SB group leaders had been identified in a few countries, namely UK (167), Germany (88), The Netherlands (50), Finland (36), France (34) and Spain (33). These figures may not represent or even be proportional to the actual number of SB groups in each country because of:

- (a) the biased methodology employed to obtain these data, and
- (b) differences in the way that research groups and institutions are named in each country (ie. to what extent the phrase *systems biology* is used).

The unbiased approach used to obtain Dataset 2 (PubMed-based), however, broadly reproduced the results observed in Dataset 1, but with a few significant differences. The two most represented countries (in researchers per country) are again UK and Germany, but in this dataset the difference in numbers between these two leading countries is rather minimal. The subsequent most-represented countries are Spain and the Netherlands, and then Italy and France. Regarding SB institutions in each country (institutions with affiliated researchers found in the dataset), the most represented countries are Germany and UK, followed by Spain, France, Italy and the Netherlands.

The differences between the two datasets may be due to the fact that some countries have, in the last years, invested significant funds to boost Systems Biology in their national research roadmaps. These countries may appear overrepresented in Dataset 1, due to the creation of centres and research-lines oriented (and named) to Systems Biology, and to the increased participation in international SB initiatives.

4.2. SB groups per expertise area (Dataset1)

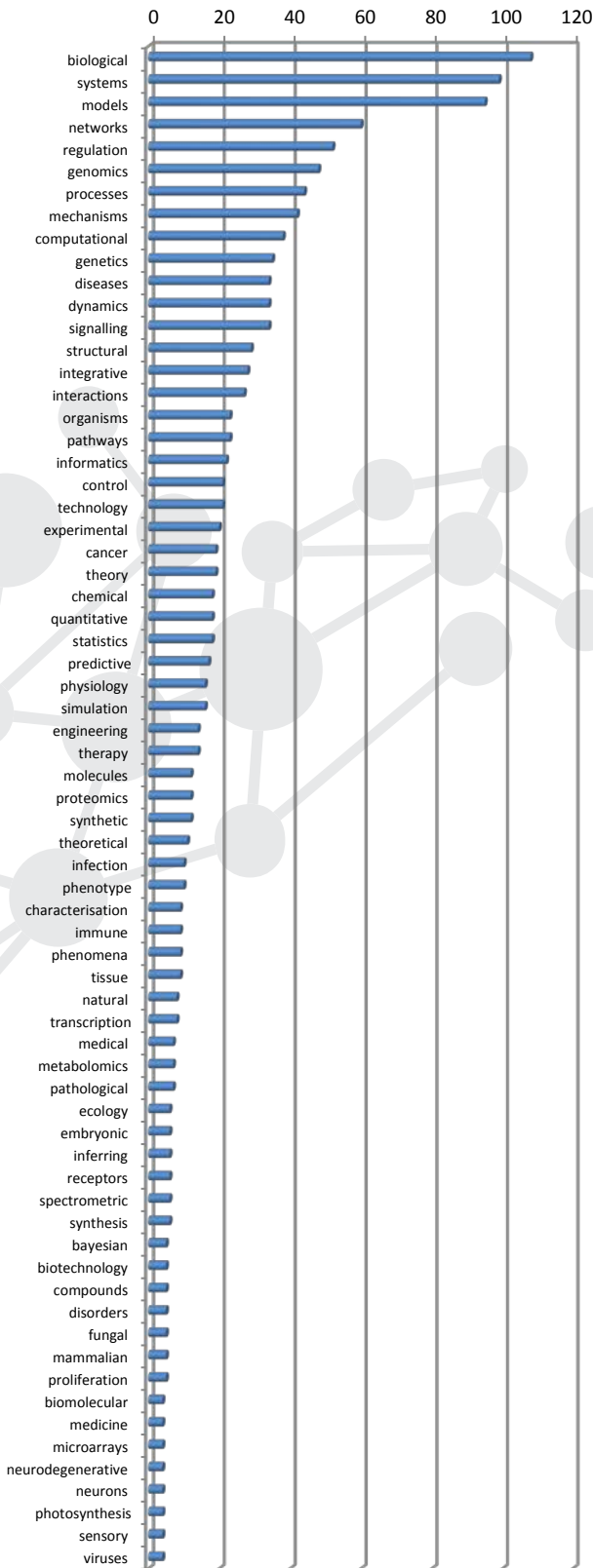


The analysis of expertise keywords based on the information collected as “Area of expertise” for each of the identified group leaders (Dataset 1), revealed that:

- The main areas of expertise for SB groups are systems, modelling and integrative biology, and computation, informatics and statistics.
- The list of expertise areas is very broad and covers almost all areas of the life sciences, and also chemistry and engineering.

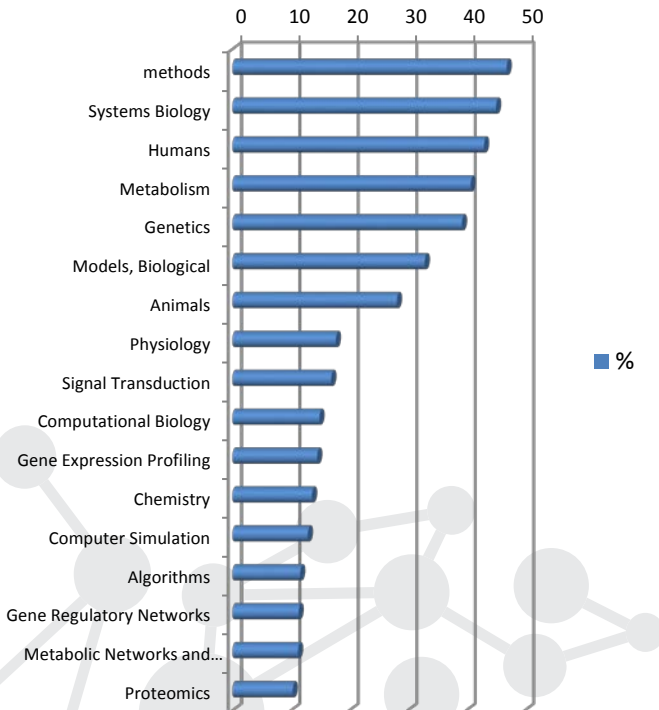
4.3. Analysis of keywords / research topics (Dataset 1)

SB Research topics terms



The analysis of the keywords found in the description of the research topic for each of the listed SB groups, shown in the figure, revealed that these groups address many different topics using SB approaches. The most frequent terms (around 20% of the SB groups) are *biological*, *systems* and *models*, followed by a number of terms directly related to the systems approach (*networks*, *regulation*, *processes*, *mechanisms*, *dynamics*, *integrative*, *interactions*, *pathways*...), with frequencies between 10% and 20%. Many other terms refer more specifically to the object of the research (*genomics*, *diseases*, *organisms*, *cancer*, *physiology*, *therapy*, *infection*...). Importantly, these specific terms are mostly found with low frequencies (<5%), confirming that SB research is very diverse in the object of study and addresses all fields of Life Sciences, although with a relatively important focus on health-related topics.

4.4. Analysis of keywords / research topics (Dataset 2)



The analysis of the keywords linked to the retrieved articles brings similar results as the obtained in the analysis of Dataset 1, with some differences due to the ontologies used by the NCI to categorize the publications. Interestingly, the top ranking term is *methods*, which is listed as keyword in almost half of the SB-related publications. This finding suggests that in the field of SB there is a major component of methodology development. Other most-used keywords, in addition to *Systems Biology*, are: *humans*, *metabolism*, *genetics*, *biological models*, and *animals*. Several terms related to systems approaches are found with frequencies between 10% and 20%: *Computational biology*, *computer simulation*, *algorithms*, *gene regulatory networks*...

5. Questionnaire-based survey

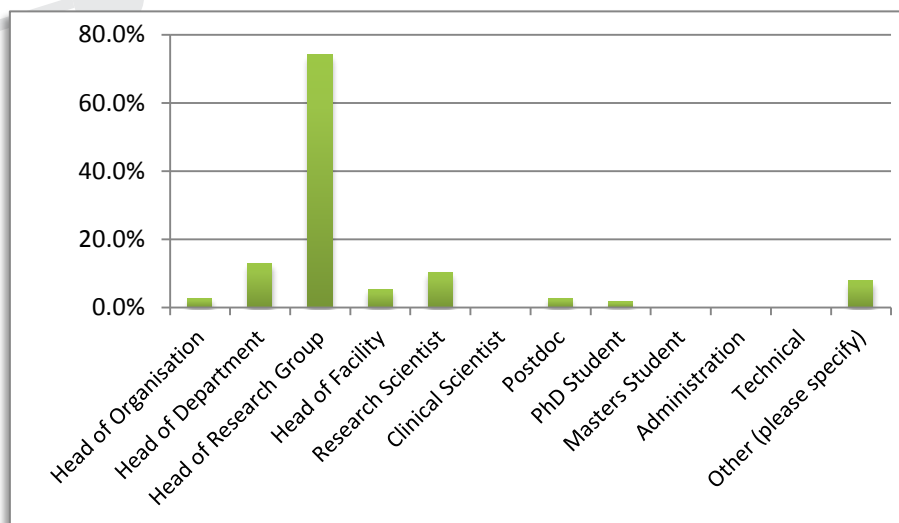
WP5 also addressed a number of questions to the research community through the [ISBE questionnaire](#) launched as a collaborative effort of several ISBE-WPs. The questionnaire has been placed as a link into the project's [website](#), and has also been distributed via e-mail to the >500 SB group leaders identified in point 1. By December 10, 2013 more than 120 respondents had completed the questionnaire (which remains open to new respondents).

Questions on use and needs of infrastructure for data generation, stewardship and modelling, and on training needs, are reported elsewhere.

The questions relevant to this survey and the obtained responses are reported in the following sections:

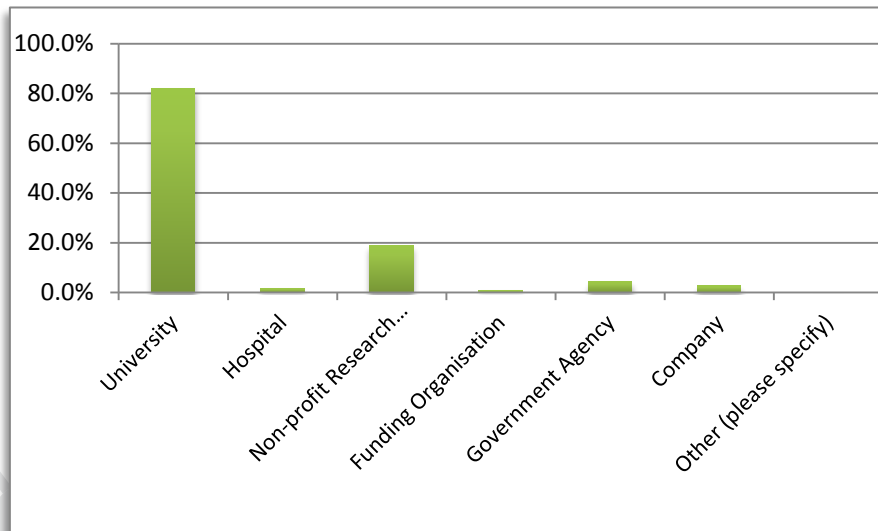
5.1. What is your position in the organisation?

In order to characterize the population of respondents and to be able to stratify the responses (when needed), we questioned about the position in the organization. Non-surprisingly, considering that the survey was sent mainly to SB group leaders identified in the Dataset 1, most of the respondents were heads of research groups and/or heads of facilities, departments or institution. Thus, we consider that the obtained responses might not fully represent the complete community, but conversely may have higher relevance in terms of knowledge of the field.



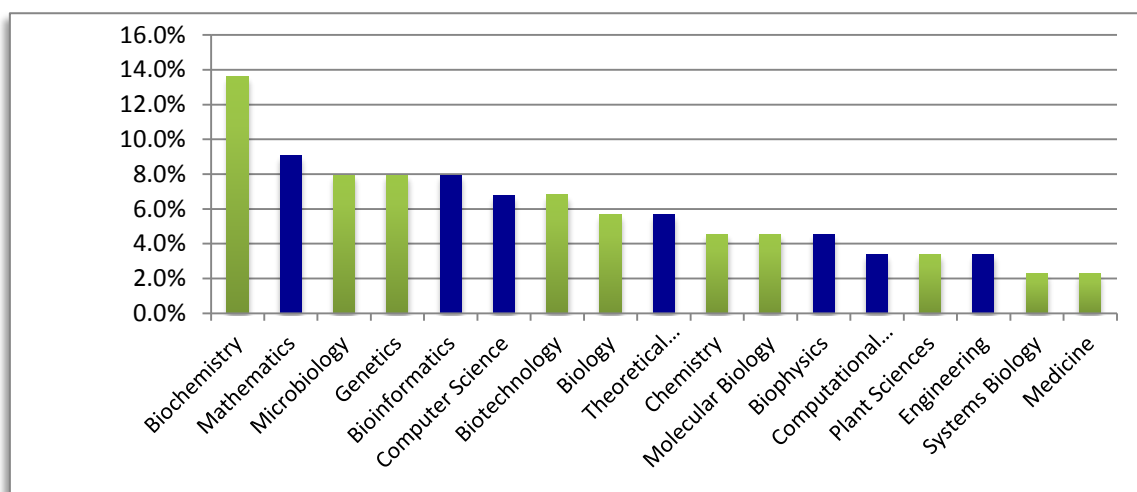
5.2. What type of organisation do you work for?

According to the questionnaire results, a vast majority of the respondents are affiliated to universities (80%) and Non-profit RPO (19%).



5.3. Please provide the scientific field for the PhD degree that you hold.

Knowing the background expertise and training of the involved researchers is also important to have a proper understanding of the community and the research field. Virtually all respondents stated to have gained a PhD degree, but as can be observed in the graph, the scientific field for the PhD degree was once more very diverse.



Key points:

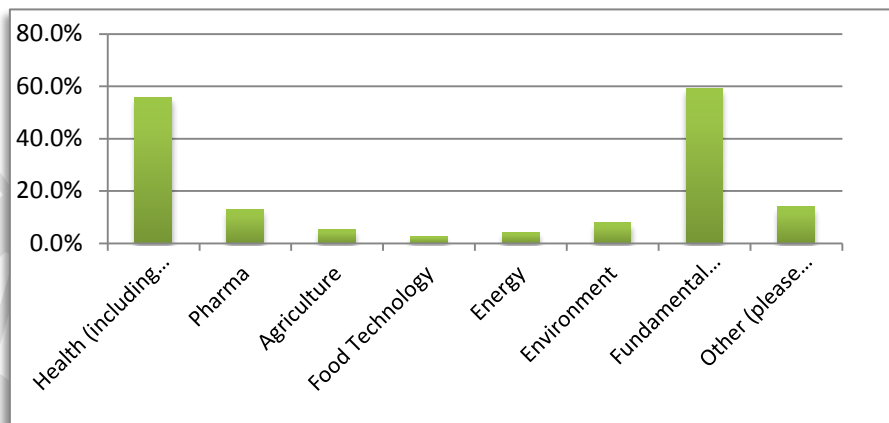
- The distribution of bio-sciences (represented in green in the graph) versus “hard”-sciences (maths, physics, computing, represented in blue in the graph) is very even.

This confirms the view of systems biology as being a highly inter-disciplinary mix of experimental and theoretical expertise.

- The topic of “systems biology” as a distinct field for a PhD qualification is much underrepresented, surely reflecting the relative youth of the field. Courses in systems biology are only being started relatively recently.

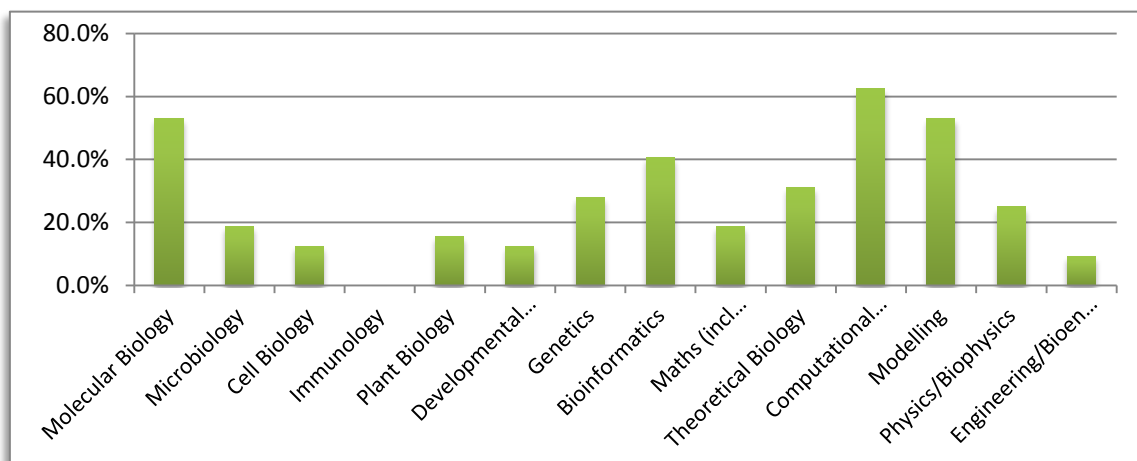
5.4. What is your current field of interest?

When asked about the (broad) field of research interest, there was a clear orientation to health and/or to fundamental research, with fewer researchers focused on translational fields (pharma, environment, agriculture, energy and food technology).



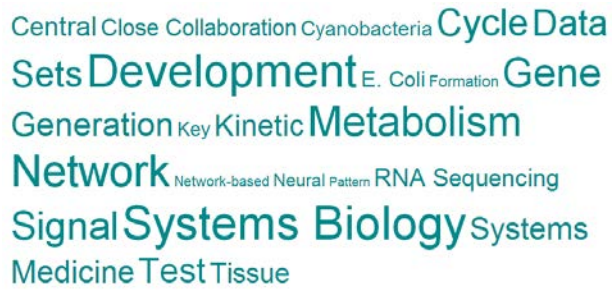
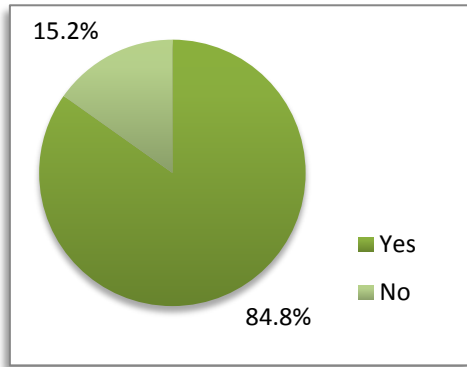
5.5. Please specify the field of fundamental research in which you currently work.

Regarding the specific fundamental-research field of interest, the responses were very diverse in agreement with data reported elsewhere in this document. Still, most of the researchers stated to be mainly focused on: 1) computational biology, modelling and bioinformatics, 2) molecular biology and genetics, and 3) maths, theoretical biology and biophysics.

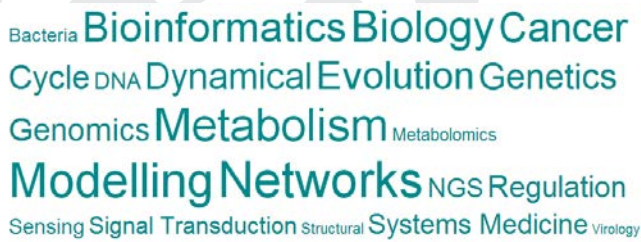


5.6. Are you currently involved in a SB project?

Most of the respondents to the questionnaire declare to be currently involved in SB projects (85%), and the analysis of the text of the short project-descriptions generated the word-cloud represented in the figure (main topics: *Network, Development, Metabolism, Gene*).

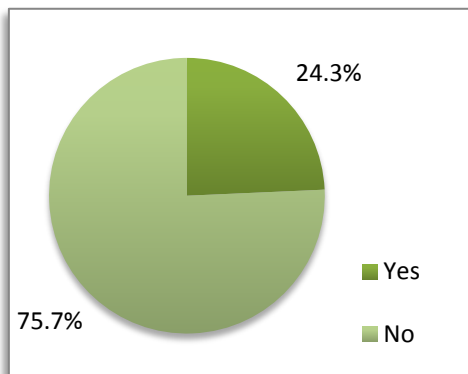


5.7. Keywords



Respondents also self-characterized their research using a few keywords, and the frequency of use of each keyword is represented in the word-cloud:

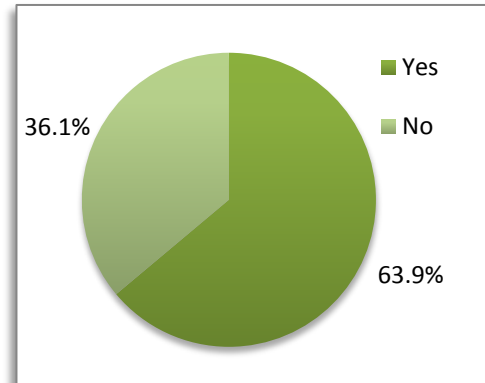
5.8. Have you received training in systems biology?



In line with the observation that very few researchers hold a PhD in Systems Biology, the answer to this question was mostly (>75%) negative. Thus, likely due to the relative novelty of this research area but also to the reduced training opportunities in the field, most of the researchers involved in SB have not received any specific SB training.

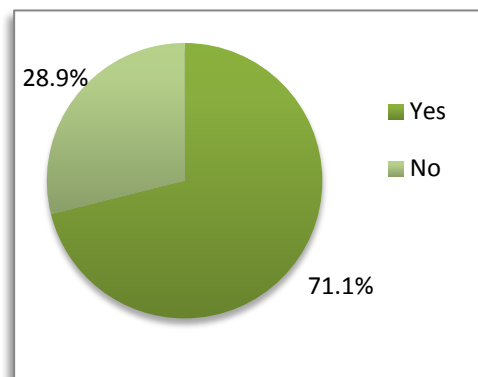
5.9. Does your institution provide training in systems biology?

In contrast to the previous question, however, most of the researchers indicated that their institutions are currently offering training in SB (mainly at the post-graduate level but also at undergraduate level and as advanced training activities). This result indicates that the discipline is highly dynamic and is growing. Most respondents did not have specific training in SB because that was not available at the time, but now training opportunities in SB are more widely available so we can expect future group leaders to have received specific training.



5.10. Are you involved in technology and methodology development (including software development and analysis tools)?

Almost $\frac{3}{4}$ of the respondents stated to be involved in methodology development, confirming the observation reported in 4.4 (analysis of keywords from the Dataset 2) that methodology development is one main activity in the field of systems biology.



5.11. Which scientific societies do you belong to?

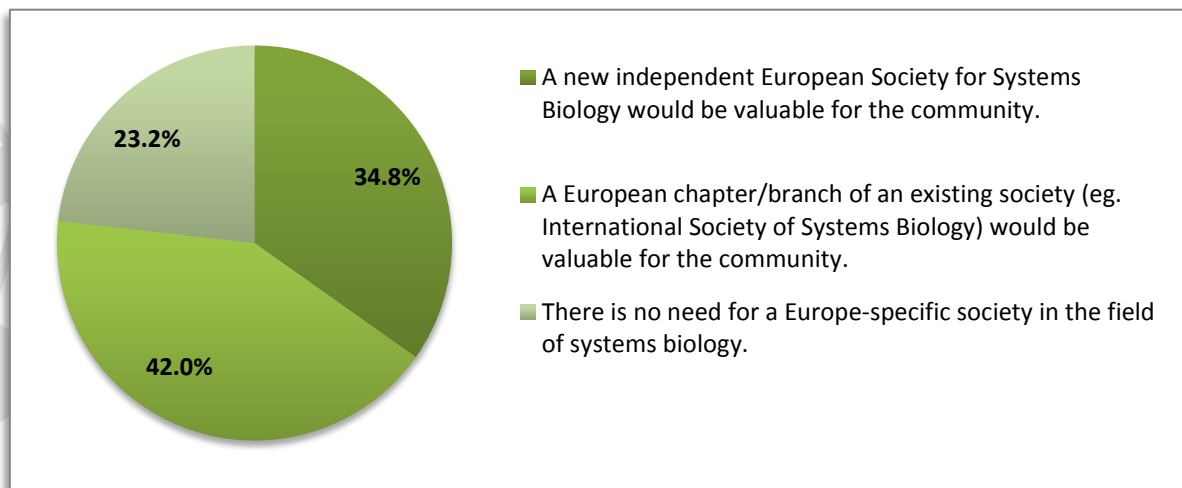
36 respondents (out of 120) stated to belong to one or more societies, listing up to 50 different ones. The only societies mentioned by at least 3 respondents are: ISCB (International Society for Computational Biology, [8]) and ASM (American Society for Microbiology, [3]).

5.12. Please list any other societies which organise meetings relevant to your research:

Regarding societies that organise Systems-Biology-relevant meetings, the 21 researchers that responded identified the following: EMBO (4), ISSB (3), ISCB (3), FASEB (2) and ISMB (2).

5.13. In your opinion, could the creation of a European Society for Systems Biology be valuable to the community?

One of the tasks of ISBE-WP5 (Community Building and Synergies) is to “Set-up a full debate on the possible development of a European SB society, especially for its potential value in lobbying funding agencies, national and European funding programmes (HORIZON2020) and regulatory bodies”. As a first activity towards this objective, we included this question in the questionnaire, and obtained the results represented in the graph. In summary $\frac{3}{4}$ of the respondents appreciate the need or usefulness of a European society for systems biologists, although more than half of them consider that creating a European branch for the International SB Society would be sufficient or more efficient. Taking these results as a starting point, we now aim to get a larger share of the SB community involved in the discussion.



6. Community website: main tool to define and understand the landscape of Systems Biology in Europe

To help in defining the European landscape of systems biology, ISBE promotes the creation of a community website which will serve a number of purposes. As part of its tasks (Deliverable 5.2), WP5 has designed and launched the community website² with the function of:

- Acting as a common point of interaction for everyone in Europe who considers themselves to be a systems biologist
- Allowing for broader discussion, showing how increasing numbers of researchers are becoming part of SB approaches
- Supporting biologists who want to learn more about systems biology

² <http://community.isbe.eu>



- Creating a “Who’s who?” – a database of the European landscape of researchers in systems biology
- Providing information about the resources available in Europe: both technical (i.e. modelling resources) and supporting actions such as conferences, education and training workshops

[The community website](#) was launched in August 30, 2013 during the ICSB2013 in Copenhagen as a Beta version. In the first 2 months the community website received 60 registrations, and it is expected to grow much more after the recent release of a functional upgrade including the option to edit several fields on the researchers’ profiles.

Thus, the community website will be a major meeting point for the SB community, and also the main tool for ISBE to gather a live picture of the European landscape of systems biology.

7. Participation to the ICSB2013 in Copenhagen: opening ISBE to the international community and identifying international state-of-the-art institutions.

WP5 participated to the International Conference on Systems Biology (ICSB) 2013, held in Copenhagen (August 30 – September 3) and organized by the International Society for Systems Biology (ISSB), and took this opportunity to:

- Present the ISBE initiative to the international community:
 - ISBE informative session – luncheon
 - ISBE flyer
 - ISBE Poster
- Gather the opinion of the European and international community on the aim of ISBE
- Identify and network with relevant European and International players (researchers and organizations – ISSB, CASyM, ICSB2014 organizing committee).

Conclusions

This survey summarizes data obtained from different sources and using different methods in order to generate an extensive picture of the European SB landscape. The Web-search approach (Dataset 1) provided detailed information for about 500 leading European SB scientists, offering the possibility for their participation in the ISBE questionnaire. This dataset was complemented by a rather unbiased, publication-based dataset (Dataset 2), providing broader information on a larger group of SB researchers (over 7000). Responses of 120 researchers to the ISBE questionnaire gave highly curated information. These three data-collection activities were complemented with dissemination (i.e. participation in the ICSB2013 in Copenhagen) and community-building (i.e. European Systems Biology Community website) activities with the aim to establish a two-way communication between ISBE and the SB community.

In spite of the diverse nature of the information sources, a number of observations were consistently obtained through the analysis of the datasets. Many of the findings reported here are confirmatory of the expected results but offer for the first time actual figures explaining the complexity of SB (community) at least in the context of the European SB community. Some of the conclusions reached support the idea that SB is a very heterogeneous and rapidly evolving field of research in terms of:

- Expertise involved
- Previous training
- Research topics
- Geographical distribution of expertise

Nonetheless, this survey identifies some common themes that help defining the identity of SB community:

- Focus on models, networks and pathways
- Equilibrated mix of expertise (computer science, maths, and biology)
- Frequent involvement of technology and methodology development
- Increasing offer in specific SB training

Our results suggest that systems biologists mainly orientate their research towards very different fields of fundamental research in Life Sciences, and also towards health-related research. However, other applied areas (energy, environment, food production) are also addressed by systems biologists. The SB community is highly dynamic and rapidly growing, as can be deduced from the increasing availability of specific SB training at the undergraduate and postgraduate levels.

The results of the extensive web-search and especially of the publication screening identified a reduced number of European countries with the higher numbers (in absolute terms) of SB researchers (UK, Germany, followed by Spain, The Netherlands, Italy and



France). To note, SB researchers are found in almost all European countries, showing the wide geographical spread of the researchers in this field. This could be one argument to sustain the creation of a European systems biology society, either as an independent entity or as a branch of the international society, an option supported by most of the questionnaire respondents.

In conclusion, the work of ISBE-WP5 on community building has generated an up-to-date portrait of the European SB community and has established communication paths in order to facilitate the understanding of the capacities and needs of the community. These tools shall be useful in the definition and future construction of the Infrastructure for Systems Biology in Europe.

