ISBE Infrastructure for Systems Biology Europe

WP7

Strategy, Vision and Advocacy

Infrastructure for Systems Biology Europe

Deliverable No: 7.2

A user concept for the ISBE

Main/responsible Author(s): Vitor Martins dos Santos, Babette Regierer
Institution: WUR
Country: The Netherlands

December 2014
(update April 2015)

Vitor Martins dos Santos, WUR
Babette Regierer, WUR
Martijn Moné, VU
Roel van Driel, UVA
Frans van Nieuwpoort, UVA
Natalie Stanford, UNIMAN
Martin Golebiewski, HITS
Adrian Pugh, BBSRC
and representatives of other WPs

Project funded by the European Commission under the Seventh Framework program for Research and Technological Development

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 24
Actual date of delivery	Month 29
Deliverable number	D7.2
Deliverable title	A user concept for the ISBE
Dissemination Level	PU
Status & version	Version 4
Number of pages	27 pages
WP relevant to deliverable	WP7
Lead Participant	WU-ATV (Vitor Martins dos Santos)
Author(s)	Babette Regierer, Vitor Martins dos Santos, Frans van Nieuwpoort, Martijn Moné, Adrian Pugh, Natalie Stanford, Martin Golebiewski, Susanne Hollmann and others
Project coordinator	Imperial (Richard Kitney)
EC Project Officer	Andrea de Candido

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

Background Information on WP7

Objectives of WP7

The objective of WP7 within the ISBE Preparatory Phase is to lay down a strategy, vision and advocacy framework for the implementation of the ISBE in a subsequent phase. This WP aims to attain a community-supported view on the needs, bottlenecks and potential of the various aspects of systems biology for area of research and development.

Deliverable 7.2 is a report on the user concept for the systems biology infrastructure designed by ISBE. Objectives are:

- Fine-tune the strategies developed in WP3 and WP13 of the activities of the physical and distributed centres and to identify and implement possible improvements
- Refine how the specific needs of ISBE users can best be met by an ISBE client concept as initialised in WP3. A strategy, built upon the surveys and assessments, will be devised to ensure that the different user interests are considered at an early stage of infrastructure design

WP7 assisted in a series of Europe-wide surveys and interviews to assess the user needs and the respective framework for measures to respond to the user needs in systems biology.

Relationship of WP7 to other work packages for D7.2:

WP7 aligns closely with the following WPs:

- WP1 Project Management and Co-ordination: WP1 is essential for the integration of results and information of all WPs and coordinates the development of the overall ISBE business case and business plan.
- WP2 Model and Data Management: Considerations around the model and data management are a central topic for the overall ISBE concept, specifically standardization and SOPs have a high relevance.
- WP3 Overall Infrastructure, Eligibility and Accessibility: This WP is the core driver for developing the infrastructure concept. The development of the ISBE user concept mainly depends on WP3 results.
- WP5 Community Building and Synergies: Community building and the establishment of cross-cutting activities with other European infrastructures enhances the development of the user concept.
- WP8 Modelling Infrastructure and Expertise: This WP is essential to provide the information about current modeling approaches and availability of the expertise across Europe
- WP9 Technology and Science Watch: This WP is relevant for the development of a concept how to use and integrate existing repositories, storage of data/information/models etc.

- WP10 Training and Education: Together with WP10, WP7 provides a strategic basis for identification of educational and training needs across the relevant fields in Europe and the development of solutions for the infrastructure concept.
- WP11 Funding, Governance and Legal: WP11 is mainly responsible for the current development of the business case and later the business plan, and is strongly contributing to the user concept from the funder's perspective.
- WP13 Connections: WP13 is working on the connections between communities and user groups and has also an essential role for the development of a user concept framework.

Introduction to an ISBE User Concept

Vision and Mission

The development of a framework for a user concept for the systems biology infrastructure for Europe was core for D7.2. The main guideline for the discussion were the vision and mission statements that set the frame for the overall design of the infrastructure. Below is a summary of the main messages relating to the development of a user concept derived from the vision and mission statements.

Vision:

- The infrastructure needs to empower "…life scientists to understand living organisms to a much higher precision and in a predictive way"
- The infrastructure needs to enable "…scientists in academia, the health sector and industry to exploit the full potential of data-driven computational modelling of complex biological systems with the required reproducibility and validation"
- The infrastructure needs to provide expertise, tools and resources to the users.

Mission:
Via a distributed infrastructure of interconnected national systems biology centres the users get access to

- data, models, biological maps, tools
- support for the generation of model-compliant data and the building and use of data-driven models
- curated and annotated datasets and models
- community standards and best practices
- expertise by training

Background

Areas of application for the systems biology infrastructure

"We are now awash with genomic, proteomic and metabolomics data. The problem is to understand it. Simply accumulating yet more data will not solve that problem". This quote of Denis Noble summarises the need for systems approaches throughout all life sciences (academic- as well as industry-based) to reach a higher level of understanding of biological systems and how to apply knowledge in fundamental biology to generate solutions for improving our lives and to help resolve urgent societal challenges.

In deliverable D7.1 an outline was given for the different areas where systems biology approaches will be relevant and increase in importance to gain a fundamental understanding of biological systems (a summary is given in fig. 1).

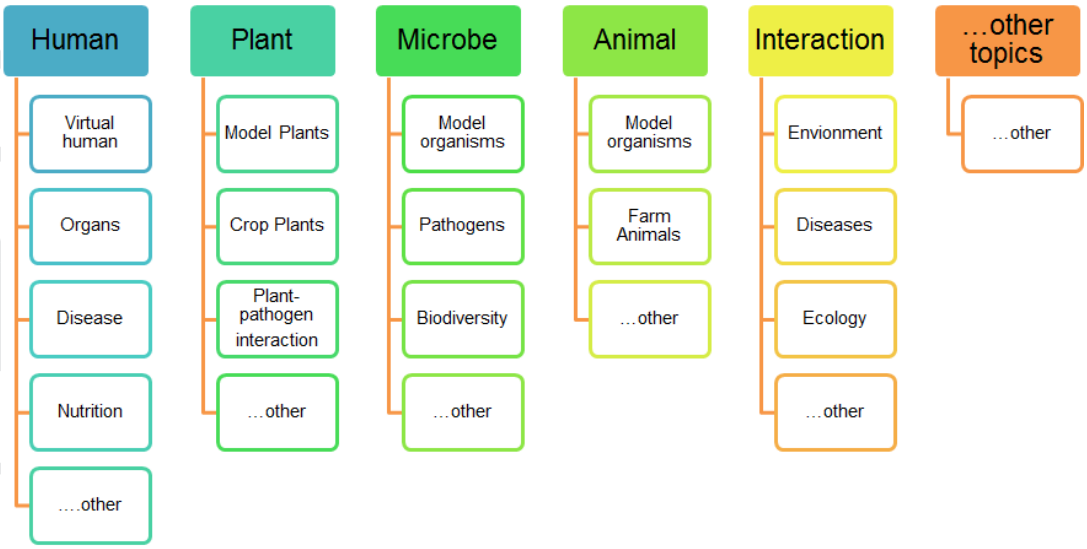

Fig. 1: Various application fields for a systems biology infrastructure. (Source: ISBE Consortium)

Problem Statement

The problem is	…to define optimal services, resources and training and education for the users given a huge variety of disciplines that need to be involved and level of knowledge.
The effect is	…that current and future grand challenges in the life sciences can be addressed by a wider community because the infrastructure for systems biology provides all necessary services and resources to enable the users to include systems biology approaches in their own research and applications.
The impact is	…that Europe develops into a world leader in systems biology by raising awareness and broadening the basis of knowledge within the life science communities (especially user communities) and encourage them to involve systems biology approaches in their research.
A solution would be	…the creation and implementation of a pan-European infrastructure for comprehensive systems biology providing all services, resources, expertise and training needed in the different areas of the life sciences.

For designing and implementing a systems biology infrastructure for Europe, some aspects need to be taken into consideration:

A) All life science areas are different: their scopes, approaches and application of models might vary considerably in different areas. We also have to consider that the availability of data, biological maps and models might be different in the various fields of application.

B) We also have to deal with the fact that not all life science areas have been developing at the same speed into the use and application of systems approaches, for many different reasons.

C) One reason for the difference in the speed of development and uptake of systems biology approaches in the life science areas is the availability of funding for the different areas. Many European countries have defined systems medicine as a high priority research area, therefore provide more funding for this field compared to other like agriculture.

D) The available funding and support for the different life science fields has also an impact on the availability of resources and tools. The field of bioinformatics provides a very good example how the strategic decisions for funding influences the development of resources and tools in specific fields: Currently we observe a big difference between the human and non-human life science fields; for the human field, the number of bioinformatics tools, web services and databases are exceeding by far the number of tools available for non-human life science fields. Not all tools, web services and databases are easily transferable to other areas, and there is an urgent need to build these resources for the non-human life science areas to respond to the needs of the users.

E) For systems biology a variety of different science disciplines is needed; all this expertise must be integrated based on a more systematic approach. The current situation shows that more attention is needed to address this problem as these science fields are not integrated to a full extend and working in silos still hampers the leverage of the existing potential to a full extent.

F) The observed fragmentation between the life science areas and the disconnection between the users and developers of systems biology/modelling is limiting the wide application of the systems biology approaches.

G) The number of life scientists who generate biological data fit for modelling is much larger than the number of experts in data generation and modelling. Across whole Europe, the number of institutions where this knowledge is efficiently combined in one place is still too low to cover the needs.

H) There is a huge variety of modelling approaches, and they vary with the biological system, the research question and the available knowledge and resources. For a non-expert user it is difficult to decide which modelling approach is appropriate and where to find the experts for each.

I) The knowledge level about systems biology approaches varies in the user community from no knowledge at all until a high level expertise.

J) Systems biology approaches could also vary considerably in size and scope:
- small projects proposed by individual researchers or research groups
- large projects proposed by national and international research consortia
- small or large projects for commercial parties

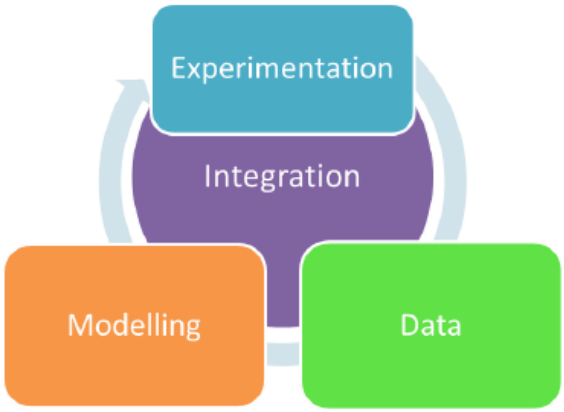

Fig. 2: Systems biology is an iterative cycle of integrating data-driven modelling and model-driven experiment. (Source: ISBE Business Case)

A relevant issue arises around the data generation resp. the use of existing data and model resources. For generating a comprehensive resource of highly curated and annotated data and models that will be provided by the infrastructure to the research communities, the quality of data and models generated for biological systems need to be re-usable. This requires a stringent implementation of standards and SOPs to produce high-quality datasets and an efficient system for stewardship ad continuous curation.

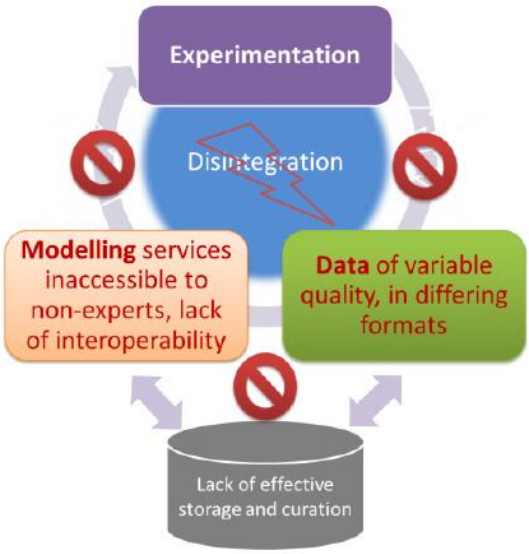

Fig. 3A: Currently, reductionism in combination with lack of modelling services and standards hampers the integration of data and getting insight in how biological systems function. (Source: ISBE Business Case)

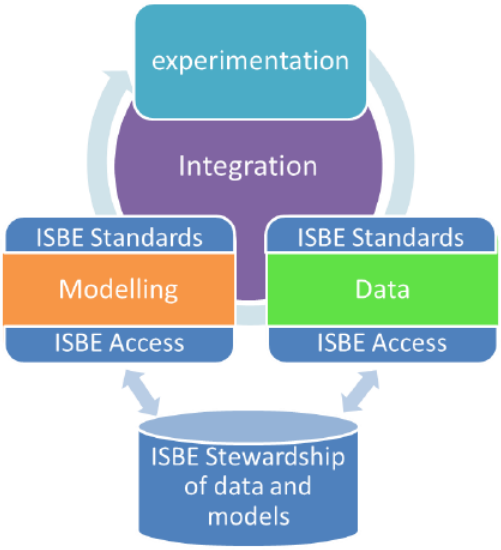

Fig. 3B: ISBE will enable a more efficient and widespread uptake of systems biology research by providing support in computational modelling, model-compliant experimentation, systems biology compliant data, model and SOP stewardship together with the development and implementation of community standards, linked to relevant training and education. (Source: ISBE Business Case)

To fulfil this task the close interaction between the infrastructure national systems biology centres and the user – and developer – communities will be essential. This task is adding a new dimension to the infrastructure-user relationship.

An appropriate user concept for the planned system biology infrastructure needs to respond to these challenges.

The approach: Persona Modelling

Description Persona Modelling

The ISBE approach to define a user concept for the systems biology infrastructure needed to consider the full iterative cycle of experimentation, data generation/use of existing data resources, modelling and the overall integration (see fig. 2). The challenge is to provide a sound basis to enable users to exploit all resources of the full cycle for systems biology approaches. All aspects that are needed to perform

systems biology research need to be integrated in a consistent approach along this cycle. Experienced users, but also new users should have full access to the resources and expertise for systems biology.
It is desired for the ISBE user concept to encompass the diversity of the relevant expertise, the different life science fields and also bridge the gap between user and developer community. Therefore, the ISBE user concept must designed to respond to the user, as well as the developer/provider community at all scales of expertise level.

The essential requirement to design an appropriate user concept is to gain inside what is the user perspective. The user perspective must include the perspective of all possible user groups that will potentially approach the infrastructure. A useful method to capture the user perspective is "persona modelling". Originating from economics where it is used to describe precisely the customers in their environment, background, their needs and limitations, we applied this method to potential user groups of the systems biology infrastructure.
EMBL-EBI uses persona modelling to optimise their services in bioinformatics or establish new services in accordance to the needs of the customers from academia and industry. As EMBL-EBI has already many years of experience, Dominic Clark and Jenny Cham from the EMBL-EBI in Hinxton provided an introduction into the method and guidelines and recommendations for the application in the ISBE context.
Henny Cham introduced Persona Modeling as an instrument to align the planned new services better to the user needs. A virtual persona helps to understand in detail the situation and context of potential users. The personas are fictitious, but would be based on examples from reality. Here is a short guideline how the personas are used to understand the customer perspective better:
Persona modelling:
(i) Personas

- Persona identification

Personas help to understand the customers (ii) Tool to turn abstract "users" into real people with real needs
(iii) They represent:

- Goals
- Behaviours
- Motivations of real users
- Provide empathy which is useful for designing services/
- products to better meet the needs of your users

Persona modelling is a structure and consecutive and iterative process that allows us to drill down level by level to fully understand the user within their context.

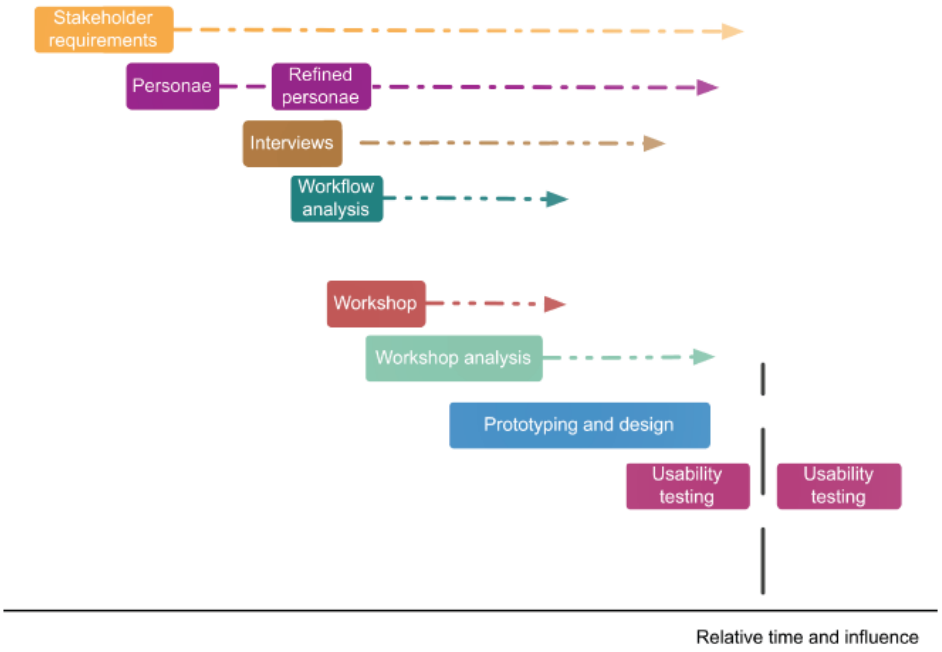

Fig. 4: Persona modelling is a process of refinement steps to fully understand the user needs and the respective framework conditions. (Source: Jenny Cham, EBI)

Persona Modelling Approach for ISBE

The ISBE persona models were obtained via a cross-WP workshops during three meetings: 17-18 February 2014, 19-20 March 2014, 22-23 April 2014. Participating WPs were: WP2, WP3, WP7, WP8, WP10, WP11 and WP13.

As a first step, we addressed a specific user group that has already defined expertise in modelling. The requirement for this user group was to extend their knowledge to include new modelling approaches in their work.

The driving question was why this group would like to use the infrastructure?

- This user group is not an expert in a certain model they want to apply for their data
- They have no access to experts in their near environment or they do not want to have shared authorship (that would exclude a direct cooperation)
- They need access to knowledge about modelling approaches they did not use so far
- They need access to material, tools, web services and curated models as examples

The first pilot and as a starting point for the persona modelling exercise we chose the fictitious persona "Neil" representing the above described user group. "Neil" was described with the following attributes:

- Experienced in systems biology
- Has a PhD/is a postdoc
- Has been working already in the systems biology context

As we are addressing here an experienced user group, the main service for this user group provided by the infrastructure would be training in the new modelling approaches. Working along this line, we could assign the need for **training** and education as a key driver for the infrastructure. The more detailed discussion resulted in the finding that with increasing intensity of training there will be a transition from training into consultancy as the intensity of expert knowledge is increasing as well, needed for direct application.

The main key word that must be the driver for the ISBE user concept is "enabling" of the users to perform systems biology projects (fig. 6).

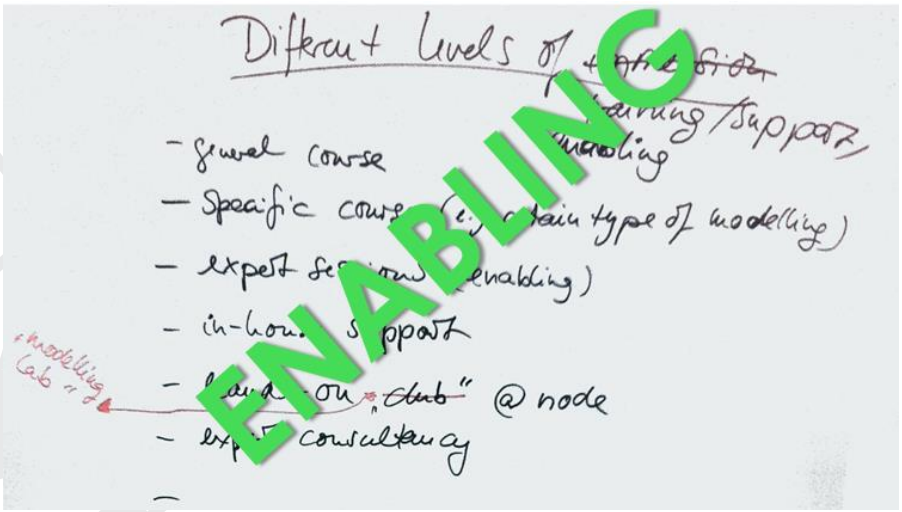

Fig. 6: The main element of the infrastructure service concept is to enable the user to perform high quality systems biology research.

The next step to obtain more information about potential ISBE users via persona modelling we performed 7 interviews (see next paragraph).In addition, we formulated potential scenarios in which multiple stakeholders would require working together (e.g. projects, consortia) that need to be looked into to explore if these type of user groups would require different activities apart from the personas we interviewed. Several persona and scenario categories have previously been proposed from discussions during plenary WP13 meetings:

Potential Personas

- SB expert, i.e. PhD students/postdocs working on systems problem
- Idem, but non-expert in SB
- Stakeholder from industry
- national systems biology centre (nSBC) manager (that needs to provide ISBE services)
- PI in research group that wants to use ISBE
- Scientist in nSBC
- Governance persona (quality control / compliance etc.)

- Non-SB end user (e.g. clinician)
- Funding persona

Potential Scenarios

- Predicting biomarkers upon drug administration to a mouse and validating this experimentally.
- Develop a personalized therapy for an inborn error of metabolism based on Recon2 map.
- Engineer a yeast cell to robustly produce octanol.

Developing personas

Given the broad range of potential personas it was decided to commence from the more prevalent and readily accessible persona types. The initial constants chosen were:

- position: postdoc/PhD student
- environment: academic or SME
- background: (at least some) modelling experience

Personas are initially shaped based on mapping their requirements (i.e. for ISBE: what they would seek in terms of systems biology), which can be reinforced through interviews with real people that resemble the persona. The finalized personas are typically amalgamations, which can then be subjected to characteristic workflow analyses. This process can be reiterated through sub-personas (similar personas with some key parametric change(s) – level of experience, presence/absence of resources, etc.).

Stakeholder requirements

Several relevant persona-like researchers (i.e. postdoc or PhD student that has some systems biology background and that needs some systems biology done) were available and therefore we assessed stakeholder requirements through direct interviews. The results were basically a coarse breakdown of the person's research environment (i.e. what expertise or resources were available) and their subsequent needs assessment (i.e. what would significantly have supported the project if ISBE was operational).

Interviews with Personas

Interviews were conducted with different prototypes of "personas"; besides the academic context, also interviews on the perspective of SMEs was included. The summary of the interviews is given below.

Persona "Natalie1"

"Natalie 1" working in a systems biology group.

- PhD student
- academic – systems biology institute

Available (skills, expertise & resources):

- Rich data
- Sufficient equipment available
- Laboratory experience
- Limited modelling skills
- Expert modelling environment
- Experiment to modelling well-linked

Needs:

- Computational training
- Standards
- Data structuring solutions
- Long-term data storage
- Need for computing power
- Knowledge of downstream processes

Remarks:

- More studying network behaviour than consistent project/research questions

Persona "Natalie2"

"Natalie 2" working at non-systems biology institute.

- Postdoc
- Academic

Available (skills, expertise & resources):

- Computer

Needs:

- Modelling software (no licensed software available)
- Lab expertise (little locally available)
- Data
- Money (no travel, no training)

Persona "Natalie3"

"Natalie 3" working in an SME providing services in the field of systems biology.

- Postdoc
- SME

Available (skills, expertise & resources):

- Equipment
- Computers
- Access to people and expertise

Needs:

- Modelling software (no MatLab)
- Analytics
- Data
- Access to databases
- Training
- Standards, SOPs
- Computing power

Persona "Daniel1"

"Daniel 1" working in a systems biology institute.

- PhD student
- Academic

Available (skills, expertise & resources):

- Bioinformatician
- Modest experimental experience
- Interface between experiment and theory
- Data access
- Software (though limited by project duration)
- Computing power
- Data management

Needs:

- Software (licenses limited)
- Access to experts/expertise
- Solutions to identify experts (seemed impossible)
- Standards, SOPs
- Training – coherent training programme

Persona "Daniel2"

"Daniel 2" working in a management position in a systems biology institute.

- Postdoc
- Manager/informatician
- Academic

Available (skills, expertise & resources):

- Bioinformatics
- Computer science
- Data access
- Software (open access)

Needs:

- Tools (no flexibility in choice of toolsets)
- Training – biological knowledge

Persona "Martin"

"Martin" working as a postdoc in a systems biology institute.

- Postdoc
- Bioinformatician (interface between experiment and models)
- SME (non-profit)

Available (skills, expertise & resources):

- Bioinformatics
- Basic funding (allows long-term projects)
- Travel money
- Collaborations (with academy and industry)
- Good computing facility

Needs:

- Data (no databases, no lab)
- Standards
- Training – modelling
- Training – experimental expertise

Persona "Matthias"

Matthias is a PhD student in a systems biology group, and is part of the consortium VLN. Matthias created a model for diabetes and performed simulations based on existing data.

- PhD student (end phase)
- Modelling expertise
- University, combination with clinics

Available (skills, expertise & resources):

- Bioinformatics
- Project funding, large consortium
- Travel money
- Collaborations within the VLN
- Access to computing facility, software, expertise etc.
- Only limited access to data (often produced very late/too late in a project); therefore models are often developed using existing data(bases)

Needs:

- Data (databases)
- Standards
- Missing is an easy-to-use standard solution for datasets and metadata description; existing solutions are too complicated and often not compatible if different datasets are used for modelling
- Access to storage
- Training must be oriented on real-world experience, best would be training on own datasets
- Support by video chats or tutorials etc. would be very helpful; courses with presence for 2-3 days only in exceptional cases reasonable for experts (or semi-experts)
- Data management is very important; it must be an easy-to-use systems with possibility for interoperability with many others and the added value must be clear for data management

- VLN: specialised solutions are available, but these are sometimes too complicated and take too long (e.g. SEEK); it would be very helpful to have a simple standard format (spreadsheets?) for data and metadata (independent from origin and type of data)

Remarks:

- Also mentioned as important element: before project start it would be beneficial for all participants if there was a meeting at the beginning with all partners to agree on strategies for data generation, data management and how to process and integrate the information including standards etc.

Persona "Johannes"

Johannes is founder and CSO of an SME for data management and bioinformatics, the company exists since more than 10 years; participates in CASyM.

- Bioinformatician (including data management, text mining etc.)
- No data generation
- SME

Available (skills, expertise & resources):

- Bioinformatics, data management
- Experience as partner in various projects (national and EC)
- Access to computing facility, software, expertise etc.
- Only limited access to data (often produced very late/too late in a project); therefore models are often developed using existing data(bases)

Needs:

- Standards (1 standard "fits all" is not possible)
- Very important: access to the information on standards and which standards are currently in development
- Training
- Viable business models in systems biology; potential business fields for SB: e.g. diagnostics, biomarker development, fermentation (synthetic biology)
- Methods for data management, and training about this (via tutorials etc.)
- Essential for modelling: understanding of the underlying biology (e.g. which parameter describe the biological system?)

Remarks:

- What is the relationship to ELIXIR?
- SME have often a relatively short half-life; that makes it complicated to build business models for SMEs
- Who is paying for the implementation and sustainable organisation (of standards, training etc.)?
- For modern data management/systems biology: as the problems get more and more complex, it needs much more interaction and involvement of the customer (no "one approach fits all" solutions in systems biology); but this interaction is very difficult in terms of communication and time.

Four more "persona profiles" are presented below that have been produced that are also included in the ISBE Business Case and include an even wider range of potential user groups not yet addressed in the previous approaches.

Persona Model 1: ISBE and the researcher

'Sarah' is the leader of a Computational Biomedicine group based in the United Kingdom. She is looking to model the changes in iron metabolism within cancerous cells. The project requires generation of 6 different data sets (a mixture of high throughput and single cell analysis) which Sarah does not have the expertise for in her group. The expertise for producing the data is distributed across 3 different nSBCs, and the data is legally sensitive. Sarah also wants to couple her model with an already available ISBE cell cycle model.

What ISBE will offer: The raw data is collected, structured, and annotated according to available and agreed SOPs in two of the nSBCs. The raw data is then stored in an embassy cloud, to be accessed and post-processed by the third nSBC, according to relevant SOPs, into sharable formats (structured and annotated according to community and ISBE defined minimal standards). The share-format data is loaded into ISBE–related repositories, and made available privately (length defined by user/ISBE/legal requirements) to Sarah in a data-unified interface.

The model is constructed by Sarah's group through consultation with her local nSBC to ensure that its structure and format is compatible with the cell cycle model Sarah wants to integrate it with. After the full model is constructed and integrated with the cell cycle model, it is uploaded into a relevant ISBE model database where it can be kept private, or shared with collaborators until publication. At the point of publication the model and data are made available to the public subject to legal restrictions governing the data. The model is curated such that all data can be directly linked and identified with model components.

Impact: 5 sets of high quality data are released into the public domain, and are available for other projects to use, subject to legal restrictions. Provenance of the data and model are available and will be tractable through the lifetime of the data and model. The public can access the model and simulate it using ISBE simulation services. Other researchers can (re-)use the data and model for their own research, and satellite work based on this work will be tractable by the community. Sarah's group can be credited for their input into new projects.

Persona Model 2: ISBE and the national research council

A national research funding (NRF) body wants to ensure that the systems biology research it funds has the highest impact possible both in Europe and globally. They have identified that one of the key weaknesses in long-term asset storage from their funded projects is accessibility and (re-)usability. They want to devise a strategy to be implemented on all future funded projects that will overcome these issues.

What ISBE will offer: The NRF can consult with ISBE about its requirements for future systems biology projects. Data handling frameworks will be established between funder and ISBE, and a full set of recommendations for data and model formatting, annotation, and storage will be defined and made available for reference by holders of future successful grants. Training courses can be designed by ISBE and made available voluntarily, or mandatorily to future grant holders.

Impact: When funding projects with public money, especially those with large budgets, it is vital that all assets of suitable quality are made available to the public. By establishing data management and stewardship practices early, and making this a requirement to researchers it improves the likelihood that funded research will achieve higher impact. The development of suitable training made available to grant holders increases the likelihood of the practices being followed correctly. A centrally managed framework means that groups do not have to waste time and resources developing their own formatting, annotation and storage procedures, and therefore reduces the burden and the cost to the researchers whilst allowing the funder to achieve its goals.

Persona Model 3: ISBE and the citizen

'Joe' is diabetic and as an avid amateur biologist is interested in how his blood sugar level impacts the metabolic behaviour of his organs.

What ISBE will offer: The Consensus Human Diabetes Model is stored in a standardised format in an ISBE managed model database. The database is searchable using key-words allowing Joe to find the model quickly. The model has several associated links including the open-access paper it was published in - with a public summary, the patient data that was used to build the model, and services for simulating the model. After reading the paper Joe can understand the basics about what the model does. After launching the simulation, he alters the blood glucose levels through many different ranges. After spotting some clear changes in behaviour, he uses identifiers in the model that link to external resources, in order to understand their function. Joe soon discovers the wide-reaching impact that deviations in his blood sugar levels can have over the short and long-term. He signs up to receive automatic notifications for when the model is updated.

Impact: An open, well managed, and easily accessible infrastructure is not just useful for research scientists; it is also a powerful resource for the enquiring public. The careful storage, annotation, and linking of resources within ISBE has allowed someone with little expert knowledge to gain access to information that impacts their understanding of a common disease.

Persona Model 4: ISBE and the scientific journal

'Systems Biology at Multi-Scale' is an open-access journal dedicated to publishing the growing number of multi-scale models developed within the systems biology community. They have strict policies for publishing models: (i) All data used to construct the model must be available in the public domain, fully annotated to ensure reproducibility, and directly traceable to and from the model; (ii) all models must be publicly available, structured and annotated according to community standards, and simulatable for (re-)use by the community; (iii) the model must be able to reproduce all the finding in the paper; (iv) the data and model must be guaranteed to be available, and (re-)usable, in the public domain for at least 10 years post-publication.

What ISBE will offer: The Journal can work directly with its local nSBC to turn the requirements into a functional set of formats and annotations for authors to follow. The nSBC can train staff from the journal in data and model curation, submission and interlinking. ISBE can provide temporary data and model areas that are private for reviewers to access. Upon publication the data and models will be referred to the trained journal staff who can ensure the formats, and metadata standards of the data and model are suitable, that acceptable cross linking is present, and that the model produces the findings in the paper correctly. This is then submitted to permanent, publicly accessible (subject to any legal restrictions) storage facilities, where the model and data can be viewed in a unified interface. The data will be stored there for a minimum lifetime of 10 years required by the Journal.

Impact: Journals want to publish high impact, highly cited research. A barrier to this is often the lack of availability of the datasets and models included in journal papers. Poor availability of these assets prevents other researchers assessing the quality of the research, and also being able to use the research to build on within their own work. This will reduce the impact of the research on the community to the detriment of the journal.

How will the ISBE consortium answer to the needs of users?

The user spectrum of ISBE is broad, and services need to be designed in accordance to the needs of these different groups. The potential user communities of ISBE include academia, hospitals, clinics, industry, and representatives from other interest groups relating to e.g. agriculture like the European Technology Platform: Plants for the Future (www.plantetp.org). Derived from the persona modelling exercise, the interviews and related other WP activities the following major issues have been identified that form a framework for the user concept for the systems biology infrastructure.

- Modelling resources and services
- Stewardship
- Standardisation
- Access to model-compliant data and tools, software and expertise
- Education, Training and consultancy

Aligned with the needs and comments from the interview and persona modeling exercise, the ISBE core concept foresees support of its users in their research through the following activities (described as in the ISBE Business Case document):

- Services in three tightly linked domains: modelling, model-compliant data generation, stewardship and standardisation.
- Diverse data and model exchange in a standardised manner that is accessible and understandable between all sectors.
- Access to data and models to ensure that researchers can progress the systems biology life-cycle by becoming an ISBE user.
- Developing and maintaining tools and software that makes it easier to model biological data.
- Meeting future-proof operational needs of systems biology.
- Developing and offering an education and training programme in systems biology.

Training is the key

The following list provides a systematic overview about the various levels of training that the infrastructure should provide for the variety of user groups and level of expertise we need to consider:

1. **Level: level of expertise**

- Basic training course
- Middle level
- High level, expert training
- Training overlap with consultancy: hands-on, visiting expert lab to develop the model

2. **Level of career**

- Students
- PhD students

- Young researchers
- PIs

3. Target group
- Individuals
- Groups
- Projects/consortia
- Communities

4. How to perform the training
- Blogs, online help by developers/experts
- Tutorials, manuals
- Courses with presence in a course
- Hands-on training on own material/model with experts (her again the overlap between training and consultancy)

5. Academia and/or industry
- Maybe a specific training is also needed for the different target groups?

Consultancy

From the interviews, but also from a joint discussion with the project "AllBio - Broadening the Bioinformatics Infrastructure to unicellular, animal, and plant science" (www.allbioinformatics.eu) we realized that the lack of careful experimental design and partner organization is often an underlying problem in cooperation projects, especially when these consortia are working in an interdisciplinary approach as it is important for systems biology. In systems biology this problem is even more prominent and causes problems when data generated during the project time need to be integrated to build a model. Experts estimate that most of the data generated in life science projects are not of sufficient quality for integration and are not re-usable due to the lack of careful experimental design and missing implementation of standards, appropriate statistics etc.
In this context, consultancy will not only be a desired service that the systems biology infrastructure needs to provide, but also consultancy at the beginning of a project would enormously contribute to the efficiency of research projects and ensure that the data generated in the projects are re-usable and can be shared with the broader community. This consultancy for the organisation and experimental design at the beginning of projects could in turn also support the wider adoption of standards throughout the communities. Funders will benefit from such targeted consultancy as they will get higher quality outcomes and results for their investment.
This element also needs to be integrated in the user concept for ISBE.

Funders might also become customers for the systems biology infrastructure for consultancy in strategic questions and issues related to data management, storage, maintenance and availability of data to the scientific community.

Capacity Building and National Smart Specialisation Strategies

The training concept also needs to consider capacity building across all areas and include all stakeholder groups. Very important issue to address here is to take into consideration the situation in the European Member States. Not all Member States have the same level of development, some countries have already established centres for systems biology representing excellently nSBCs for the infrastructure, other countries are lagging behind this development and need specific measures for capacity building in their countries.
It would be highly recommendable to align the efforts in capacity building to the national structural funds strategies (ESIF) and also to the Smart Specialisation Strategies (RIS3) in each country to increase efficiency.

Storage and Accessibility of Data and Models for Other Stakeholder Groups

(Leadership WP2/WP8)

Based on the persona modelling activities it became clear that also funders, publishers, but also interested laypersons would be future users of the systems biology infrastructure. Needs resulting from the persona modeling and discussions with stakeholders are:

- the storage of data, models and information and make them accessible for a broad range of users, including non-scientific user groups
- service for publishers to store data and models connected to publications

The service for publishers might develop into an essential service that the systems biology infrastructure in cooperation with ELIXIR might need to provide in the near future. In a workshop on standardization which took place on 18 September 2014, the Chief Editor of Molecular Systems Biology, Dr. Thomas Lemberger, presented the upcoming challenges for publishers dealing with large datasets and models:
Journals/publishers are currently challenged by fundamental changes in publishing strategies. They need to deal with the problem that today large datasets are the basis for the publications and increasingly models are included in the publications.
A major need is to reorganise the process how to share research data with the users/research community. There are two ways: Research data are deposited in databases where the users can access the information. Research data are also published in journals and made available via this route. The question is how these two ways can be linked to each other: research data -> journals -> databases -> users – and how can the systems biology infrastructure support this?
The systems biology infrastructure might play an essential role in data deposition (how to structure data? Which information is important?) and support the development of a new system for data citation (provide credit for data providers; how to allow citation of datasets/databases?). The models or figures for journal articles are normally based on datasets deposited in remote places. A new way to present figures is the deposit of the raw data in a common repository where also metadata and additional relevant information is linked to the raw data. To publish research results based on a standardised process to deposit ad store raw and metadata would make research results from publications much better accessible for the whole research community. The development, establishment and maintenance of such a data repository could be a future service package of the systems biology infrastructure in cooperation with ELIXIR.

Another essential service that systems biology infrastructure needs to provide for the scientific community is the curation of models that are published, either via journal articles or by uploading to repositories like BioModels and JWS.

Standardisation

(Leadership WP2)

Standardization is a highly important topic for ISBE because standards and SOPs enable researchers to exchange and integrate their data and resulting models and to relate corresponding data to each other. In addition to the results produced in WP2 we organised workshops and activities around the question how to connect the users to the standardisation world.
A workshop was organised together with the COST Action SeqAhead (www.seqahead.eu) and NORM-SYS (www.normsys.de), a new project funded by the German Federal Ministry for Economic Affairs and Energy. This workshop took place on 9-10 July 2014 where a mixed group of participants comprising of members from the developer side as well as from the (to a large extent unexperienced) user side. The aim of the workshop was to discuss which stakeholders need to be involved to address the challenges in standard development and adoption of standards and SOPs in the communities.

The discussion was leading to more general considerations what are community needs:

- a change in attitude towards acceptance of standards in the scientific communities
- Publishing of developed standards can be highly rewarding and connected with a high citation rate (= attractiveness to participate in the development)
- Strategic alliances must be established, e.g. with funders, industry, public bodies
- Important question is how we get the commitment and engagement from institutions to participate resp. let the researchers participate in these standardization activities?
- It is needed to establish the right timing and the pipelines for the development and the implementation of standards
- Funding must be raised for the development of standards, but also for the adoption and implementation of standards in the institutions

A major question for ISBE and related systms biology projects and initiatives is how we can serve these end-users

- We need a clear definition of requirements
- We need the acceptance of standards
- Certification could link to bodies and promote career development
- We need training
- We need to promote the inclusion of standards in publications
- We need to promote a new career path towards curators or similar
- We need to influence calls to raise funding for the development of standards (important work, but almost no funding available for this type of work for the service for the whole community)
- We need accreditation systems
- We need to intensify the contact to the community of biocurators
- What could be the role of infrastructures? In principle, infrastructures (especially European infrastructures) could be an excellent ally for the engagement of user communities, for the

validation and promotion of standards; how could we strategically engage with the infrastructures in Europe? Which ones are relevant? E.g. BioMedBridges, ELIXIR

As a summary, it was concluded that the field of standardisation is facing many challenges that need long-term engagement and leverage of resources (time and money). Different stakeholder groups need to be engaged in the activities to allow an efficient and fast process of standardisation and adoption of the developed standards. Infrastructures in Europe appear to be excellent hubs for promoting the communication between different groups involved (e.g. developers of standards and end-users), and also for the validation and adoption of standards.
Here, a joint efforts is needed especially between ISBE and ELIXIR to address the (user) needs, develop a strategy to promote adoption of standards in the life science communities and how to promote the development and maintenance of necessary standards in a long-term perspective.

A joint effort to address user needs: CORBEL - Coordinated Research Infrastructures Building Enduring Life-science services

(Leadership WP1/WP5)

Together with 11 BMS RIs, ISBE under the leadership of WP1 and WP5 developed a project application for the Horizon 2020 call "Implementation and operation of crosscutting services and solutions for clusters of ESFRI and other relevant research infrastructure initiatives" (INFRADEV-4-2014/2015).
The consortium developed a strategic plan how to better serve the user needs via the infrastructures and how to combine efforts to avoid duplication, fragmentation and allow a very convenient access for the user to all BMS-RI infrastructures. The application had the title: "CORBEL - Coordinated Research Infrastructures Building Enduring Life-science services".

CORBEL combines the expertise and effort of:

- BBMRI
- EATRIS
- ECRIN
- ELIXIR
- Infrafrontier
- Instruct
- EU-OPENSCREEN
- EMBRC
- Euro-BioImaging
- ISBE
- MIRRI

The objectives of the proposal are:

1. Forge effective partnerships with user communities.
2. Develop unique solutions to user needs.
3. Implement a portfolio of generic, shared services.

The impact is expected to affect different groups:

- Users, no matter which infrastructure they first approach, will benefit from a coherent set of processes and collaborative interfaces that support access and use of advanced research tools, samples and facilities.
- Funders and other stakeholders will benefit from effective engagement mechanisms and overview of cost-effective research infrastructure services, harmonised legal and ethical expert services.
- Society will benefit from a focused, pan-European effort that supports industry engagement and delivers innovative treatments from scientific discoveries through to applications in health and patient care.

The results of the community efforts will contribute considerably to the development of a highly efficient infrastructure for systems biology.