



## OPINION ARTICLE

**REVISED** **Systems Biology in ELIXIR: modelling in the spotlight**  
**[version 2; peer review: 1 approved, 2 approved with reservations]**

Vitor Martins dos Santos<sup>1</sup>, Mihail Anton<sup>2</sup>, Barbara Szomolay<sup>3</sup>,  
Marek Ostaszewski<sup>4</sup>, Ilja Arts<sup>5</sup>, Rui Benfeitas<sup>6</sup>, Victoria Dominguez Del Angel<sup>7</sup>,  
Elena Domínguez-Romero<sup>8</sup>, Polonca Ferk<sup>9</sup>, Dirk Fey<sup>10</sup>, Carole Goble<sup>11</sup>,  
Martin Golebiewski <sup>12</sup>, Kristina Gruden<sup>13</sup>, Katharina F. Heil <sup>14</sup>,  
Henning Hermjakob<sup>15</sup>, Pascal Kahlem <sup>16</sup>, Maria I. Klapa <sup>17</sup>, Jasper Koehorst<sup>18</sup>,  
Alexey Kolodkin <sup>19,20</sup>, Martina Kutmon <sup>5,21</sup>, Brane Leskošek<sup>9</sup>,  
Sébastien Moretti <sup>22</sup>, Wolfgang Müller<sup>12</sup>, Marco Pagni<sup>22</sup>, Tadeja Rezen<sup>23</sup>,  
Miguel Rocha<sup>24</sup>, Damjana Rozman<sup>23</sup>, David Šafránek <sup>25</sup>, William T. Scott <sup>18,26</sup>,  
Rahuman S. Malik Sheriff <sup>15</sup>, Maria Suarez Diez <sup>18</sup>, Kristel Van Steen<sup>27,28</sup>,  
Hans V Westerhoff <sup>20</sup>, Ulrike Wittig <sup>12</sup>, Katherine Wolstencroft <sup>29</sup>,  
Anze Zupanic <sup>13</sup>, Chris T. Evelo <sup>21</sup>, John M. Hancock <sup>23</sup>

<sup>1</sup>Laboratory of Bioprocess Engineering, Wageningen University & Research, Wageningen, 6708 PB, The Netherlands

<sup>2</sup>Department of Biology and Biological Engineering, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Chalmers University of Technology, Gothenburg, SE-41258, Sweden

<sup>3</sup>Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, UK

<sup>4</sup>Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, L-4367, Luxembourg

<sup>5</sup>Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, 6200 MD, The Netherlands

<sup>6</sup>National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

<sup>7</sup>Research and Platforms Department, Genopole, Evry-Courcouronnes, 91030, France

<sup>8</sup>Department of Bioinformatics, Maastricht University, Maastricht, The Netherlands

<sup>9</sup>Faculty of Medicine, Institute for Biostatistics and Medical Informatics, Centre ELIXIR-SI, University of Ljubljana, Ljubljana, SI-1000, Slovenia

<sup>10</sup>Systems Biology Ireland, School of Medicine, University College Dublin, Dublin, 4, Ireland

<sup>11</sup>Department of Computer Science, The University of Manchester, Manchester, M13 9PL, UK

<sup>12</sup>Heidelberg Institute for Theoretical Studies - HITS, Heidelberg, 69118, Germany

<sup>13</sup>Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, SI-1000, Slovenia

<sup>14</sup>ELIXIR Hub, Hinxton, Cambridge, CB10 1SD, UK

<sup>15</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, CB10 1SD, UK

<sup>16</sup>Scientific Network Management SL, Barcelona, 08015, Spain

<sup>17</sup>Metabolic Engineering & Systems Biology Laboratory, Institute of Chemical Engineering Sciences, Foundation for Research & Technology - Hellas (FORTH/ICE-HT), Patras, 26504, Greece

<sup>18</sup>Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, 6708WE, The Netherlands

<sup>19</sup>Competence Center for Methodology and Statistics; Transversal Translational Medicine, Translational Medicine Operations Hub, Luxembourg Institute of Health, Strassen, L-1445, Luxembourg

<sup>20</sup>ISBE.NL, VU University of Amsterdam, Amsterdam, The Netherlands

<sup>21</sup>Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, Maastricht, 6200 MD, The Netherlands

<sup>22</sup>SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>23</sup>Faculty of Medicine, University of Ljubljana, Ljubljana, SI-1000, Slovenia

<sup>24</sup>Centre of Biological Engineering, University of Minho, Braga, Portugal

<sup>25</sup>Faculty of Informatics, Masaryk University, Brno, 602 00, Czech Republic

<sup>26</sup>UNLOCK, Wageningen University & Research, 6708 PB Wageningen, The Netherlands

<sup>27</sup>BIO3 - Laboratory for Systems Medicine, Department of Human Genetics, KU Leuven, Leuven, 3000, Belgium

<sup>28</sup>BIO3 - Systems Genetics, GIGA-R Medical Genomics, University of Liege, Liege, 4000, Belgium

<sup>29</sup>Leiden Institute of Advanced Computer Science, Leiden University, Leiden, 2333 CA, The Netherlands

**V2** First published: 07 Nov 2022, 11(ELIXIR):1265  
<https://doi.org/10.12688/f1000research.126734.1>

Latest published: 31 May 2024, 11(ELIXIR):1265  
<https://doi.org/10.12688/f1000research.126734.2>

## Abstract

In this white paper, we describe the founding of a new ELIXIR Community - the Systems Biology Community - and its proposed future contributions to both ELIXIR and the broader community of systems biologists in Europe and worldwide. The Community believes that the infrastructure aspects of systems biology - databases, (modelling) tools and standards development, as well as training and access to cloud infrastructure - are not only appropriate components of the ELIXIR infrastructure, but will prove key components of ELIXIR's future support of advanced biological applications and personalised medicine.

By way of a series of meetings, the Community identified seven key areas for its future activities, reflecting both future needs and previous and current activities within ELIXIR Platforms and Communities. These are: overcoming barriers to the wider uptake of systems biology; linking new and existing data to systems biology models; interoperability of systems biology resources; further development and embedding of systems medicine; provisioning of modelling as a service; building and coordinating capacity building and training resources; and supporting industrial embedding of systems biology.

A set of objectives for the Community has been identified under four main headline areas: Standardisation and Interoperability, Technology, Capacity Building and Training, and Industrial Embedding. These are grouped into short-term (3-year), mid-term (6-year) and long-term (10-year) objectives.



## Keywords

Systems Biology, Systems Medicine, ELIXIR Communities, Biomolecular Models, Network Biology, FAIR, Biological data, Biotechnology

## Open Peer Review

Approval Status ? ? ✓

	1	2	3
<b>version 2</b> (revision) 31 May 2024			
<b>version 1</b> 07 Nov 2022	? <small>view</small>	? <small>view</small>	✓ <small>view</small>

1. **Yoram Vodovotz**, University of Pittsburgh, Pittsburgh, USA
2. **Herbert Sauro** , University of Washington, Seattle, USA
3. **Ioannis Androulakis** , Rutgers, The State University of New Jersey, New Brunswick, USA

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Bioinformatics** gateway.



This article is included in the **ELIXIR** gateway.

**Corresponding authors:** Vitor Martins dos Santos ([vitor.martinsdossantos@wur.nl](mailto:vitor.martinsdossantos@wur.nl)), John M. Hancock ([john.hancock@elixir-europe.org](mailto:john.hancock@elixir-europe.org))

**Author roles:** **Martins dos Santos V:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Anton M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Szomolay B:** Writing – Original Draft Preparation, Writing – Review & Editing; **Ostaszewski M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Arts I:** Writing – Original Draft Preparation, Writing – Review & Editing; **Benfeitas R:** Writing – Original Draft Preparation, Writing – Review & Editing; **Dominguez Del Angel V:** Writing – Original Draft Preparation, Writing – Review & Editing; **Domínguez-Romero E:** Writing – Original Draft Preparation; **Ferk P:** Writing – Original Draft Preparation, Writing – Review & Editing; **Fey D:** Writing – Original Draft Preparation, Writing – Review & Editing; **Goble C:** Writing – Original Draft Preparation, Writing – Review & Editing; **Golebiewski M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Gruden K:** Writing – Original Draft Preparation, Writing – Review & Editing; **Heil KF:** Funding Acquisition, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Hermjakob H:** Writing – Original Draft Preparation, Writing – Review & Editing; **Kahlem P:** Writing – Original Draft Preparation, Writing – Review & Editing; **Klapa MI:** Writing – Original Draft Preparation, Writing – Review & Editing; **Koehorst J:** Writing – Original Draft Preparation, Writing – Review & Editing; **Kolodkin A:** Writing – Original Draft Preparation, Writing – Review & Editing; **Kutmon M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Leskošek B:** Writing – Original Draft Preparation, Writing – Review & Editing; **Moretti S:** Writing – Original Draft Preparation, Writing – Review & Editing; **Müller W:** Writing – Original Draft Preparation, Writing – Review & Editing; **Pagni M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Rezen T:** Writing – Original Draft Preparation, Writing – Review & Editing; **Rocha M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Rozman D:** Writing – Original Draft Preparation, Writing – Review & Editing; **Šafránek D:** Writing – Original Draft Preparation, Writing – Review & Editing; **T. Scott W:** Writing – Original Draft Preparation; **Sheriff RSM:** Writing – Original Draft Preparation, Writing – Review & Editing; **Suarez Diez M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Van Steen K:** Writing – Original Draft Preparation, Writing – Review & Editing; **Westerhoff HV:** Writing – Original Draft Preparation, Writing – Review & Editing; **Wittig U:** Writing – Original Draft Preparation, Writing – Review & Editing; **Wolstencroft K:** Writing – Original Draft Preparation, Writing – Review & Editing; **Zupanec A:** Writing – Original Draft Preparation, Writing – Review & Editing; **Evelo CT:** Writing – Original Draft Preparation, Writing – Review & Editing; **Hancock JM:** Conceptualization, Funding Acquisition, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** The ELIXIR Greece Node was funded by the grant EPAnEk-NSRF 2014-2020 ELIXIR-GR project (MIS 5002780).

**Copyright:** © 2024 Martins dos Santos V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Martins dos Santos V, Anton M, Szomolay B *et al.* **Systems Biology in ELIXIR: modelling in the spotlight [version 2; peer review: 1 approved, 2 approved with reservations]** F1000Research 2024, 11(ELIXIR):1265 <https://doi.org/10.12688/f1000research.126734.2>

**First published:** 07 Nov 2022, 11(ELIXIR):1265 <https://doi.org/10.12688/f1000research.126734.1>

**REVISED Amendments from Version 1**

We have made some improvements in the revised paper responding to the reviewers' comments and updating it with reference to changes within ELIXIR, for example the emergence of a Single Cell Community within ELIXIR. In particular we have added more discussion of PBPK models and AI/Machine Learning. We have also added 2 new authors for this version.

**Any further responses from the reviewers can be found at the end of the article**

**Executive summary**

This white paper presents the future strategy of the new ELIXIR Systems Biology Community. This emerging ELIXIR Community was established upon the recommendation of ELIXIR's Systems Biology Focus Group to develop and coordinate ELIXIR's interactions with the broader systems biology community. **The infrastructure aspects of systems biology - databases, tools and standards development, as well as training and access to cloud infrastructure - are not only appropriate components of the ELIXIR infrastructure, but will prove effective drivers of ELIXIR's future support of advanced biological applications and personalised medicine.**

Systems biology is defined here as: modelling and understanding living systems in terms of their very large numbers of molecular interaction properties using a wide range of approaches, which can be further classified as bottom-up (starting from the molecular components) or top-down (starting from system behaviour). A key feature of systems biology is that, because of the complexity of the systems that are studied and the variety of data that is collected, it often requires collaboration between different laboratories and between computational biologists and experimentalists, often asynchronous in both space and time (e.g. through the literature and databases). It is essential that all the components that go towards building a systems model, from datasets and data collection methods to the models themselves, are FAIR (Findable, Accessible, Interoperable and Reusable) (Wilkinson *et al.*, 2016).

Historically, in Europe and worldwide, there have been significant investments in the development and applications of systems biology (see Table 1 & Table 2). In Europe, this culminated in the establishment of the Infrastructure Systems Biology Europe (ESFRI ISBE); however, this initiative fell short of obtaining sufficient support from member states. Meanwhile ELIXIR, the European Research Infrastructure for life science data, has become operational and established a range of **user Communities**. ELIXIR's new Systems Biology Community will build upon some of the strands of the work in ISBE, as well as work that has been taking place within ELIXIR's **Communities** and **Platforms**.

The Community has identified seven key challenges for systems biology in the short to medium term that can be addressed in part by ELIXIR, each with their own identified sub-challenges:

- 1. Barriers to the wider uptake of systems biology** (the challenge of providing well-parameterized systems biology models across the breadth of the life sciences; lack of data that can be used to accurately define the model parameter values; lack of standardisation and interoperability of systems biology; the resistance to mathematical modelling that is still present in biological and medical sciences)
- 2. Linking new and existing data to systems biology models** (FAIR generation of data well-suited to use in modelling and linking data and models in a FAIR way)
- 3. Interoperability of systems biology resources** (descriptions and annotations of data, models and their content need to follow coherent terminologies and ontologies; models, workflows, and data require FAIR, state-of-the-art computational infrastructure and tools for storage, access, and efficient use; interlinking standard models with models expressed in scientific programming and general purpose languages; interlinking descriptive and predictive models; trained experts to develop and curate resources that are user friendly and accessible; availability and accessibility of existing computational analysis methods and their interoperability with relevant modelling approaches)
- 4. Further development and embedding of systems medicine** (identification and modelling of network structures that are prognostic and predictive; integration with related systems such as microbes or exosomes; personalising models using patient data; clinically validating models; addressing ELSA (Ethical, Legal and Social Aspects) for clinical, sensitive data; interfacing with epidemiology)
- 5. Provisioning of modelling as a service** (provisioning of sophisticated data resources, tools and rich standards that are useful both for data mining and experimental design; availability of experts; availability of blueprint models to help with building new models)
- 6. Capacity building and training** (different trainee backgrounds; need for a systems biology learning path and broad promotion of the integrated systems biology framework; availability and use of standardised datasets in training materials; need for a broad training expertise)
- 7. Need to support industrial and societal embedding** (addressing, among others, pharmacology; toxicology; diagnostics; synthetic biology; agronomy; including microbial biotechnology and the bio-economy)

Many areas of relevance to systems biology are already embedded into activities of ELIXIR's **Platforms**, **Communities** and **Focus Groups** (see section 4 below) and have the potential for further developments. In particular we identify the following:

**Data Platform:** ELIXIR Core Data Resources and Deposition Databases are already highly used by the systems biology community. For example, data from BRENDA (Chang *et al.*, 2021),

**Table 1. Summary of European initiatives in Systems Biology.**

<b>H2020 projects</b>	<p>Projects EmpowerPutida, P4SB, Shikifactory100, DD-DECAF, SINFONIA, BIOS and SafeChassis (all industrial biotechnology), BioRoboost (standards SynBio) were / are all efforts connected to Metabolic Engineering / Industrial Biotech; PoLiMeR (Polymers in the Liver: Metabolism and Regulation, ITN); MESI-STRAT (Systems Medicine of Metabolic-Signaling Networks - A New Concept for Breast Cancer Patient Stratification); EPIPredict-(Systems Biology around the epigenetics of Estrogen Receptor-mediated breast cancer).</p> <p>REPO-TRIAL (Systems Medicine and drug repurposing)</p> <p>ADAPT (Accelerated Development of multiple-stress tolerAnt PoTato)</p> <p>EERA <b>CoBioTech</b> cofund action for <i>systems biology</i> and <i>synthetic biology for industrial biotechnology</i>; previous ERANets include <b>ERASysAPP</b> and SysMO.</p> <p>ERA-Net Cofund Scheme ERACoSysMed aims to enhance the implementation of Systems Biology approaches in medical concepts, research, and practice throughout by structuring, coordinating, and integrating national efforts and investments.</p> <p>CORBEL (<a href="https://www.corbel-project.eu/">https://www.corbel-project.eu/</a>) - Coordinated Research Infrastructures Building Enduring Life-science Services, where ISBE provided modelling services coordinated with other Research Infrastructures.</p> <p>Recon4IMD (Accelerating the diagnosis and personalising the management of inherited metabolic diseases).</p>
<b>ESFRI &amp; other (national) research infrastructure</b>	<p><b>ELIXIR Microbial Biotechnology Community</b>, on the specific ESFRI project <a href="http://www.IBISBA.eu">www.IBISBA.eu</a> (industrial biotechnology, many aspects related to workflow, data and models intertwined very tightly)</p> <p>ISBE (Infrastructure Systems Biology Europe, see below)</p> <p><b>UNLOCK</b> (An open infrastructure for exploring new horizons for research on microbial communities), a Dutch infrastructure on the national roadmap.</p> <p>A worldwide network of biofoundries of special relevance to synthetic biology and industrial biotechnology, <a href="https://biofoundries.org/">https://biofoundries.org/</a></p>
<b>IMI H2020 projects</b>	<p>Drug Disease Model Resources (<b>DDMoRe</b>)</p> <p>Enhancing Translational Safety Assessment through Integrative Knowledge Management (<b>eTRANSafe</b>)</p> <p>Translational quantitative systems toxicology to improve the understanding of the safety of medicines (<b>transQST</b>)</p>
<b>EASYM</b>	<p><b>EASyM</b> is a charitable association open to everyone with an interest in personalised medicine and Systems Medicine.</p>
<b>VPH</b>	<p>The virtual physiological human (<b>VPH</b>) is a long-standing and successful activity focusing on modelling physiology, such as that of the heart and of diabetes. It is much less molecular than mainstream Systems Biology (but of course not less relevant thereby).</p>
<b>EU-STANDS4PM</b>	<p><b>European standardization framework</b> for data integration and data-driven <i>in silico</i> models for personalised medicine.</p>
<b>Disease Maps</b>	<p>An open community effort to comprehensively <b>represent disease mechanisms</b> for various diseases.</p>
<b>ITFoM</b>	<p>For a long time now initiatives aiming to make comprehensive, molecules-up, models of entire organisms (<a href="http://www.siliconcell.net">www.siliconcell.net</a>), including the human, have been brewing. These even made it into a candidate flagship programme of the EC but did not receive funding. Yet the European Commission is still drafting a roadmap for such a program, called EC initiative 'Human Digital Twin', the European Commission is currently drafting a roadmap</p>
<b>BioModels</b>	<p><b>BioModels</b> is an EMBL-EBI based repository for curated models, with a focus on SBML-based models, but also providing models in other representations.</p>
<b>Metabolic Atlas</b>	<p><b>Metabolic Atlas</b> is a freely available repository and tool for visualisation and exploration of open-source genome-scale metabolic models (GEMs), particularly for human and model organisms, developed by the Nielsen Lab at Chalmers University of Technology. The web portal is developed open-source, it integrates GEMs benefiting from community-driven curation towards FAIR models, and it further presents a number of tissue-, cell line- and disease-specific GEMs for usage in systems medicine approaches.</p>

<b>MetaNetX/MNXref</b>	<a href="#">MetaNetX</a> is a unified namespace of metabolites and biochemical reactions developed to bring together models and resources published by other groups. It is distributed under open-source licence as a database (MNXref). In addition, a standalone software suite (MNXtools) will soon be released to critically assess and suggest improvement for Genome-Scale Metabolic Models, w.r.t to biochemistry.
------------------------	---

**Table 2. Summary of international initiatives and resources in Systems Biology.**

<b>COMBINE</b>	Computational Modelling in BIology NEtwork ( <a href="#">COMBINE</a> ) that coordinates the development of modelling standards: SBML, SBGN, SBOL, SBOL Visual, CellML, NeuroML SBGN, SED-ML and Biopax.
<b>COVID-19 Disease Maps</b>	<a href="#">COVID-19 Disease Map</a> , aims to establish a knowledge repository of molecular mechanisms of COVID-19 as a broad community-driven effort with contributions from the Disease Maps community, WikiPathways, and Reactome.
<b>LiSyM (Germany) LiSyM-Cancer</b>	Liver Systems Medicine Network ( <a href="#">LiSyM</a> ): Striving to develop non-invasive methods for diagnosing and treating NAFLD by combining mathematical modelling and biological research. Follow-up initiative LiSyM-Cancer started in July 2021.
<b>Center for reproducible biomedical modelling</b>	<a href="#">Reproducible Biomedical modelling</a> aims to enable larger and more accurate systems biology models, as well as their applications to science, bioengineering, and medicine, by enhancing their understandability, reusability, and reproducibility.
<b>Interagency Modelling and Analysis Group (IMAG) - Multiscale Modelling Consortium</b>	<a href="#">IMAG</a> is a government group of program officials from multiple federal government agencies supporting research funding for modelling and analysis of biomedical, biological, and behavioural systems. The IMAG wiki supports the activities of the Multiscale Modelling (MSM) Consortium and other IMAG agency-supported research consortia that focus on modelling and analysis projects. MSM consortium is focused on multiscale modelling of biomedical, biological, and behavioural systems. An example of an MSM subgroup is the recently formed <a href="#">Multiscale modelling and Viral Pandemics</a> to tackle ongoing and future viral pandemics.
<b>INCOME</b>	<a href="#">INtegrative COLlaborative modelling in systems MEdicine</a>
<b>BioSys ANR call in France</b>	Funding of several excellence initiatives in France, including Institutes of convergence and Laboratories of Excellence closely related to Systems Biology.
<b>ERAnet Sysbio call</b>	To promote multidimensional and complementary European Systems Biology projects, programmes, and research initiatives on a number of selected research topics on applied translational Systems Biology.
<b>JWS Online</b>	A transnational live model repository initiated in Stellenbosch (South Africa), and then extended to Amsterdam and Manchester. <a href="#">JWS Online</a> has been integrated into FAIRDOME and has been funded by various South African, Dutch, German, and UK grants.
<b>COPASI</b>	A software tailor-made for (stochastic) rate-balance-equation modelling and analysis of biochemical reaction and signalling networks. COPASI has been and is funded by various grants mostly from the UK, German, and US governments, and is an ELIXIR service, fully integrated with Systems Biology standardisation.
<b>FAIRDOM</b>	A community ( <a href="#">FAIRDOME</a> ), software platform ( <a href="#">FAIRDOME-SEEK</a> ) ( <a href="#">Wolstencroft et al., 2015</a> ), and public resource ( <a href="#">FAIRDOMHub</a> ) ( <a href="#">Wolstencroft et al., 2017</a> ) serving the asset management needs of Systems Biology projects are created and hosted by several ELIXIR nodes.  Of the 140+ instances of the FAIRDOME-SEEK many serve Sys Bio (e.g. Leipzig Health Atlas, LiSyM, IBISBAHub). The FAIRDOME-SEEK is the platform used by the Industrial Biotech ESFRi IBISBA (IBISBAHub).
<b>WikiPathways</b>	<a href="#">WikiPathways</a> is an open database of biological pathways maintained by and for the scientific community ( <a href="#">Martens et al., 2021</a> ). WikiPathways is managed by the Gladstone Institute in San Francisco and Maastricht University in the Netherlands. The project is supported by grants in the US and NL, and it is also an ELIXIR service.
<b>COLOMOTO</b>	<a href="#">COLOMOTO</a> is a consortium of research groups interested in logical modelling: modellers, curators and developers of methods and tools. The consortium works on the definition of standards for model representation and interchange (especially the SBML qual format), and on the comparison of methods, models and tools.
<b>Avicenna Alliance</b>	<a href="#">The Avicenna Alliance</a> advocates for the regulation and deployment of computer modeling and simulation (also known as <i>in silico</i> methods). The Avicenna Alliance's mission aims to complement traditional methods (bench, animal and clinical testing and trials) to deliver faster safer and more affordable health care to the patient.

STRING (Szkarczyk *et al.*, 2021), and Reactome (Gillespie *et al.*, 2022) are essential both for the construction of molecular pathways and the parameterization of molecular reactions. There is potential for the systems biology community to add new data resources to the ELIXIR list of services. Engagement of the Platform with curation efforts like BioModels (Malik-Sheriff *et al.*, 2020) can improve the FAIRness of systems biology-related data resources.

**Tools platform:** bio.tools (Ison *et al.*, 2016) and the workflow hub (Goble *et al.*, 2021) are invaluable resources for finding computational tools for systems biology. The ELIXIR Systems Biology Community aims to ensure relevant tools and workflows are added to these resources.

**Compute platform:** This Platform has strong links to the European Open Science Cloud which will be important for future instantiation and simulation of extensive, multiscale models in the cloud.

**Interoperability platform:** The Interoperability Platform's Recommended Interoperability Resources help to improve the FAIRness of systems biology-related data, tools and models, while the Platform's standards mapping resources such as BridgeDb (van Iersel *et al.*, 2010), identifiers.org (Wimalaratne *et al.*, 2018) and OLS (Jupp *et al.*, 2015) facilitate better interoperability and integration of data and models. The systems biology community at large has also developed its own standards, like SBGN and SBML, which we want to better connect with the abovementioned resources. The FAIRDOM platform (Wolstencroft *et al.*, 2017) (part of the ELIXIR CONVERGE data management toolkit) provides a collaborative community space for the FAIR integration of data and models in their experimental context.

**Training platform:** Better and more extensive training in systems biology tools and methods, e.g. including the development of new training materials, providing a repository and an annotation process for training materials, will be essential for the wide uptake of systems biology methodologies.

**Communities:** Many Communities already cover systems biology aspects and may well benefit from and provide data, tools and expertise for the modelling of various aspects in their systems. From a technological perspective, the *Galaxy Community* will play an important role in the integration of omics data and systems biology tools into workflows. Related to this, the *Metabolomics Community* has already begun to work on standardising fluxomics workflows. The *Microbial Biotechnology Community* has a strong interest in systems biology as it aims to contribute to addressing standardisation and other issues in relation to models and their applicability. Other Communities, such as *Plant Sciences*, *Microbiome*, *Food and Nutrition* and *Toxicology*, have potential to develop and deploy systems biology applications as part of their work towards the understanding of their systems under study. It is foreseen that the Human Data Communities, especially the *Federated Human Data* and *Rare Diseases* Communities, will provide data, tools and expertise for the modelling of human disease.

**Focus groups:** The *Machine Learning*, *EOSC* and *Registries* Focus Groups will be instrumental in the advancement of new

techniques for model development and the implementation of systems models in a cloud environment (making use of registries to make data, tools and workflows FAIR). FAIRness will also be improved by working with the *Biocuration* and *FAIR Training* Focus Groups. Large-scale systems biology modelling of interactions and evolution of populations and ecosystems is likely to increase, and it is expected that the *Biodiversity* Focus Group will mediate the interactions of the Community with such efforts.

The Community has developed a plan for future aims and objectives on short (3 years), mid (6 years) and long-term (10 years) timescales, revolving around four pillars: Standardisation & Interoperability, Technology, Training & Capacity building, and Embedding.

#### Short term (3 years):

- Better support existing standards in model repositories;
- Build upon systems biology models to improve the design of experiments that lead to the generation of higher quality, quantitative, FAIR data;
- Address specific challenges for human modelling, which include: working with compartments; model validation through standardised phenotypes; initial interfaces for multi-level modelling and integration across scales; multi-tissue evaluations; extrapolations from single cell analysis to tissue level; microbiome - host interactions; integrating sensitive personal data into models for personalised medicine;
- Establish approaches for model exchange, building on existing resource developments in the FAIR data landscape (e.g. FAIR data points; JWS-Online), BioModels and ModeleXchange;
- Understand how big data and AI meet models meaningfully;
- Intertwine temporal and spatial modelling appropriately;
- Improve the interoperability of modelling, simulation and analysis tools;
- Interface to synthetic biology through model-based design and model-based-learning strategies;
- Pre-screen trainees prior to training events to make recommendations for courses to be followed in the context of the event;
- Integrate new systems biology courses into TeSS and co-promote them with existing TeSS courses;
- Strengthen synergies with the other ELIXIR Communities, e.g. via joint training events;
- Together with the training platform set up a "gap analysis survey" to find out the strengths and needs for each ELIXIR Node;
- Implement turnkey solutions for different Nodes (specific training, capacity building, staff exchanges, knowledge exchanges and so on);

- Establish Key Performance Indicators (KPIs) to measure the impact of different actions.

#### Mid-term (6 years):

- Improve the standardisation of meta-data to describe time-series, functional and imaging data such as to facilitate their integration in computational models; Providing a link between existing models, datasets, and analysis tools for easy access to relevant data;
- Develop good strategies (including training) to improve reproducibility, credibility, and validation of models and to assess the efficiency of tools, leading to the development of quality marks, thereby increasing the quality of workflow outcomes;
- Develop the basis for theoretical and practical multi-scale modelling frameworks;
- Provide the basis for developing Digital Twins (microbes, bioreactors, organs, organisms, ecosystems);
- Extend the use of synthetic and standardised datasets in most systems biology training events;
- Support current and future trainers via Train the Trainer ELIXIR events;
- Continue and review the process of gap analysis survey and KPIs to include SME and industry;
- Involve small and medium-size enterprises in the capacity building process;
- Identify and advertise the success stories of systems biology to encourage its wider uptake
- Identify new areas, themes and challenges in systems biology.

#### Long-term (10 years):

- Improve the interoperability of data and models to enable FAIR model connection and integration (at different scales) so as to facilitate the development of multi-scale modelling frameworks;
- Develop Digital Twin methodologies that provide sufficiently accurate, real-time and dynamic depictions of physical biosystems (microbes, bioreactors, organs, organisms, ecosystems);
- Steer and modify processes, stratify patients and thereby support medical decision-making;
- Create a centralised repository of systems biology training materials aggregated by TeSS;
- Systematically review trends in systems biology and update training resources accordingly;
- Automate the capacity building process for new partners (communities, countries, etc.).

- Increase uptake of systems biology methodologies by the communities of biologists, bioengineers and physicians;
- Increase the uptake of standards (e.g. for model and data reporting) by the world wide systems biology communities.
- Develop the integration of machine learning/AI approaches with more traditional mechanistic modeling approaches
- Increase international cooperation in systems modeling not only within Europe but also with global players on other continents, such as US, South Africa and Japan.

### Introduction

Systems biology aims to understand how biological functions **emerge** from interactions between the multiple components of living systems by modelling the (dynamics of) interactions and processes. It studies what makes **the whole (the system) different from the sum of its parts**.

Since modelling and understanding of living organisms (“systems”) in terms of their thousands of molecular properties cannot be achieved in one step, systems biology comprises a variety of diverse, complementary approaches. **Bottom-up** systems biology starts from a limited number of components and studies the nonlinear mechanisms through which new properties emerge that are important for function. **Top-down** systems biology searches in multiple functional dimensions of the whole living system for correlations and patterns that clarify where functions have arisen from coherent behaviour of components. Top-down and bottom-up models should ultimately converge and predict all experimentally measured behaviour. **Multiscale** systems biology links systems biology at a smaller or faster scale (e.g. the cell) to systems biology at a greater size or longer time (the organism or the population). A key feature of systems biology is often the integration of data, often in large volumes that may be heterogeneous, at multiple scales, both in space and in time.

Due to their internal and external connectivity and nonlinearities, the study objects of systems biology are generally too complex to be handled by a single laboratory. They require a ‘worldwide laboratory without walls’, akin to the open science community. For instance, data and models that come from different expert groups dealing with a cell’s nucleus need to be integrated with data and models of that cell’s cytosol, as well as with data and models of neighbouring cells or tissues. The **data integration** inherent in systems biology requires metadata **standards** and ontologies, for computer-aided data exchange and integration, for computational software, for quality control, for experimental methods and for the reporting of experimental data. **Models** are required for experimental design and experiments are needed to validate them, whilst standards are required for their communication (Nickerson *et al.*, 2016; Stanford *et al.*, 2015; Stanford *et al.*, 2019; Waltemath *et al.*, 2016). **High Performance Computing** (HPC) hardware with dedicated **software** is required to run complex



models. In addition, a **capacity building** strategy is key in order to train dedicated experts in the field.

Model-driven **integration of heterogeneous and distributed data and knowledge** is instrumental to the reconstruction of emergent behaviour in systems biology. For this integration to be scientifically tractable or even possible, the data, as well as the processes of integration and the models themselves, must be **FAIR** to begin with. For the integration of data and knowledge in models, continual improvements in interoperability are needed to improve the link between the entities in the model and the FAIR data, e.g. from a gene expressed to an active enzyme with supporting literature evidence.

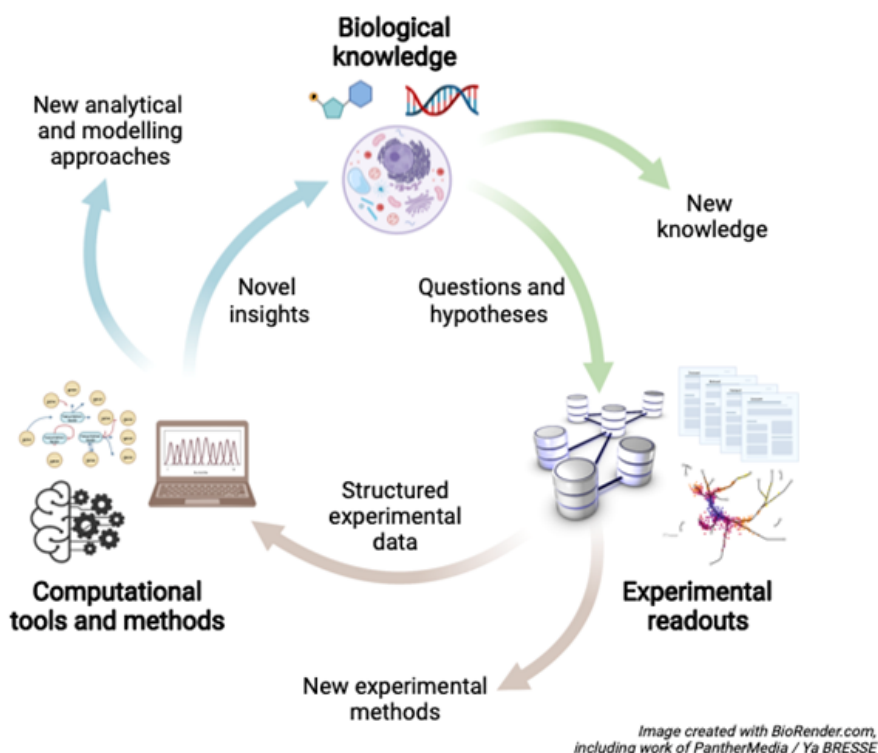
Many functions are critical to an organism's evolutionary fitness and competitiveness. Because all components of any living organism are, often indirectly, connected, systems biology requires the **experimental collection, and then integration**, of results obtained from high-throughput genome-wide, or other holistic, methodologies (see [Figure 1](#)). Indeed, the **accelerating increase of omics data generation** has been a major driver of modern systems biology. In order to understand the complexity of living systems in realistic terms, much

of systems biology engages in experimental analyses co-designed and analysed by mathematical modelling in a physical-chemical context.

The dependence on data from omics experiments, biochemistry and physiology mirrors the 'ecosystem' of data resources, tools, and Communities represented in ELIXIR. Data resources provide standardised and FAIR data of a particular kind. These data can be analysed through high-quality and well-characterised tools, often encapsulated in workflows. ELIXIR's Communities then specialise in the standardisation and analysis of particular types of data, in their application to particular types of biological problems, or in underpinning technologies.

#### Historical context: large-scale investment in systems biology in Europe and worldwide

In the early 2000s, the challenges and benefits of integrating experimental molecular biology with mathematics and informatics had already been demonstrated in systems biology. Reflecting its perceived impact on both life science research and economic development, systems biology research was supported by numerous European and global efforts. In Europe, an Infrastructure for Systems Biology in Europe (ISBE) was added to the [ESFRI roadmap](#) in 2010. In



**Figure 1. The systems biology "cycle".** To address a particular biological question, the cycle may start by data- and/or hypothesis-driven modelling of the system being addressed, followed by the generation of testable hypotheses, sets of predictions and design of the subsequent laboratory experiment. The experiment is carried out, data are generated and this is followed by computational analysis of these data, comparison of prediction to experiment and refinement of the model, which ultimately leads to gaining insights and generating new hypotheses. This cycle builds upon essential disciplines from Biology (the scientific question), Technology (experimentation and generation of data) and Computation (data analysis, modelling and prediction). Although the diagram is a cycle, it represents a spiral: with every turn data, knowledge and models increase, which could be seen as the cycle spiralling out of the plane of the paper/screen.

subsequent years, and in the scope of an ESFRI preparatory phase, a science case and business plan were formulated with the goal of ISBE becoming a legal entity upon gathering support from the various member states involved. The efforts to establish ISBE as a stand-alone entity received the support of too few member states however, and ISBE was removed from the [ESFRI roadmap](#) in December 2021.

Whereas the focus of **ISBE** could be summarized as “**Models for Life**”, **ELIXIR**, which is the European Research Infrastructure for life science data, focuses on **Data for Life**. However there is a continuum between the two: models rely on experimental data, and are often very tightly integrated with their supporting data. The same ambivalence applies to specific resources. An example is the **FAIRDOM** initiative and the **FAIRDOMHub**, which was developed within ISBE as a follow-up from initial efforts within the ERA-Net Systems Biology, among others. Other examples are the **BioModels repository**, an ELIXIR Deposition Database, which stores models for Life, **JWS Online**, and perhaps the most visible part of ISBE, its Make me My Model component, mostly at ISBE.NL (see the ISBE deliverables<sup>1</sup>).

### Development of the ELIXIR Systems Biology Community

In response to the issues faced by ISBE and its expected loss from the research infrastructure landscape in Europe, and the awareness that many aspects of systems biology were already features of ELIXIR resources, a Focus Group on Systems Biology was established in ELIXIR in April 2020, which held its first meeting in June of that year. The aim of this Focus Group was to recommend how ELIXIR should respond to the situation around ISBE and support systems biology. The Focus Group presented its report to the ELIXIR Heads of Nodes in May and September 2021 with the recommendation that a new Community in Systems Biology should be established within ELIXIR. This suggestion was supported by the ELIXIR Heads of Nodes and this white paper represents the agreed priorities and aims of this new grouping.

### Developing an ELIXIR Systems Biology Roadmap Key barriers for wider adoption of systems biology

The overarching long-term goal of the ELIXIR Systems Biology Community is to make systems biology modelling a central pillar of research in biology. In this vision, systems biological models are developed based on the understanding of the biological problem, are used to design biological experiments, and help with the interpretation of collected data. The combined results then allow the development of actionable solutions to the original biological problem. Importantly, therefore, future developments need to improve the FAIRness of systems models, making them more accessible to and usable

by individual laboratories as well as large consortia. By combining improvements in standardisation with wider training efforts the Community will contribute proactively to this goal.

While systems biology is present in all biological disciplines, it has so far not reached its full potential. To achieve this, systems biology needs to become more accessible to enable a broader community to benefit from its approaches. The ELIXIR Systems Biology Community has identified the main barriers that prevent wider adoption of systems biology tools and it aims to gradually eliminate them through its future activities.

One barrier is the lack of availability of well-parameterised systems biology models that are of immediate value for researchers in the wider community. Biologists study a great variety of organisms, each with a great variety of pathways, and physicians study a great variety of diseases in a great variety of tissues. While biological pathway/network resources and genome-scale metabolic maps have grown over the last decade and their usability for flux balance analyses has greatly improved, other approaches additionally require the setting of the much larger number of parameters inherent in rate equations. The values of these parameters differ between species, individuals and tissues. Consequently the probability that any of the available kinetic, stochastic or multiscale models fits the experimental object a particular researcher is interested in is extremely small. Kinetic, stochastic, and multiscale models continue to be used only by modelling experts. There are many additional facets to this problem. The first is that systems biology models are rather difficult to build - both the understanding of the underlying biology and of skillful modelling are rarely present in the same person. While we can partly solve these issues via better training, the former is also due to a limitation that can only be diminished in incremental steps by increasing our understanding of biology. The second facet is the lack of data that can be used to accurately define a particular use case. As there are (almost) no two particular cases that have the same parameter values, this results in model predictions with wide confidence intervals and low predictive value. A resource such as **BioModels Parameters** ([Glont \*et al.\*, 2020](#)) provides quick access to parameter values and ranges but is limited to models existing in the BioModels repository.

Another barrier lies in the insufficient standardisation and interoperability of systems biology, which often makes sharing and reuse a time-consuming challenge. The last decade has seen the creation and increased uptake of new systems biology standards, and increased sharing of standardised models with cross-referenced annotations in repositories, such as MetaNetX ([Moretti \*et al.\*, 2021](#)) and BioModels. However, these do not yet sufficiently cover the areas of model parameter sharing and linking with experimental data. There has also been a huge increase in the development of systems biology software and tools that make use of the standards, but switching between tools remains cumbersome. Better interoperability would enable the building of efficient systems biology modelling pipelines. Here, the ELIXIR Systems Biology Community can learn from the recent advances in the standardisation of sharing of biological data and

<sup>1</sup> ISBE deliverables: [ISBE D2.2 Preliminary Recommendations for a Data and Model Management Framework: Data and Model Stewardship for the ISBE Infrastructure \(2014\)](#); [ISBE D2.3 Final Recommendations for a Data and Model Management Framework: Data and Model Stewardship for the ISBE Infrastructure \(2015\)](#); [ISBE D2.4 Implementing Recommendations for ISBE Asset Stewardship Data and Model Stewardship for the ISBE Infrastructure \(2015\)](#).

data analysis pipeline development, where the wider ELIXIR community has played an important role.

Finally, we need to overcome the conceptual barrier to mathematical modelling that is still present in biomedical sciences. While biology is a mature science, in the sense that we have a decent understanding of the major processes that make it work, we mostly do not understand the details to make it truly applicable. To get at those details, and to take advantage of all the big and small datasets collected in the last decades, we need to build more and better systems biology models for better data interpretation, such that systems biology consistently demonstrates utility and becomes a part of all new biomedical studies. This needs to be done in ways that make the resulting models maximally accessible and usable by biomedical scientists. Until then, the next best thing we can do is to collect systems biology success stories and share them across the life sciences.

### Linking new and existing data to mathematical models

While discussions of systems biology are often focused on the modelling part, there is very little that can be achieved without making use of high-quality data. Data is a source of knowledge necessary to develop the systems biology models, to parametrize them and to evaluate how well they describe the biological question at hand. However, beyond the data produced in their own group or by close collaborators, a modeller rarely finds data perfectly suited for the task at hand. Mostly, this is because the data needed simply do not exist, which is not surprising given the small section of biology we have analyzed experimentally so far. Often the data that would inform modelling are there, but extremely difficult to find or are not annotated well enough for subsequent integration with models. Data for parametrization of dynamic systems biology models are particularly problematic, as these need to be time courses that cover the entire dynamics of the modelled process: biological networks adapt over time.

There are therefore two types of challenges that need to be tackled when discussing data in systems biology. The first is a cultural challenge on the data generating side. If data are to be useful for systems biology modelling, then this needs to be taken into account already when the experiments are designed, or better yet, when the grant applications are written. The more systems biology modelling is included in the early stages of the project, the more we can count on a one-on-one connection between model and data.

The second challenge, and one that the ELIXIR is very well positioned to help meet, is linking the data to the models in a FAIR way. Ideally, in the future, this would mean that a model developer would be able to search and find data they need and an experimental scientist would be able to easily find that an existing model can be used to analyse the data obtained. In the last decade, a giant leap forward has been achieved by collecting increasingly large amounts of experimental data in FAIR repositories; however, the data has rarely been linked to modelling repositories. One potential way forward should

be to make the data and modelling repositories mutually searchable. Another is to create data repositories that are aimed specifically at storing data that are useful for systems biology modelling. The first attempt in this direction is the recently established *datanator* repository (Roth *et al.*, 2021). A second is FAIR joint model/data repositories, such as FAIRDOM-Hub (Wolstencroft *et al.*, 2017) (with strong links to both ISBE and ELIXIR), which allows interlinking of experimental data, computational models and simulation results in a project-centred approach.

### Interoperability of systems biology resources

The essential components of systems biology integrate well with the ELIXIR infrastructure. Although systems biology models have not been, until now, at the heart of the ELIXIR infrastructure, they share many essential properties with the components of ELIXIR. ELIXIR already makes workflows, in essence programs and tools, part of their portfolio. Models, workflows, and data require state-of-the-art computational infrastructure and tools for storage, access, and efficient use. All of this needs to be FAIR. They require highly educated and trained experts to develop and curate resources that are user-friendly and accessible. Additionally, systems biology models depend greatly on the quality and quantity of data for their construction and simulation.

### Data and metadata

Metadata describing data and models with their items and entities need “minimal information” checklists for attributes to be listed. Descriptions and annotations of data, models and their content need to follow coherent terminologies and ontologies. This is a prerequisite for their integration into systems biology models. Models need standardised formatting and description (for comparison, for modularization, for integration/interlinkage into complex multiscale models, etc.). Standardised visualisation of models helps to share visual information (e.g. pathway diagrams, activity flows, entity relations) on the models in a consistent way.

There is no single tool that can encompass all aspects of systems biology. Being an interdisciplinary endeavour, systems biology projects span multiple specialities, multiple scales, multiple experimental methods, multiple modelling systems. This puts interoperability and flexibility at the centre of building tools and resources. As a consequence, systems biology standards and resources must aim for improved interoperability.

The international COMBINE initiative, with its standards SBML, SBGN, SBOL, SBOL Visual, CellML, NeuroML, SED-ML and Biopax (Hucka *et al.*, 2015; Waltemath *et al.*, 2020), has developed a range of interoperability standards for systems biology models, their visualisation, combination, and execution. COMBINE standards form the basis of model resources like the ELIXIR deposition database BioModels, as well as JWS online (Peters *et al.*, 2017). The ELIXIR Core Data Resource BRENDA, and the ELIXIR Node service SABIO-RK (Wittig *et al.*, 2018) both offer export of mainly literature-based enzymology/reaction kinetics data in SBML format. ELIXIR

has strong links to COMBINE, enabling and embodying information exchange between the organisations.

**Challenged interoperability where there are different descriptions of the same or related data.** When analysing data one often stumbles upon the problem that different data models still yield different descriptions, even though both descriptions are FAIR; there is more than one way for a model to be FAIR. FAIR identifiers from different databases, such as those provided by BridgeDb and MetaNetX), ontology terms from different domain ontologies (e.g. Oxo versus meta-ontologies like EFO) and connections between different levels of precision in chemical (sub)structures and chemical names (e.g. the ChEBI ontology, the Chemistry Development Kit (Willighagen *et al.*, 2017), and the various chemistry resources provided by the ELIXIR-CZ Node) should become interoperable with one another. These needs are a challenge not just for systems biology but in fact for ELIXIR, and especially the Interoperability Platform, as a whole, we will need to add the resources needed for interoperability as part of modelling and analysis to the initial work on FAIR descriptions.

The FAIRDOME-SEEK (Wolstencroft *et al.*, 2015) project data-management system for systems biology emphasizes integration of data and models. It supports researchers and collaborative projects to catalogue, organize, share, interlink and publish local and remote data files, models, protocols, workflows, etc., enabling in particular linking models and their supporting data. FAIRDOME-SEEK's 'Search' includes an external search in the BioModels repository. It integrates the BiVeS tool for describing differences between model versions (Scharm *et al.*, 2016). Integration of the tools JWS online and COPASI (Mendes *et al.*, 2009) enables users to run simulations of SBML models directly in the system. Integration with e.g. NeLS (Tekle *et al.*, 2018) and the Swiss openBIS (Barillari *et al.*, 2016) data management systems aim at bringing more ELIXIR data close to the models.

**Using descriptive models and linking them to predictive models.** An ongoing challenge of modelling is the linking of standard models with models expressed in scientific programming and general-purpose languages such as Matlab, Python or Julia. Researchers argue that innovative types of modelling precede standardisation, e.g. in the case of languages for exchange of multicellular agent models. Building bridges between different types of model specifications appears a worthy challenge for a network as wide as ELIXIR.

The integrative systems biology community has developed various resources for descriptive models (primarily molecular pathways), e.g. in Reactome, KEGG (Kanehisa *et al.*, 2008), MetaCyc (Caspi *et al.*, 2020), and WikiPathways (Martens *et al.*, 2021). Some early attempts to harmonise the content of these built on the BioPAX standard (Demir *et al.*, 2010) or simply on gene lists (e.g. Pathway Commons (Rodchenkov *et al.*, 2020) and the Molecular Signatures Database, MSigDb, at the Broad Institute (Subramanian *et al.*, 2005)). More recently, dedicated converters allow translation between

Reactome, WikiPathways, and the Disease Map resource MINERVA (Gawron *et al.*, 2016). This supports integrated analysis using the various resources including the conversion of these pathways into biological networks and exploration with network biology tools like Cytoscape (Shannon *et al.*, 2003). Such networks, combined with experimental data, uncover relevant aspects of molecular biology, like strongly connected parts or strongly regulated parts in the biological system. They also enable linking with gene regulation databases (e.g. transcription factor and microRNA target linking databases) and with databases of chemical interactions with molecular biological targets (e.g. drug-target or molecular toxicology databases). Moreover, such networks can be mapped to biological ontologies (e.g. from Gene Ontology (The Gene Ontology Consortium, 2019)), disease-related genes (e.g. from OMIM (Amberger *et al.*, 2019)), and variants (known from e.g. dbSNP (Sherry *et al.*, 2001) or observed experimentally). Of course, the possibility of linking multiple resources creates new standardisation and interoperability challenges (cf. 5.1).

A new development is the link from descriptive models to predictive models in SBML. Since pathway models already support standard descriptions of reactions (e.g. in SBGN and MIM) it is possible to convert them into SBML and that again allows their use as predictive models. This development was recently catalysed by the COVID-19 Disease Map project (Ostaszewski *et al.*, 2021) that is also supported by the FAIRDOME initiative with its FAIRDOMEHub platform to share models and corresponding data. **The basic infrastructure now exists and needs to be further developed, tested, and disseminated (e.g. as training modules).** In essence, this aspect of model interoperability can form the bridge between biological data analysts and predictive modellers working on the same biological systems. This idea was also the driver for the recent fluxomics Implementation Study of the Metabolomics Community. Fluxes are what is predicted by Flux Balance modellers while concentrations of gene products and metabolites are what is usually determined experimentally; the latter are at best a proxy of the former. Measuring fluxes or extending to dynamic modelling, provides the extra linkage.

The MetaNetX reconciliation (Moretti *et al.*, 2021) of metabolites and reactions aims at providing cross-references between major public resources for metabolism (e.g. KEGG, Rhea, ChEBI) and genome-scale metabolic models published by different groups (e.g. BiGG, Metabolic Atlas); only such a reconciliation can lead to true genome-scale metabolic models. The computation of the reconciliation considers three lines of evidence: the detailed metabolite chemistry, the description of gene-protein-reaction complexes inclusive of their kinetics, and the dynamic properties of the models, i.e. the distribution of permissible fluxes. Discrepancies, imprecisions, and mistakes in the metabolite chemistry are detected, possibly corrected and converted into cross-references that preserve at best the dynamic properties of the models. The seamless integration of existing models, and their improvement, require more software development and dedicated (application-specific)

databases and are in line with the requirements promulgated by the FAIR data consortium.

### Modelling as a service

Models generated via a systems biology approach should be concrete enough for their newly proposed molecular network mechanisms to be validated/invalidated experimentally. This may in part be automated (King *et al.*, 2009) at the high-throughput level, enabling advanced systems biology research to formulate the crucial questions for understanding the system under study. We expect this to enable a new type of bioengineering, which will develop new production processes as well as more effective medical therapies. The role of ELIXIR in supporting this encompasses the provision of sophisticated data resources and tools and rich standards that are useful both for data mining and experimental design.

Although we may anticipate the entire automation of model building in the foreseeable future, for the time being we still need a hand-crafted approach, where every model is unique and personalised for a certain customer. ISBE developed the “Make Me My Model” (M4) service (Kolodkin *et al.*, 2018) which was provided to various customers. For example, within the framework of the CORBEL project, ISBE has built a physiologically-based pharmacokinetic (PBPK) model for a customer from the Environmental Engineering Laboratory in Spain (Sharma *et al.*, 2020). The experience at M4 service highlighted that every new model requires a new configuration of the modellers team with experts from various modelling approaches. Taking advantage of a large research infrastructure where a flexible team of experts from different modelling areas could be quickly assembled for the needs of a specific project would be very advantageous in this sense.

Another advantage of using a large infrastructure and interoperable platforms is to minimise the need for new models to be built *de novo*. Already existing models for similar systems could be used as a starting point. We should also notice that, due to the similarities in the biological organisation, and as the same building blocks (biomolecules) and similar biochemical processes are used in all organisms, a common blueprint model could be used. Every new organism and every new case will be instantiations of this more generic model. Along with the development of ‘Silicon cells’ or ‘Digital Twins’ of Biological Systems (discussed in chapter 5.2), Models as a Service can become customer-specific tailoring of the already available blueprint model.

These various efforts are, altogether, necessary to make systems biology a true enabler of advancements in a variety of fields, from understanding host-microbiome interactions, to the development of comprehensive ecosystem models, biobased production processes, or systems medicine. In the following section we exemplify how systems biology plays a crucial role in enabling systems medicine.

### Systems biology underpinning systems medicine

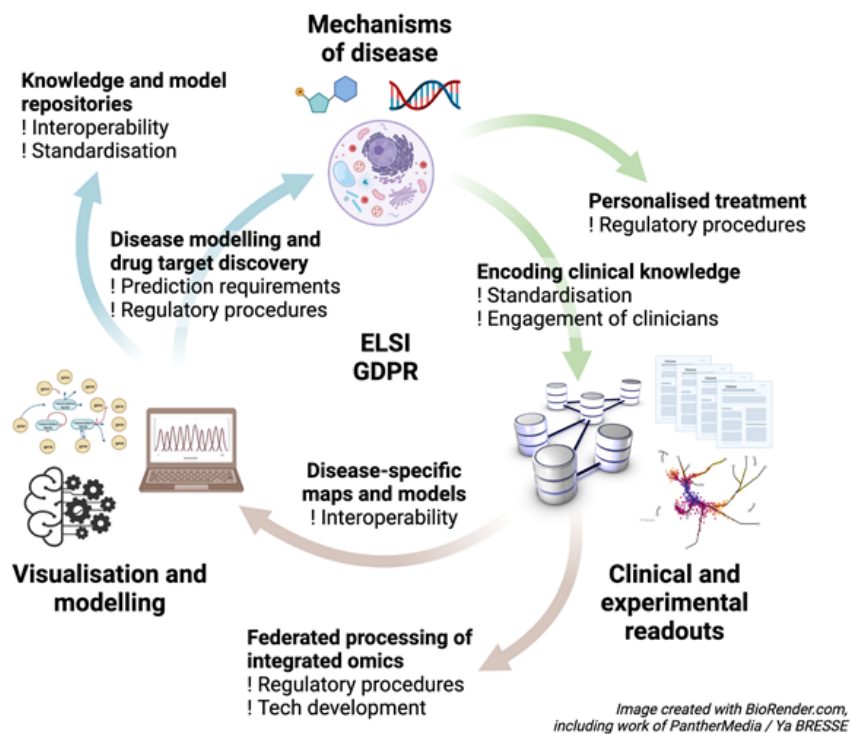
Systems medicine is an implementation of systems biology in the areas of clinical research and practice. It employs

computational, statistical and mathematical multiscale analysis and modelling methods to study disease mechanisms towards improved diagnosis, prevention and treatment. The framework of systems medicine is closely related to the concepts of personalised or precision medicine, where the systems approach informs decisions about tailored actions to improve the health of individual patients or patient subgroups. Numerous publications in high-profile journals have confirmed the benefits of systems medicine approaches in promoting precision in diagnostics and personalised therapies in cancer and other diseases, both rare and common. There are multiple examples of systems medicine showing its success and applicability, ranging from providing predictive models of multifactorial diseases, such as cancer or metabolism associated disorders (e.g. Ivanovic & El-Kebir, 2023; Kezer *et al.*, 2021), to pharmacokinetics and pharmacodynamics modelling approaches that can describe how a drug travels through the body and is metabolized, which is of particular interest for the pharmaceutical industry.

Systems biology studies of human disease have to encompass additional levels of complexity compared to other implementations in systems biology (Apostolopoulos *et al.*, 2020; Wolkenhauer, 2020; Zanin *et al.*, 2021), as highlighted in Figure 2. On one hand, human diseases are phenotypically and mechanistically better understood than diseases of other organisms, but on the other hand, there are difficulties studying these molecular mechanisms in accordance with ELSI and GDPR rules. Additionally, whereas we tend to ignore individual differences when we study model organisms, we do recognise the importance of individual differences in humans, obliging systems medicine tools and approaches to personalise disease interventions. Further, disease therapies, which are themselves complex, are also important areas of study and modelling. Implementations in the broad domain of systems medicine can range from semantic representations of diseases and disease maps through mathematical modelling of diseases to applications aimed at supporting P4 (predictive, preventative, personalised, participatory) medicine and at managing individuals’ health, including linkage to clinical monitoring devices and the use of big data and AI. An important requirement for clinical applications in the future will be the link to personal health data records.

Challenges associated with building personalised systems medicine models are: identification and modelling of network structures that are prognostic and predictive; integration with related systems such as microbes or exposomes; personalising models using patient data; and clinically validating these models.

Solving these challenges will require access to well-annotated patient data, developing standards for personalising and validating models, and a dedicated repository to exchange these models. Importantly, building models fit for clinical use in precision and personalised medicine should start with a clinical problem, and requires close collaboration between modellers and clinicians. An important aspect to addressing these challenges will be linking maps and models to sensitive data held in the Federated European Genome-phenome



**Figure 2. Illustration of the key areas of systems medicine (building on Figure 1).** Encoding clinical knowledge; federated processing of integrated omics data; construction of disease-specific maps; construction of knowledge and model repositories in the area; and disease-specific modelling and drug target discovery leading to the development of personalised treatments based on models of disease mechanisms. Underpinning all of these is the need to work within ELSI and GDPR regulations.

Archive (FEGA), potentially making use of Beacons to identify relevant information (e.g. variation associated with relevant phenotypes). This will become increasingly important with the availability of data from the 1+MG project<sup>2</sup>. Beyond this, it will be important to improve the interoperability of models with other data sources and tools. ELIXIR can help to address the challenges of linking models to clinical information through the work of its Health Data Focus Group. A general technical issue, shared by systems medicine and other applications that will make use of sensitive data in the human health and genomics domain, will be the development of federated learning algorithms that are able to learn models based on data held on a number of restricted-access servers. Furthermore, standards and solutions will need to be developed to link dispersed health information to integrating models such as disease maps and digital twins. At the same time, it will be important to improve interoperability between maps, models and digital twins and ELIXIR's Core Data Resources and Deposition Databases.

At the core of systems medicine lies the involvement of the user communities: the clinicians and the patients. The patient-oriented approach with the integration of personalised data, both clinical and omics, into network-based analyses and

models will enable tailored stratification, therapies, disease management strategies and monitoring. A key element that systems medicine can bring to ELIXIR is its engagement with clinical researchers collecting disease-relevant data and the application of its approaches close to the clinic.

Visual exploration and analytics of computable disease models, which bridges the expertise of clinical experts and the methods of bioinformaticians, enables knowledge about molecular disease mechanisms and relevant clinical data to be brought together for meaningful interpretation, thereby reducing the complexity of the knowledge and the scale of data (Satagopam *et al.*, 2016). This is greatly aided by disease maps (Mazein *et al.*, 2018), an emerging methodology for building human and machine-readable models of molecular disease mechanisms. They offer online and interactive exploration of diagrams describing molecular and cellular hallmarks of different disorders, with detailed annotations of participating molecules, and citations of articles describing the encoded mechanisms.

The construction of these maps is a challenge, as it requires close interactions with clinical experts, continuous quality checking against emerging facts and data, and persistent evaluation to support downstream modelling approaches. ELIXIR is in a prime position to support building such visual and computable repositories via its newly established

<sup>2</sup> <https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes>

**Disease Maps Node Service** and by engaging relevant communities, e.g. Rare Diseases, Federated Human Data, 3D-Bioinfo, Metabolomics, Proteomics, Galaxy, and others, many of which themselves engage in visualisation and modelling of different kinds of biological entities.

An example of such an ecosystem supported by ELIXIR is the COVID-19 Disease Map, engaging clinicians, life scientists and computational biologists to set up a graphical and computable repository of SARS-CoV-2 mechanisms. The engagement of a highly motivated community has resulted in an interoperable repository of curated diagrams following systems biology standards (Schreiber *et al.*, 2020), integrated with interaction databases and text mining platforms. We can consider this effort as a blueprint for building qualitative systems medicine models, which will feed downstream, detailed modelling workflows (Fröhlich *et al.*, 2018). A future aim for systems medicine within the Systems Biology Community will be to facilitate the construction of disease maps for currently unrepresented disease areas.

Incorporating individual differences into systems biology models can be used to create personalised models akin to a “digital twin” of a patient with a disease (Fey *et al.*, 2015). Modern omics technologies, such as genome sequencing, allow molecular profiling of individual patients, with unprecedented resolution down to single cells. A prominent example is *The Cancer Genome Atlas*, an enormous repository of multi-omics data from over 11,000 cases across 33 cancer types (Hutter & Zenklusen, 2018). A Single-Cell Omics Community has recently been founded within ELIXIR (Czarnewski *et al.*, 2022) which we expect to provide us with invaluable collaborations in future. Populating systems biological models with personal data can yield highly individualised models that can help simulate disease evolution and response to therapy with high sensitivity and specificity (Barrette *et al.*, 2018; Béal *et al.*, 2021; Bhinder & Elemento, 2017; Crawford *et al.*, 2018; Eduati *et al.*, 2020; Fey *et al.*, 2015; Hastings *et al.*, 2020). These systems medicine models are knowledge-based and thus able to offer insight into the disease and drug response mechanisms of a patient (Ebata *et al.*, 2022; Hutter & Zenklusen, 2018). For example, a personalised model of the JNK stress-response network resulted in refined patient-stratification for neuroblastoma, a common childhood cancer, and revealed an impairment of the JNK apoptotic switch in high-risk cases (Fey *et al.*, 2015). As more and more clinically-validated models of this kind arise we foresee a need for a repository for them which might, for example, be built in conjunction with Reactome and the proposed ModeleXchange initiative (see Section 5.1).

Personalised medicine, and therefore systems medicine, is an active area of industrial development in Europe (European Biopharmaceutical Enterprises, 2015). Facilitating linkage of systems medicine into the ELIXIR ecosystem of data, tools and standards can therefore be expected to have an impact on the development of this sector in the future.

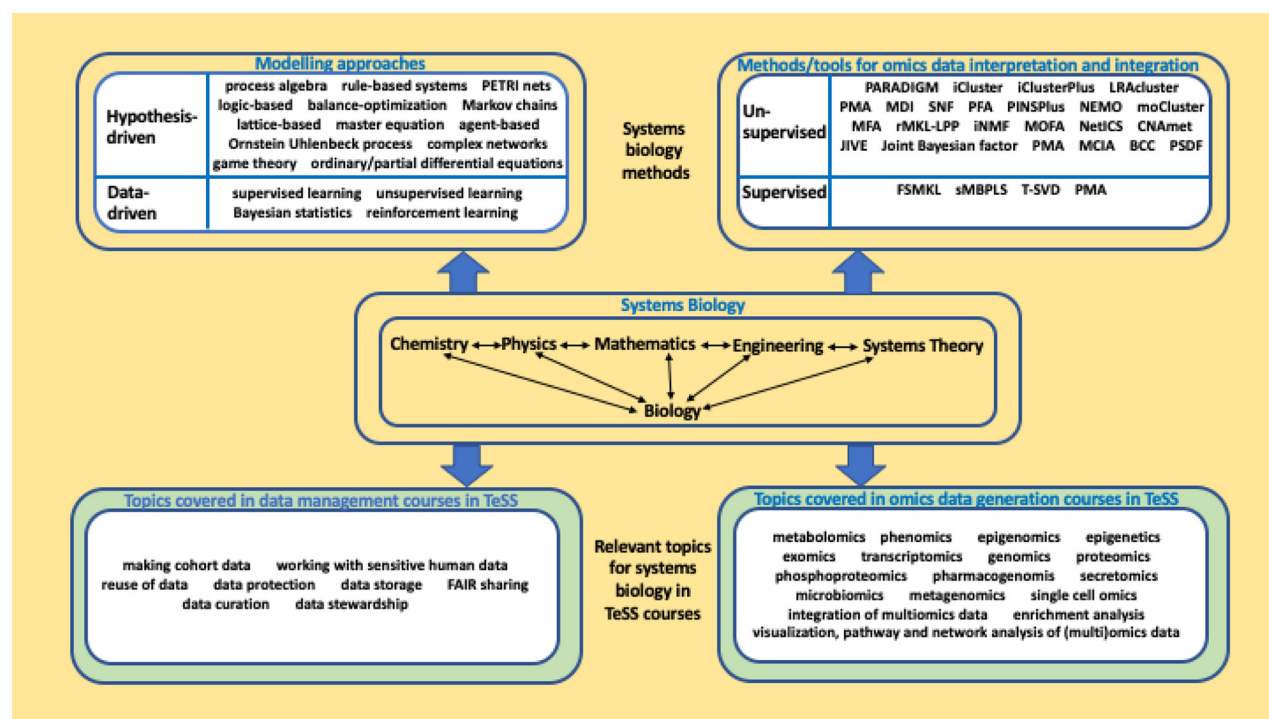
A need that systems medicine shares with other applications in the Human Data domain is to address ethical challenges of (i) accessing data for constructing models, (ii) the use of models and (iii) the use of their outputs. For example, constructing models using a combination of datasets describing an individual might increase their identifiability and aspects such as data ownership and the ability to withdraw consent from datasets could affect the persistence of models. The Community will therefore engage with ELSI efforts in the Human Data Communities, within ELIXIR and beyond, to investigate, for example, the impact of ELSI issues on building, constructing and sharing models based on human data. Close collaboration is also envisioned with the **Genome of Europe initiative**, which will start by collecting ELSI compliant human datasets from the general European population. Another ELSI aspect is engagement with industry, to ensure ethically acceptable uptake of the techniques and infrastructures it develops from lab to the bedside. Finally, the Community will work on the dissemination and outreach of developments resulting from the Community’s work to clinicians and patients.

### Capacity building and training

The stark rise of high-throughput technologies and related datasets describing the complexity of biological systems in health and disease, has been a driver of the emergence of systems biology, which comes with inherent training challenges. The ELIXIR Systems Biology Community has identified several challenges that need to be addressed for developing and delivering successful systems biology training that could be achieved within different time frames. These are detailed below.

**Diversity in trainee backgrounds.** Owing to the diversity of systems biology approaches, heterogeneity of trainee backgrounds is common. Trainees often differ greatly in their knowledge of biology and mathematical modelling, and their levels of experience with software tools or systems biology methods also varies from ‘novice’, to ‘competent practitioner’ and ‘expert’ user. Finding the optimal balance between providing sufficient information on systems biology databases and software and not overloading trainees with too many new concepts, represents one of the main training challenges (Figure 3).

To accommodate different trainee backgrounds, an option might be to pre-screen trainees well-ahead of the training event to understand their expectations of what they will achieve at the event. This could be in the form of a questionnaire, hosted by TeSS (Beard *et al.*, 2020) or sent out on a mailing list, with the objective to (1) assess the trainee’s abilities, (2) find out their training needs, and (3) make recommendations on existing courses based on (1) and (2). This may prove a good opportunity to synergize training with other biomedical communities (e.g. fluxomics, microbiomics) and offer general bioinformatics courses (e.g. on reproducibility and data management, omics data analysis). Such information will not



**Figure 3. Training courses available in TeSS can be fully integrated under the systems biology umbrella.** Methods/tools for omics data interpretation and integration are categorised based on unsupervised/supervised approaches (Subramanian *et al.*, 2020).

only help the trainers to adjust the course material, but also establish uniform criteria for pre-requisites of the course. Building a systems biology concept map covering the basic areas of systems biology research and perspectives should be of value. The organisation of hackathons to decide on the concepts that are to be included in more specialised training courses depending on the level of maturity of the trainees in systems biology, has proven to be a successful practice for setting up an ELIXIR training school in a rather diverse area of expertise.

**Creation of a systems biology learning path and broad promotion of the integrated systems biology framework.** Although some independent ELIXIR training activities already deliver systems biology-related training modules (e.g. single-cell omics, metabolic modelling, data integration), many may well be missing or unknown. There is a need for integrating new systems biology courses into TeSS and for integrating them with existing courses under the systems biology umbrella.

In line with the efforts of the ELIXIR Training Platform in Task 2, the ELIXIR Systems Biology Community will define a learning path, identify gaps on missing topics and suggest suitable courses. For example, there may be a need for a course on reproducibility of systems biology models, on handling sensitive human data, or on specialised omics topics. As depicted by Figure 3, the systems biology framework can be well-complemented by omics data generation courses and data management courses, already covered in TeSS.

Coordinating educational events across borders is an important part of training. The community will engage in facilitating the collaboration of ELIXIR Nodes for training purposes. Additionally, the ELIXIR Systems Biology Community will initiate joint events with other ELIXIR communities via workshops and hackathons. A complementary approach may be to collaborate with non-profit organisations in the advancement of systems biology-related areas and education, such as iGEM.

We will systematically review the interest in novel systems biology topics and will propose new courses, based on demand. Dissemination of existing (see Table 3) and new systems biology courses will be enabled by the TeSS platform, with some of the systems biology-related ELIXIR training courses already reported in TeSS. New systems biology courses will be integrated into TeSS and further promoted by organising ELIXIR-level international courses. To reduce the technical issues associated with broken web links on TeSS, we will perform regular checks of these and will also encourage our trainees to report any issues to the site's administrator. Moreover, we will investigate the overlap between TeSS and other training resources such as the Galaxy Training Network. E-learning approaches for courses and training educational resources will be incorporated in collaboration with the Training Platform and its services.

**Availability and use of standardised datasets in training materials.** With the ever-changing systems biology landscape, new tools and massive amounts of data become available



**Table 3. Examples of systems biology courses across the ELIXIR Nodes.**

Course	ELIXIR Nodes involved	Length of the course	Training material publicly available
Systems biology: From large datasets to biological insight	EMBL-EBI	5 days	yes
Mathematics of life: Modelling molecular mechanisms	EMBL-EBI	5 days	no
ELIXIR Omics Integration and Systems Biology	ELIXIR-SE	5 days	yes
Course portfolio Bioinformatics and Systems Biology	ELIXIR-NL	5 days	no
Tools for Systems Biology modelling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK	ELIXIR-DE	3 days	no
Computational Systems Biology for Complex Human Disease: from static to dynamic representations of disease mechanisms	ELIXIR-FR, ELIXIR-UK	5 days	yes
ELIXIR Fluxomics Training School	ELIXIR-GR, ELIXIR-ES	5 days	partially
Hands-on tutorial Systems Biology/Medicine	ELIXIR-SI	3 days	no

and so, trainers need to continuously adapt to these developments. One of the training challenges relates to the lack of standardised and FAIR datasets underpinning specific topics and the lack of suitable documentation for some of the tools or databases. This makes it difficult for trainers to keep their materials up-to-date. In addition, training materials are often focused on specific software or databases, rather than providing an overview of a given topic.

Sharing real or synthetic datasets could prove especially useful, for example, as a part of practical hands-on exercises or use cases. The ELIXIR Systems Biology Community will aim to (1) promote the use of appropriate data sets for testing and/or validation in different omics areas, underpinned by ELIXIR Core Data Resources such as [The Human Protein Atlas](#) (Uhlen *et al.*, 2010) or [MetaboLights](#) (Haug *et al.*, 2020), (2) provide a web-based platform for self-education via online tutorials, (3) establish a network of regular trainers and invited speakers, in order to readily communicate best practices and share materials.

***To provide support to trainers, in order to build a variety of training expertise that can meet the demands of this fast-growing field and the training needs of its users.*** Supporting the needs of current and future trainers is a key objective of the ELIXIR Systems Biology Community. Identifying specialised trainers from the TeSS network and creating a centralised repository of training materials, would help enormously with this task. In addition, joining the [train-the-trainer](#) events organised by ELIXIR will help improve the quality of training.

There is a lack of teaching courses that would be designed for trainees who wish to learn particular systems biology tools or methods in order to train others. The ELIXIR Systems Biology Community will facilitate the delivery of courses with the aim of (1) training trainees on how to use systems biology resources and (2) providing them with good training material

to disseminate the knowledge acquired during the course. The network of trainers will have access to a centralised repository of systems biology training materials, hosted by TeSS, which could save trainers a large amount of material preparation time. Such repositories may contain downloadable PowerPoint slides, lesson plans, high-resolution images in editable format, a list of web links to systems biology resources related to training, ranked based on relevance to a given topic, and models that can be run at a click. This may sound ambitious, but once created, the resources can be updated (e.g. replacing a tool with a newer one).

### Industrial embedding

Systems biology is a multidisciplinary endeavour to gain insights into the complexity of biological systems, pervasive to all colours of biotechnology (Kafarski, 2012)<sup>3</sup>. It is, thus, also bound to play a major role in the translation of this knowledge into applications of industrial, medical, agricultural and environmental interest. However, the systematic deployment of systems biology varies greatly across sectors, and its potential often remains insufficiently exploited. In this subsection, we map some of the major embeddings of systems biology in various industrial sectors and pinpoint key challenges.

Various biotechnological and pharmaceutical companies started their own systems biology programs years ago and stimulated academic parallels. Already in the 2000's, AstraZeneca instated a research chair in systems biology at the University of Manchester, UK, which then became instrumental in setting up largely academic research and doctoral training centres at that same and at other UK universities. In parallel, AstraZeneca and Pfizer developed and published the first systems biology models of signal transduction, relevant for the targeting of anticancer drugs. Bayer spun out a smaller

<sup>3</sup> See also <https://omgmopodcast.com/definition-of-biotechnology/>

company focusing on systems biology and data analysis. These initiatives also led to systems biology penetrating adjacent fields and setting up Systems Medicine, systems pharmacology, and systems toxicology.

The EU's Innovative Medicines Initiative (IMI), co-funded by industry and now superseded by the [Innovative Health Initiative](#), brought the necessity of these new disciplines to the fore. Many of the supported IMI projects lacked systematic long-term stability, owing to the short innovation life cycle in the pharmaceutical industry. Initiatives of this kind would benefit from long-term sustenance by a European research infrastructure. [TransQST](#), an IMI-funded consortium with industrial-academic partnership with links to ELIXIR's Toxicology Community, built novel systems toxicology models that enable translation from non-clinical to human safety during clinical trials. Quantitative data and resources generated within this initiative are sustainably disseminated through ELIXIR Core Data Resources, such as ChEMBL ([Gaulton et al., 2017](#)), ArrayExpress ([Athar et al., 2019](#)), and deposition repositories, including BioStudies ([Sarkans et al., 2018](#)) and BioModels.

As a systems medicine application in industry PK-PD (pharmacokinetics-pharmacodynamics) modelling, which involves mathematical approaches to study pharmacokinetics (PK), pharmacodynamics (PD), and their relationship ([Danhof et al., 2005](#); [Peck et al., 1992](#)), represents an essential component of the drug discovery and development pipeline. In particular, physiologically-based PK (PBPK) models, which include a realistic representation of physiology and its impact on PK, have a major interest in risk assessment ([Adler et al., 2011](#)). The process of drug absorption, distribution, metabolism, and elimination by the body (pharmacokinetics, PK) can be represented and better understood by using quantitative PK and notably PBPK models. Polymorphisms in drug metabolism enzymes and transporters can be taken into account by personalized modelling approaches, where models can be adapted to individual kinetics parameters. The pharmacological effects of a drug on the body taking into account the mechanism of drug action can be quantitatively described by using PD modelling. Although the most frequent approach in PK/PBPK modeling lacks biochemical details, combined PBPK-PD models may require for example active drug efflux pumps and kinetic details of drug metabolism in the liver by P450s [e.g. [Ploemen et al., 1997](#); [Sharma et al., 2020](#)]. It is also possible to include PBPK models as part of broader, multiscale, systems biology models ([Sluka et al., 2016](#)). PBPK modelling is also a key element of a safe-by-design approach for material development (for instance in the [VHP4Safety](#) EU project). The European Medicines Agency (EMA) and OECD have stipulated guidelines on qualification and reporting of PBPK analysis ([OECD, 2021](#); [Zhao, 2017](#)).

Quantitative Systems Pharmacology (QSP) is a converging point of biochemical pathway analyses and pharmacological modelling, and falls under the broader umbrella of systems biology. Much of such work has been funded by the IMI. The UK-QSP

network with the UK and international scientists in industry and academia is jointly funded by the Engineering and UK Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC) with financial assistance from AstraZeneca, Pfizer and GlaxoSmithKline.

There is a strong parallel between systems biology approaches described above for pharmacology and those used in toxicology. The same kinetics ("does it get there?") and dynamics ("what does it do?") modelling approaches apply. For example, PBPK models can be developed for drugs or for environmental pollutants. Mechanistic models used in toxicology like adverse outcome models and quantitative approaches for read-across, predicting endpoint information for one substance by using data from the same endpoint from (an)other substance(s), also have their counterparts in pharmacology.

Bioinformatics approaches towards mining literature and databases for comprehensive metabolic data have led to the development of maps based on genome-scale metabolic models (GEMs) for a multitude of organisms, including humans ([Robinson et al., 2020](#); [Thiele et al., 2013](#)). Constraint-based modelling ('FBA' and derivatives) of these GEMs is a systems biology product that enables the understanding of a variety of processes such as new metabolic ramifications of tumours ([Damiani et al., 2017](#)).

Analogous applications to microorganisms have led to multiple new insights into how these organisms may be engineered towards higher productivity ([Wehrs et al., 2019](#)). These types of activity have, at the European level, been linked with industry in various settings, including explicitly the ERA-net [CoBioTech Action](#) and implicitly the [Biobased Industries Joint Undertaking](#) [and its [successor](#)] promoting systems biology and synthetic biology as technology drivers to speed up research and innovation in industrial biotechnology. This is particularly relevant when addressing key priorities of the European Union with regards to a [green transition to a biobased, environmentally sustainable economy](#).

The [bio-economy](#) comprises the production and use of renewable resources from land and sea, and the use of waste to make value-added products, such as food, feed, bio-based products and bioenergy. In the EU, the bio-economy is worth an estimated €2 trillion, employing 9% of the workforce. It provides a unique opportunity for Europe to develop [sustainable processes](#) and address grand challenges, such as tackling the effects of climate change, achieving a clean environment, fostering industrial innovation, tailor-producing chemicals, contributing to healthy living and ensuring food security. Among the ways to address these challenges, the use of microbial organisms offers a valuable and powerful alternative to the fossil fuel-based economy as they can be engineered into cell factories producing fuels, high-value compounds, chemical building blocks, nutraceuticals and novel medicines (see [bioconsortium.eu](#) and the [ELIXIR Microbial Biotechnology Community](#)). This is best done through synthetic biology, which is frequently defined as the application of engineering principles to

biology. Such principles [model-driven design, modularization, standardisation (for example through consistent use of the synthetic biology open language - SBOL), separation of design and fabrication] enable streamlining the practice of biological engineering, to shorten the time required to Design, Build, Test and Learn (DBTL) biological systems. This streamlining of iterative design cycles facilitates the construction of more robust microbes that are better adapted to the target application and behave in a more predictable fashion. The streamlining benefits also hold for the engineering of consortia of interacting microorganisms, which allows to explore the wealth of microbial diversity. Furthermore, the individual processes have to be tackled by taking into account the whole value chain if they are to result in economically feasible innovations which can truly contribute to a shift from a petrochemical to a biobased economy (Kampers *et al.*, 2022)<sup>4</sup>.

**The principles, methods and (data and model) resources underpinning systems biology are crucial to reach these objectives.** Over the last 5 years, there has been a substantial increase in the number and variety of companies that both adopt and build upon these principles and technologies for their business development. A few examples include DSM, Lanzatech, Zymergen, Ginko Bioworks, Amyris, and Genomatica and, recently, a major push has been made by the USA government on this with substantial investment<sup>5</sup>. Crucial to these developments is the increasing focus on advanced, (semi-)automated infrastructures that enable rapid, tailored biomanufacturing of chemicals and materials, such as those by ESFRI programme IBISBA<sup>6</sup> and other biofoundries<sup>7</sup>.

The same basic principles and many methodologies are relevant when tackling challenges in agriculture, in particular in plant and animal sciences, as well as in ecology and water and soil management, and, crucially, in addressing climate change issues, such as reduction of emissions at source (see industrial biotechnology above), carbon capture and nutrient recycling.

### Systems Biology within ELIXIR

**The infrastructure aspects of systems biology and systems medicine - databases, tools and standards development as well as training and access to cloud infrastructure - are not only appropriate components of the ELIXIR infrastructure, but will be essential for ELIXIR's future support of advanced biological applications and personalised medicine. Our vision of how the different aspects of the systems biology life cycle map to the different entities within ELIXIR is represented in Figure 4.**

Because in systems biology experimental data relate to the emergence of function through computational models, and

<sup>4</sup> [https://www.cell.com/trends/biotechnology/fulltext/S0167-7799\(21\)00108-6](https://www.cell.com/trends/biotechnology/fulltext/S0167-7799(21)00108-6)

<sup>5</sup> <https://www.forbes.com/sites/johncumbers/2022/09/12/white-house-inks-strategy-to-grow-trillion-dollar-us-bioeconomy>

<sup>6</sup> [www.ibisba.eu](http://www.ibisba.eu)

<sup>7</sup> <https://biofoundries.org>; <https://agilebiofoundry.org>; <https://www.biomade.org>

because computational models need to be based on realistic experimental data, integrated data-and-model repositories such as *JWS Online*, *BioModels* and *Metabolic Atlas* are essential. Models and data need to be accessible to external users, for curation, validation and critical evaluation: the model-data repositories need to be 'live' in that they readily simulate the spatial-temporal behaviour of living systems. They should enable 'what-if' computational experiments *in silico*, in order to serve bioengineering, medicine, pharmacology and basic biology. Further, these repositories should allow for constant refinement and update of the models that they host. We can build on the landscape of data resources that ELIXIR already has in the form of Core Data Resources for knowledge used to build the models, and Deposition Databases for parameter tuning, validation, etc.

Standardised model annotation of the described components such as genes, reactions, or metabolites is required to ensure interoperability between models and FAIR repositories hosting large omics datasets. While extensive guidelines and test suites for such annotation have been developed in the COMBINE context and MEMOTE (Lieven *et al.*, 2020), respectively, their evolution and consistent application would benefit from a strong infrastructure context.

In relation to ELIXIR, the Systems Biology Community will enable a wide range of researchers to benefit from existing and proven systems biology and bioinformatics approaches, rather than to reinvent wheels. A number of applied Communities in ELIXIR, such as Rare Diseases, Food and Nutrition and Toxicology, form bridges to large research fields. We offer to support these ELIXIR Communities with the standards and best practices of the systems biology 'ecosystem' and build on ELIXIR Platforms to enable better understanding and reproducible analysis of complex systems.

### Synergies with ELIXIR Platforms

ELIXIR Communities and Platforms cover topics, including:

- Technology, like the Galaxy Community and Bioschemas (and Compute in general);
- Biological entity descriptions, in Core Data Resources, and in Communities like Metabolomics, Proteomics, Copy Number Variation and Intrinsically Disordered Proteins.
- Applied Communities like Rare Diseases, Plant Science, Food and Nutrition, and Toxicology.

The Systems Biology Community intends to become the link between these, providing the model ecosystem to the applied Communities, allowing the usage of data from those concerned with biological entities, and using the computational technology approaches and data infrastructure experiences and tools to do so. That linkage again builds on activities in the Interoperability Platform.

There are clear synergies between the objectives and activities of the Systems Biology Community and the ELIXIR Platforms, as described in the following sections.



simulations can be and are executed on general-purpose computational platforms. In the same way that ELIXIR is enabling its members to use national and European compute service providers for their data management needs, it could provide access for systems biology modelling. Through the Compute Platform we will also be able to link to EOSC activities related to Cloud data and workflow execution and also provide convenient access options via the Life Science AAI. Implementation of LS-AAI single sign-on user management for curation activity could help to broker efforts between biodatabases and to link data types, a potential point of interest for systems biology.

**The interoperability platform.** Standards are as big an issue in models as in data. Only data and models that use standardized file formats, metadata and vocabularies can truly be FAIR and used by the whole community. The Interoperability Platform's Recommended Interoperability Resources help to improve this FAIRness while the Platform's standards mapping resources, such as BridgeDB, identifiers.org and OLS, facilitate better interoperability and integration of data and models. In systems biology, the COMBINE consortium (Hucka *et al.*, 2015) is the international initiative that coordinates the development of standards, and the use of standards in systems biology is continually increasing. The harmonization between data and model standards, therefore, presents a great opportunity in the next decade, and ELIXIR is in a perfect position to catalyse such a development. A more systematic collaboration with COMBINE should lead to harmonisation of the standards landscape for systems biology. Practical collaboration between the hackathon initiatives from COMBINE, HARMONY, and the Biohackathon can also be expected to be fruitful. Because systems biology combines all the different aspects mentioned above, it is a testbed and impact case for interoperability approaches, further fueling new interoperability developments.

**The training platform.** The ELIXIR Training Platform remains instrumental in the training of the next generation of data science experts. While there have been plenty of smaller-scale initiatives for systems biology training in Europe at the same time, the overall scale of training in systems biology has been much smaller, mainly caused by a more fragmented systems biology community. To lift-up systems biology training Europe-wide, we propose the joining of the Systems Biology Community and data resources under a single roof. Upscaling training can benefit from FAIRification of training materials, a process which is well supported by the [ELIXIR FAIR Training Focus Group](#). Embedding systems biology training within the ELIXIR Training Platform via [TeSS](#) will also enable their integration into Learning Paths directly connecting to the activities of the [Learning Paths Focus Group](#). The [Training Platform](#) will help the ELIXIR Community to develop new training materials in a co-production model (i.e. resources allocated from both Community and the Training Platform) and through mini hackathons that will be hosted together with existing training resources, fully utilizing the standardized lesson template and the [Training Platform's GitHub](#) repository, as well as annotating them using BioSchemas.

## ELIXIR communities and focus groups

As indicated already in the section "Current Systems Biology activities in ELIXIR" above, systems biology is already a key component of the work of a number of ELIXIR Communities, and it can play an important role in others. Moreover, systems biology also plays a role in various Focus Groups. Some possible contributions of the Systems Biology Community to the ELIXIR Communities and Focus Groups are described below in a non-exhaustive manner.

### Technology-oriented Communities:

- *Galaxy*: In a recent Community-led Implementation Study, ELIXIR supported the integration of omics data access and analysis tools into Galaxy workflows. Building on this work, an initiative to improve connectivity between Galaxy and systems biology model repositories, development and simulation resources should provide a strong, practical boost to the ELIXIR Community-driven integration between Data and Models for Life.

### Biological Entity Communities:

- *Metabolomics*: Aimed towards the chemical, mechanistic and reaction flux-related aspects relevant for biomedical applications, microbial biotechnology, plant sciences, toxicology and nutrition, and the workflows and data interoperability aspects needed for those. Getting from metabolite levels to fluxes strongly links to quantitative dynamic models in systems biology. The Implementation Study "[Standardising the fluxomics workflows](#)" of the Metabolomics Community includes aspects of how ELIXIR could contribute to systems biology methods and workflow standardisation.

### Applied Communities:

- *Biodiversity*: This Community is focused on understanding and cataloguing the capabilities, interests and ongoing projects we have in this area across the ELIXIR Nodes. It develops appropriate connections with key external partners in the field. Large-scale systems biology models of interactions and evolution are bound to strengthen the area, as exemplified by various specific research projects across Europe.
- *Microbial biotechnology*: Aims to support the computational infrastructure underlying the Design - Build - Test - Learn (DBTL) cycle in the design of industrial microbes. It aims to contribute to addressing standardisation and other issues in relation to models in microbial biotechnology (e.g. semantic ontologies) and thereby contribute to a knowledge-based infrastructure for biotechnology.
- *Plant sciences*: An interdisciplinary group of researchers very active on the border between experimental and computational approaches that aims at building a tools service bundle that will support plant scientists in the integration and linking of diverse datasets with an

extension towards systems biology. Many of the current activities of the Community are in the field of experimental data management tools. Expertise from the systems biology community would in the short term enable the extension of these activities to the management of plant systems biology models which are becoming increasingly prominent in the field.

- *Food and nutrition community*: Aims to support the research towards the effects food choice and nutrition have on health and well-being. This typically is a system-wide effect where many small changes work in combination.
- *Toxicology*: Aims to support risk assessment of chemicals, drugs, cosmetics ingredients and nanomaterials to lead to safer products. This includes the combination of toxicokinetic (exposure, uptake, distribution, metabolism) and the toxicodynamic (molecular interactions and complex connected events in adverse outcome pathways). Both aspects and their combination strongly lean on models and modelling.
- *Microbiome*: Aims to develop a sustainable metagenomics infrastructure to enhance research and industrial innovation within the marine domain. It develops standards and best practices for the marine domain, provides databases specific to marine metagenomics and develops tools and pipelines to enhance metagenomics analyses. These goals will benefit from the standardisation efforts of systems biology elements and, in particular, from the deployment of metagenome scale (metabolic) models.

#### Human data communities:

- *Federated human data*: Activities in the domain of human data focus on the sharing of human data, predominantly but not exclusively in genomics making use of the increasingly sophisticated Federated EGA and Beacon infrastructures. Although focusing on genomic sequences, FEGA also accommodates phenotypic and disease information which has potential uses in systems biology and systems medicine.
- *Rare diseases*: Implementation Study “[ELIXIR Rare Diseases Infrastructure \(2019-21\)](#)”, although not explicitly addressing the needs of systems biology or systems medicine, addresses the linking of infrastructures needed to interpret data on rare diseases as well as collecting data that is “FAIR at source”. There is an opportunity to link these objectives to a systems medicine perspective on rare diseases.

ELIXIR also runs a number of focus groups (FGs) that are relevant to the Systems Biology Community:

- *Machine learning (ML)*: This Focus Group was initiated in October 2019 to address needs related to the application of ML in mining large omics datasets to uncover new insights in the field of medicine. These complement

activities in systems medicine. Goals of the ML focus group relate to the development of controlled terminology/ontology and services for ML model description. There is an opportunity to align these developments with efforts on standardisation of models in systems biology.

- *EOSC*: Connects ELIXIR’s EOSC-related activities, which run along the axes: a) Consolidation of e-Infrastructure services and positioning these services as an embedded “supply-chain” for data-intensive scientific collaborations; b) Open Science in practice; c) Integrating the user-focussed services from research infrastructures with their user communities to enable interdisciplinary research aligned with the major societal challenges. Systems biology and in particular its focus on model standards, workflows and model-driven activities, can contribute and help to expand the aforementioned EOSC ELIXIR activities.
- *Biocuration*: This group was established in 2021 and builds a network of database developers and curators in ELIXIR. The Systems Biology Community benefits from the interaction with the International Society for Biocuration and the engagement for a better visibility and recognition of the work of biocurators, e.g. community curation efforts in model resources.
- *FAIR training*: The FAIR Training FG was formed in 2018 with the aim of improving the production and diffusion of FAIR training materials across Nodes. These activities closely complement those of the Systems Biology Community whose applications often rely on multi-omic and holistic datasets and respective descriptors and identifiers for effective delivery of materials and tools of systems biology.
- *Learning Paths*: This group aims to foster the exchange of knowledge, ideas, and experiences with the aim of identifying needs, devising solutions, and advocating the widespread adoption of learning paths across ELIXIR, its Nodes, Communities, and beyond. This effort will address the current lack of guidance in developing curricula or structured training programs, with the ultimate goal of enhancing the learning experience for end users.

#### The community’s objectives

To illustrate the importance of infrastructure for systems biology in a forward-looking way, we have identified a number of potential challenges for the Systems Biology Community that would rely strongly on the infrastructure ELIXIR provides. These challenges align with those of other ELIXIR Communities, as outlined previously in this paper. They sit on a scale comparable to the grand challenges that ELIXIR tackles in its programme, namely:

- To deal with the increasing volume, complexity and heterogeneity of data,

- To enable the interoperability between data resources,
- To effectively use large, complex and heterogeneous data sets to generate actionable knowledge,
- To make it easier to find and deploy the right tools and to undergo training,
- To build data interpretation and modelling infrastructure following FAIR principles,
- To drive innovation and industry usage.

A major remit of systems biology is to quantitatively describe the dynamic, emergent interactions among the many components of biological systems. The goals are hence the generation of insights and knowledge, which can be translated into applications of industrial, environmental, nutritional, medical, ecological, and agricultural interest. For instance, predictive dynamic models of cells, organs and organisms have a potential for pre-testing drugs, producing new materials and chemicals for food & feed, understanding biogeochemical cycles, tackling carbon storage, and accelerating the shift from a petrochemical to bio-based economy. Models are crucial to understanding and fostering human and animal nutrition, host-microbiome interactions (plants, insects, animals, environment) and a range of other areas. These include the biochemical brain; systems ecology, agriculture and environment; individualised medicine enabling the prediction of the effects of 2,4-Dinitrophenol (DNP) in physiology and pathology; systems pharmacology enabling the individualised prediction of drugs' effects and toxicity, as well as model-driven production of tailored pharmaceuticals; systems epidemiology enabling the critical prediction of how government and therapeutic measures affecting pandemics such as that of COVID-19.

All these areas will benefit substantially from systematic and comprehensive bioinformatics and mechanistic and realistic modelling. Below we lay down five pillars around which this ELIXIR community aims to contribute to strengthen systems biology.

### Strengthening standardisation & interoperability

The standardisation needs in systems biology remain diverse. Data and models, as well as metadata of both, need consistent structuring following standardised formats. Close collaboration of ELIXIR with standardisation communities dedicated to modelling in the life sciences, such as COMBINE, as well as with relevant committees of standardisation bodies like CEN/CENELEC and ISO, such as the ISO committee for biotechnology standards (ISO/TC 276) with its working group WG5 "Data Processing and Integration" or the ISO committee for health informatics (ISO/TC 215), will ensure further development and adaptation of existing modelling standards to the needs of models shared via ELIXIR resources. Based on modelling standards like the ones from COMBINE (Golebiewski, 2019), model validation becomes more realistic. An advancement in this direction has been the release of the standardised genome-scale metabolic model testing tool

MEMOTE. Retrospectively applying this tool to existing models, however, remains an open challenge. We will seek to contribute to the standardisation and interoperability of data, operations and models.

**Modelling repositories.** Current work across ELIXIR nodes has begun to partly address these challenges. For example, Metabolic Atlas is promoting the use of a template code repository called *standard-GEM* for open-source genome-scale metabolic models. Conceptually similar to the COMBINE archive, standard-GEM establishes a folder and file format structure that fits with the iterative, versioned model maintenance process. In turn, such a standard structure enables future automatic validation with tools such as the aforementioned MEMOTE, and opens the door towards packaging with RO-Crate (Soiland-Reyes *et al.*, 2022) and potential integration with BioModels, JWS-Online and OpenEBench (Capella-Gutierrez *et al.*, 2017) via COMBINE.

Another ELIXIR effort is the established service MetaNetX, which cross-checks model annotation and reports inconsistencies in identifier mapping with respect to chemistry, in addition to facilitating cross-model mapping. Building on this knowledge-base, a potential future direction is the development of a service focused on assessing the quality of the annotation, which would complement the quantitative assessment that is already covered by MEMOTE.

On a wider-reaching level, model repositories like BioModels, FAIRDOMHub and its JWS-Online and others collaborate on the development of community standards, but they are only starting to co-ordinate their curation and dissemination activities. Currently users still need to access multiple repositories to discover all models potentially relevant to them. Moreover, a recent large-scale study (Tiwari *et al.*, 2021) of 455 published models showed that about half of the models could not be reproduced using the information in the manuscripts. Without coordination, many researchers might independently try and fail to reproduce a published model, wasting a lot of time and effort. Recognising these challenges, in the emerging ModeleXchange consortium, repositories are starting to coordinate model curation and discovery. This activity should be strengthened in the context of an ELIXIR Systems Biology Community, with significant user benefits.

**Standards for design and modelling.** Less developed are the standards related to the design of experiments and the description of these designs (which are required to generate standardised, FAIR data to be subsequently capitalised on by models). This is of particular importance for complex, multistep operations that are required for many processes, such as those pertaining to biobased production of chemicals, pharmaceuticals or materials. The ESFRI programme IBISBA ([www.ibisba.eu](http://www.ibisba.eu)) works on the interoperability and deployment of such concepts and standards and workflows, but many challenges remain. This is relevant for ELIXIR since these designs in the end lead to data covered in the infrastructure.

Standards for linking models with sensitive data, including electronic health records, as well as other person-related data and commercial data, with ELIXIR Core Data Resources and Deposition Databases are an identified need. ELIXIR and members of national ELIXIR Nodes are already active in defining such standards and connecting them to existing research projects in the domain as partners in the European standardisation initiative **EU-STANDS4PM** (European standardisation framework for data integration and data-driven in silico models for personalised medicine) that very recently has published guidelines and recommendations for data integration and model validation for computational models in the domain of clinical applications in personalised medicine (Collin *et al.*, 2022). Through the EU-STANDS4PM initiative, ELIXIR also supports the development of a series of ISO standards with recommendations and requirements for predictive computational models in personalised medicine research (ISO TS 9491). Such standardisation efforts need to be intensified and extended, given the increasing importance of modelling in the health domain. For this purpose, close collaborations with European initiatives, such as EU-STANDS4PM and those developing infrastructures for human digital twins (including corresponding standards for data and models) that are currently forming and will be funded in the near future, will help to jointly create a pan-European cloud-infrastructure for data integration and modelling in health research and personalised medicine.

**Interoperability between various forms of descriptive models and predictive models.** Some relevant model connections, which basically connect predictive analysis with data analysis have been developed and a few are in production. For now,

most exist as proof of concept rather than production ready services. The work done in the COVID-19 Disease Map project involved much manual curation and improvement of both the converter and the source (pathway) model and can even lead to updates for the standards used. The challenge is to streamline that process and to support curators to come to more interoperable models and FAIR descriptions of provenance and evidence.

A breakdown of the short, mid and long term objectives for the Standardisation & Interoperability theme is given in **Table 4**.

### Developing and deploying data and modelling technologies

Although the interconnectedness between modelling and experimentation is a hallmark of systems biology, its practical implementation remains challenging. Partly this is due to culture and insufficient training, but it also stems from difficulties in generating adequate, quantitative dynamic data that can support modelling and from the lack of models that are sufficiently accurate to handle the generated data. Addressing these challenges will require 1) an interface between big data and modelling frameworks, 2) integration of modelling approaches, including temporal and spatial modelling, and 3) application of the modelling results in a plethora of relevant domains, ranging from bioengineering at various scales to precision and personalised medicine. Integrating AI algorithms may significantly enhance the analysis of complex, quantitative dynamic data, enabling more accurate models that can effectively process and interpret large data sets. Further, developing AI-driven interfaces between big data and modelling frameworks

**Table 4. Breakdown of short, mid and long term objectives for the Standardisation & Interoperability theme.**

	Aims and objectives
<b>Short term (~3 years)</b>	<ul style="list-style-type: none"> <li>• Better support of existing standards in model repositories</li> <li>• Build upon systems biology models to improve the design of experiments that lead to the generation of higher quality, quantitative, FAIR data;               <ul style="list-style-type: none"> <li>• Address specific challenges for human modelling, which include:                   <ul style="list-style-type: none"> <li>◦ working with compartments;</li> <li>◦ model validation through standardised phenotypes;</li> <li>◦ initial interfaces for multi-level modelling and integration across scales; multi-tissue evaluations;</li> <li>◦ extrapolations from single cell analysis to tissue level;</li> <li>◦ microbiome - host interactions;</li> <li>◦ integrating sensitive personal data into models for personalised medicine</li> </ul> </li> </ul> </li> <li>• Establish approaches for model exchange, building on existing resource developments in the FAIR data landscape (e.g. FAIR data points; JWS-Online), BioModels and ModeXchange.</li> <li>• Improved interoperability of modelling, simulation and analysis tools</li> </ul>
<b>Mid term (~6 years)</b>	<ul style="list-style-type: none"> <li>• Improved standardisation of generation of complex data incl time-series, functional and imaging data for integration in computational models</li> <li>• Providing a link between existing models and datasets for easy access to relevant data</li> <li>• Good strategies (including training) to improve reproducibility, credibility, and validation of models and to assess the efficiency of tools, leading to the development of quality marks, thereby increasing the quality of workflow outcomes</li> </ul>
<b>Long term (~10 years)</b>	<ul style="list-style-type: none"> <li>• Improved interoperability of data and models to enable FAIR model connection and integration (at different scales) so as to facilitate the development of multi-scale modelling frameworks</li> </ul>



can streamline the integration of diverse data sources, improving the precision of models in systems biology.

**Systems medicine.** In the systems medicine domain, additional challenges result from 1) the sensitive nature of many of the data sources, 2) the complexity of disease phenotypes and mechanisms, especially in the context of precision medicine, and 3) the ethical and legal implications of using models and model predictions in clinical decision support. Solutions will be needed that enable the use of sensitive data to build models in a manner consistent with requirements for sensitive data. The community will also facilitate the development of new models and disease maps and of improved repositories to enable their sharing in a FAIR manner, in order to address the challenge of disease complexity. Finally, the Community will engage in ELSI activities to explore the challenges of using systems medicine models close to clinical practice.

**Models as a service.** Owing to the importance of modelling resources, methods, models, data, and expertise across the board (from dynamic to constraint-based, stochastic, statistical, and data modelling) it should be an aim to enhance applications in particular by systems biology novices, many of whom are deep experts in medicine, biology or biotechnology. This should be done by providing assistance to those novices in their use of the facilities offered by the Systems Biology Community as well as by a larger number of ELIXIR Communities and as per recommendations above. It should be made easy for the novices to find the most relevant model, to adjust it to their needs, to extend it, and to even make their own new model. This relates, most immediately, to activities on Make-Me-My

Model, Data for Modelling, Modelling for Data, multiscale modelling discussed above, but it will readily expand to many other ELIXIR areas.

**Digital twins of biological systems.** Systems biology models build upon data for predictions and thereby truly bring data to life. The **combination of models, big data and AI** could enable designing Digital Twins of biological systems and thereby enhance the possibilities to explore, understand, design and predict biological behaviour. For many years, perhaps decades, mechanistic models have provided a clear-box approach to modelling, allowing researchers to be in control of the abstractions, and facilitating a transparent mapping to biological processes. AI models, in comparison, are prone to create difficulties in understanding their inner workings, and much more so in the large language models developed in recent years. Interfacing the two approaches could be a way to bridge the gap between the explainable and unexplainable modelling approaches, e.g., by leveraging hypotheses. Therefore, it is proposed to develop dedicated projects across Platforms, Communities, and Focus Groups (such as the Machine Learning Focus Group) to smoothly integrate systems biology models with Big Data analytics and to thereby stimulate dedicated activities underpinning the development and deployment of Digital Twins for the whole range of applications in the Life Sciences. This extends from the design of highly efficient cell-based industrial and pharmaceutical processes through decision-support systems in health and disease, to integrated farming systems and ecosystem management.

A breakdown of the short, mid and long term objectives for the Technology theme is given in [Table 5](#).

**Table 5. Breakdown of short, mid and long term objectives for the Technology theme.**

	Aims and objectives
<b>Short term (~3 years)</b>	<ul style="list-style-type: none"> <li>• How big / smart data meets models meaningfully</li> <li>• Intertwining temporal and spatial modelling appropriately</li> <li>• Interfacing to synthetic biology through model-based design and model-based-learning strategies</li> </ul>
<b>Mid term (~6 years)</b>	<ul style="list-style-type: none"> <li>• Developing good strategies (and training therein) for validating models and for checking the efficiency of tools leading to quality marks; and tools assessing predictions;</li> <li>• Developing theoretical and practical multi-scale modelling frameworks;</li> <li>• Providing the basis for developing Digital Twins (microbes, bioreactors, organs, organisms, ecosystems)</li> </ul>
<b>Long term (~10 years)</b>	<ul style="list-style-type: none"> <li>• Deploying Digital Twin methodologies that provide sufficiently accurate, real-time and dynamic depictions of physical biosystems</li> <li>• Steer and modify processes, stratify patients or support decision-making</li> <li>• Increase uptake of systems biology methodologies by the communities of biologists, bioengineers and physicians;</li> <li>• Increase the uptake of standards (e.g. for model and data reporting) by the world wide systems biology communities.</li> </ul>

### Building capacity and providing training

Whereas the separate curricula in physical sciences, data-focused life sciences, and computer science have become quite effective in training in their own disciplines, they are becoming inadequate to train multidisciplinary teams which need to tackle increasingly complex problems. Global health challenges such as non-communicable diseases, pandemics, or diseases stemming from environmental factors have all been addressed by systems biology models. When supplemented with ELIXIR earmarked information, these models should soon become ready for use in the clinic, especially because they are the only tools truly to handle the increased call from the general public for personalised medicine. This call has also been realised by the European Parliament and the European Commission. A vast increase particularly in transdisciplinary training is necessary now and ELIXIR-training is well-posed to set this up. The [joint ELIXIR-ISBE course on Corona \(SARS-CoV-2\) epidemiology](#) may serve as an example.

Several independent, ongoing training activities already deliver systems biology training modules within ELIXIR (e.g. [Table 3](#)). With the exponential growth of biological data, there is a lag in the identification and generation of new ones. National Training Coordinators may assist in flagging when such courses are missing from TeSS, and further promote Node-interaction when such competencies are not present by organising ELIXIR-level international courses. A possibility might be interacting with non-profit organisations working in systems biology-related areas and education such as iGEM.

Systems biologists often have to cope with scattered knowledge resources. Hence, a well-balanced and consistent set of competences are required that are compatible across the ELIXIR Nodes. We propose to implement a programme of organisational capacity building, including specific training in gap areas, advanced training, knowledge sharing and staff exchanges to build a well-developed and interconnected Systems Biology Community. We will make use of the ELIXIR's training portal TeSS by integrating different tools and services relevant to systems biology and making it available to all Nodes. Synergising the training resources, the needs of trainees and trainers, and the communication with other ELIXIR communities is an overarching aim for the Capacity building/Training task. This will be achieved by integrating new and existing systems biology-related courses under a single umbrella towards the different objectives listed in [Table 6](#). The corresponding activities will support the needs of current and future trainers long term, and centralise the use of systems biology materials.

### Fostering industrial and societal embedding

Systems-level understanding and analysis of huge amounts of experimental data is required for different 'industries'

such as hospitals, pharmaceutical industry, biotechnological companies, health care institutions, regulatory agencies, and government. However, the potential significance of systems approaches in industrial sectors has been exploited insufficiently. Systems biology aims to develop quantitative and conceptual understanding of biological phenomena. This comes with the modelling and prediction of complex processes such as functions of the human brain, ecosystem function or host-microbiome interactions. The ability to model and predict what happens to a biological system under some conditions may have a profound impact on industrial applications as diverse as the identification of the best drug candidates using PK-PD models, sustainable production of biobased chemicals and materials through model-driven designs, improvement of crop production strategies, or COVID-19 management.

Quantitative data and resources generated within IMI-funded consortia with industrial academic partnerships like TransQST, which aims to build novel systems toxicology models, have been made available through ELIXIR Core Data Resources and deposition repositories. These collaborations are valuable to ELIXIR's Toxicology Community. Similar initiatives for systems pharmacology have been supported by IMI, major UK funding bodies and pharmaceutical companies like AstraZeneca, Pfizer and GlaxoSmithKline. Systems biology and synthetic biology will be particularly relevant to address the five out of the [seven key challenges](#) identified by the European Union that are related to health and environmental sustainability. Hence, efforts towards transition into more environmentally sustainable economies offer unprecedented opportunities for a range of subfields of systems biology.

The ELIXIR Systems Biology Community will take an active role in the use of ELIXIR resources and in the definition of activities aiming to develop new industrial collaborations and to strengthen existing ones. These activities will be aligned to [ELIXIR's Industry Strategy](#) by (1) facilitating collaborations between researchers in academia and industry, (2) enabling the use of ELIXIR resources by industry and (3) engaging effectively yet appropriately with the private sector. One way to engage with the industrial sector is through joint workshops and collaboration with other ELIXIR Communities interested in working with industry. Other objectives to achieve industrial embedding are listed in [Table 7](#). An important aspect of these objectives is the use of Key Performance Indicators (KPIs) and "gap analysis surveys" to measure overall long term performance and to identify priorities for improvement. The potential is enormous, given the capabilities of the deployment of data and models to describe biological systems and to enable actionable knowledge for a vast range of translational applications.

**Table 6. Breakdown of short, mid and long term objectives for Capacity building/Training theme.**

	Aims and objectives
<b>Short term (~ 3 years)</b>	<ul style="list-style-type: none"> <li>• Pre-screen trainees prior to training events to make recommendations for courses to be followed in the context of the event</li> <li>• Integrate new systems biology courses into TeSS and co-promote them with existing TeSS courses;</li> <li>• Strengthen synergies with the other ELIXIR Communities, e.g. via joint training events</li> </ul>
<b>Mid term (~ 6 years)</b>	<ul style="list-style-type: none"> <li>• Extend the use of synthetic and standardised datasets in most systems biology training events</li> <li>• Support current and future trainers via Train the Trainer ELIXIR events</li> </ul>
<b>Long term (~ 10 years)</b>	<ul style="list-style-type: none"> <li>• Create a centralised repository of systems biology training materials aggregated by TeSS</li> <li>• Systematically review trends in systems biology and update the training resources accordingly</li> </ul>

**Table 7. Breakdown of short, mid and long term objectives for the Industrial embedding theme.**

	Aims objectives
Short term (~ 3 years)	<ul style="list-style-type: none"> <li>• Together with the Training Platform set up a “gap analysis survey” to find out the forces and needs for each ELIXIR Node</li> <li>• Implementation of turnkey solutions to different Nodes (specific training, staff exchanges, knowledge exchanges etc.)</li> <li>• Set up KPIs to measure the impact of different actions</li> </ul>
Mid term (~ 6 years)	<ul style="list-style-type: none"> <li>• Continue and review the process of gap analysis survey and KPIs to include SME and industry</li> <li>• Involve small and medium-size enterprises in the capacity building process</li> <li>• Identify new themes and challenges in systems biology</li> </ul>
Long term (~ 10 years)	<ul style="list-style-type: none"> <li>• Consolidation of the capacity building process for new partners (communities, countries etc.)</li> </ul>

## Data availability

No data are associated with this article.

## Acknowledgments

We thank ELIXIR for financial and administrative support during the development of this white paper and its support of the ELIXIR Systems Biology Focus Group.

## References

- Adler S, Basketter D, Creton S, *et al.*: **Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010**. *Arch Toxicol.* 2011; **85**(5): 367–485.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Amberger JS, Bocchini CA, Scott AF, *et al.*: **OMIM.org: leveraging knowledge across phenotype-gene relationships**. *Nucleic Acids Res.* 2019; **47**(D1): D1038–D1043.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Apostolopoulos Y, Lich KH, Lemke M: **Complex systems and population**

**health: a primer**. Oxford University Press, New York, NY, 2020.

### Reference Source

- Athar A, Füllgrabe A, George N, *et al.*: **ArrayExpress update - from bulk to single-cell expression data**. *Nucleic Acids Res.* 2019; **47**(D1): D711–D715.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bairoch A: **The cellosaurus, a cell-line knowledge resource**. *J Biomol Tech.* 2018; **29**(2): 25–38.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Barillari C, Ottoz DSM, Fuentes-Serna JM, *et al.*: **OpenBIS ELN-LIMS: an**

- open-source database for academic laboratories. *Bioinformatics*. 2016; **32**(4): 638–640.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Barrette AM, Bouhaddou M, Birtwistle MR: **Integrating transcriptomic data with mechanistic systems pharmacology models for virtual drug combination trials.** *ACS Chem Neurosci*. 2018; **9**(1): 118–129.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Béal J, Pantolini L, Noël V, et al.: **Personalized logical models to investigate cancer response to BRAF treatments in melanomas and colorectal cancers.** *PLoS Comput Biol*. 2021; **17**(1): e1007900.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Beard N, Bacall F, Nenadic A, et al.: **TeSS: a platform for discovering life-science training opportunities.** *Bioinformatics*. 2020; **36**(10): 3290–3291.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bhinder B, Elemento O: **Towards a better cancer precision medicine: systems biology meets immunotherapy.** *Curr Opin Syst Biol*. 2017; **2**: 67–73.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Capella-Gutierrez S, de la Iglesia D, Haas J, et al.: **Lessons Learned: Recommendations for Establishing Critical Periodic Scientific Benchmarking.** *Bioinformatics*. bioRxiv. 2017.  
[Publisher Full Text](#)
- Caspi R, Billington R, Keseler IM, et al.: **The MetaCyc database of metabolic pathways and enzymes - a 2019 update.** *Nucleic Acids Res*. 2020; **48**(D1): D445–D453.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chang A, Jeske L, Ulbrich S, et al.: **BRENDA, the ELIXIR core data resource in 2021: new developments and updates.** *Nucleic Acids Res*. 2021; **49**(D1): D498–D508.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Collin CB, Gebhardt T, Golebiewski M, et al.: **Computational models for clinical applications in personalized medicine—guidelines and recommendations for data integration and model validation.** *J Pers Med*. 2022; **12**(2): 166.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crawford N, Salvucci M, Hellwig CT, et al.: **Simulating and predicting cellular and in vivo responses of colon cancer to combined treatment with chemotherapy and IAP antagonist Birinapant/TL32711.** *Cell Death Differ*. 2018; **25**(11): 1952–1966.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Czarnewski P, Mahfouz A, Calogero RA, et al.: **Community-driven ELIXIR activities in single-cell omics [version 1]; peer review: 2 approved with reservations].** *F1000Res*. 2022; **11**(ELIXIR): 869.  
[Publisher Full Text](#)
- da Veiga Leprevost F, Grüning BA, Aflitos SA, et al.: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics*. 2017; **33**(16): 2580–2582.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Damiani C, Di Filippo M, Pescini D, et al.: **popFBA: tackling intratumour heterogeneity with Flux Balance Analysis.** *Bioinformatics*. 2017; **33**(14): i311–i318.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danhof M, Alvan G, Dahl SG, et al.: **Mechanism-based pharmacokinetic-pharmacodynamic modeling—a new classification of biomarkers.** *Pharm Res*. 2005; **22**(9): 1432–1437.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Demir E, Cary MP, Paley S, et al.: **The BioPAX community standard for pathway data sharing.** *Nat Biotechnol*. 2010; **28**(9): 935–942.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- du Lac M, Duigou T, Hérisson J, et al.: **Galaxy-SynBioCAD: synthetic biology design automation tools in Galaxy workflows.** *Bioengineering, bioRxiv*. 2020.  
[Publisher Full Text](#)
- Ebata K, Yamashiro S, Iida K, et al.: **Building patient-specific models for receptor tyrosine kinase signaling networks.** *FEBS J*. 2022; **289**(1): 90–101.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Eduati F, Jaaks P, Wappler J, et al.: **Patient-specific logic models of signaling pathways from screenings on cancer biopsies to prioritize personalized combination therapies.** *Mol Syst Biol*. 2020; **16**(2): e8664.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- European Biopharmaceutical Enterprises: **EBE White Paper on Personalised Medicine**. 2015.  
[Reference Source](#)
- Fey D, Halasz M, Dredax D, et al.: **Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients.** *Sci Signal*. 2015; **8**(408): ra130.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fröhlich F, Kessler T, Weindl D, et al.: **Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model.** *Cell Syst*. 2018; **7**(6): 567–579.e6.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gaulton A, Hersey A, Nowotka M, et al.: **The ChEMBL database in 2017.** *Nucleic Acids Res*. 2017; **45**(D1): D945–D954.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gawron P, Ostaszewski M, Satagopam V, et al.: **MINERVA—a platform for visualization and curation of molecular interaction networks.** *NPJ Syst Biol Appl*. 2016; **2**: 16020.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gillespie M, Jassal B, Stephan R, et al.: **The reactome pathway knowledgebase 2022.** *Nucleic Acids Res*. 2022; **50**(D1): D687–D692.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Glont M, Arankalle C, Tiwari K, et al.: **BioModels parameters: a treasure trove of parameter values from published systems biology models.** *Bioinformatics*. 2020; **36**(17): 4649–4654.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Goble C, Soiland-Reyes S, Bacall F, et al.: **Implementing FAIR digital objects in the EOSC-life workflow collaborative.** 2021.  
[Publisher Full Text](#)
- Golebiewski M: **Data formats for systems biology and quantitative modeling.** In: *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019; **2**: 884–893.  
[Publisher Full Text](#)
- Hastings JF, O'Donnell YEI, Fey D, et al.: **Applications of personalised signalling network models in precision oncology.** *Pharmacol Ther*. 2020; **212**: 107555.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hatos A, Quaglia F, Piovesan D, et al.: **APICURON: a database to credit and acknowledge the work of biocurators.** *Database (Oxford)*. 2021; **2021**: baab019.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Haug K, Cochrane K, Nainala VC, et al.: **MetaboLights: a resource evolving in response to the needs of its scientific community.** *Nucleic Acids Res*. 2020; **48**(D1): D440–D444.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hucka M, Nickerson DP, Bader GD, et al.: **Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative.** *Front Bioeng Biotechnol*. 2015; **3**: 19.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hutter C, Zenklusen JC: **The Cancer Genome Atlas: creating lasting value beyond its data.** *Cell*. 2018; **173**(2): 283–285.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ison J, Rapacki K, Ménager H, et al.: **Tools and data services registry: a community effort to document bioinformatics resources.** *Nucleic Acids Res*. 2016; **44**(D1): D38–D47.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ivanovic S, El-Kebir M: **Modeling and predicting cancer clonal evolution with reinforcement learning.** *Genome Res*. 2023; **33**(7): 1078–1088.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jupp S, Burdett T, Malone J, et al.: **A new Ontology Lookup Service at EMBL-EBI.** *SWAT4LS*. 2015; **2**: 118–119.  
[Reference Source](#)
- Kafarski P: **Rainbow code of biotechnology.** *Chemik*. 2012; **66**(8): 811–816.  
[Reference Source](#)
- Kampers LFC, Asin-Garcia E, Schaap PJ, et al.: **Navigating the Valley of Death: perceptions of industry and academia on production platforms and opportunities in biotechnology.** *EFB Bioeconomy J*. 2022; **2**: 100033.  
[Publisher Full Text](#)
- Kanehisa M, Araki M, Goto S, et al.: **KEGG for linking genomes to life and the environment.** *Nucleic Acids Res*. 2008; **36**(Database issue): D480–D484.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kezer CA, Shah VH, Simonetto DA: **Advances in predictive modeling using Machine Learning in the field of hepatology.** *Clin Liver Dis (Hoboken)*. 2021; **18**(6): 288–291.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- King RD, Rowland J, Oliver SG, et al.: **The automation of science.** *Science*. 2009; **324**(5923): 85–89.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kolodkin A, Alberghina L, Snoep JL, et al.: **Infrastructure Systems Biology Europe (ISBE): emergence of innovative systems biology servicing.** In: *BioSB-2018 4th Dutch Bioinformatics & Systems Biology Conference Congressentrum De Werelt*. abstract book, Lunteren 15–16 May 2018; 71.  
[Publisher Full Text](#)
- Lieven C, Beber ME, Olivier BG, et al.: **MEMOTE for standardized genome-scale metabolic model testing.** *Nat Biotechnol*. 2020; **38**(3): 272–276.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Malik-Sheriff RS, Glont M, Nguyen TVN, et al.: **BioModels-15 years of sharing computational models in life science.** *Nucleic Acids Res*. 2020; **48**(D1): D407–D415.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Martens M, Ammar A, Riutta A, et al.: **WikiPathways: connecting communities.** *Nucleic Acids Res*. 2021; **49**(D1): D613–D621.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mazein A, Ostaszewski M, Kuperstein I, et al.: **Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms.** *NPJ Syst Biol Appl*. 2018; **4**: 21.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Mendes P, Hoops S, Sahle S, *et al.*: **Computational modeling of biochemical networks using COPASI**. In: Maly, I.V. (ed), *Systems Biology. Methods Mol Biol.* Humana Press, Totowa, NJ, 2009; **500**: 17–59.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moretti S, Tran VDT, Mehl F, *et al.*: **MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models**. *Nucleic Acids Res.* 2021; **49**(D1): D570–D574.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nickerson D, Atalag K, de Bono B, *et al.*: **The human physiome: how standards, software and innovative service infrastructures are providing the building blocks to make it achievable**. *Interface Focus.* 2016; **6**(2): 20150103.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- OECD: **Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes**. *OECD Series on Testing and Assessment*, No. 331, Environment, Health and Safety, Environment Directorate, OECD, 2021; **331**.  
[Reference Source](#)
- Ostaszewski M, Niarakis A, Mazein A, *et al.*: **COVID19 disease map, a computational knowledge repository of virus-host interaction mechanisms**. *Mol Syst Biol.* 2021; **17**(10): e10387.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peck CC, Barr WH, Benet LZ, *et al.*: **Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development**. *J Pharm Sci.* 1992; **81**(6): 605–610.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Peters M, Eicher JJ, van Niekkerk DD, *et al.*: **The JWS online simulation database**. *Bioinformatics.* 2017; **33**(10): 1589–1590.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ploemen JPHTM, Wormhoudt LW, Haenen GR, *et al.*: **The use of human in vitro metabolic parameters to explore the risk assessment of hazardous compounds: the case of ethylene dibromide**. *Toxicol Appl Pharmacol.* 1997; **143**(1): 56–69.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Robinson JL, Kocabaş P, Wang H, *et al.*: **An atlas of human metabolism**. *Sci Signal.* 2020; **13**(624): eaaz1482.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rodchenkov I, Babur O, Luna A, *et al.*: **Pathway commons 2019 update: integration, analysis and exploration of pathway data**. *Nucleic Acids Res.* 2020; **48**(D1): gkz946.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roth YD, Lian Z, Pochiraju S, *et al.*: **Datanator: an integrated database of molecular data for quantitatively modeling cellular behavior**. *Nucleic Acids Res.* 2021; **49**(D1): D516–D522.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sarkans U, Gostev M, Athar A, *et al.*: **The bioStudies database—one stop shop for all data supporting a life sciences study**. *Nucleic Acids Res.* 2018; **46**(D1): D1266–D1270.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Satagopam V, Gu W, Eifes S, *et al.*: **Integration and visualization of translational medicine data for better understanding of human diseases**. *Big Data.* 2016; **4**(2): 97–108.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Scharm M, Wolkenhauer O, Waltemath D: **An algorithm to detect and communicate the differences in computational models describing biological systems**. *Bioinformatics.* 2016; **32**(4): 563–570.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schreiber F, Sommer B, Czauderna T, *et al.*: **Specifications of standards in systems and synthetic biology: status and developments in 2020**. *J Integr Bioinforma.* 2020; **17**(2–3): 20200022.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shannon P, Markiel A, Ozier O, *et al.*: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res.* 2003; **13**(11): 2498–2504.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sharma RP, Kumar V, Schuhmacher M, *et al.*: **Development and evaluation of a harmonized whole body Physiologically Based Pharmacokinetic (PBPK) model for flutamide in rats and its extrapolation to humans**. *Environ Res.* 2020; **182**: 108948.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sherry ST, Ward MH, Kholodov M, *et al.*: **dbSNP: the NCBI database of genetic variation**. *Nucleic Acids Res.* 2001; **29**(1): 308–311.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sluka JP, Fu X, Swat M, *et al.*: **A liver-centric multiscale modeling framework for xenobiotics**. *PLoS One.* 2016; **11**(9): e0162428.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Soiland-Reyes S, Sefton P, Crosas M, *et al.*: **Packaging research artefacts with RO-Crate**. *Data Sci.* 2022; **5**(2): 97–138.  
[Publisher Full Text](#)
- Stanford NJ, Scharm M, Dobson PD, *et al.*: **Data management in computational systems biology: exploring standards, tools, databases, and packaging best practices**. In: Oliver, S.G. and Castrillo, J.I. (eds), *Yeast Systems Biology*. Springer New York, New York, NY. *Methods Mol Biol.* 2019; **2049**: 285–314.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Stanford NJ, Wolstencroft K, Golebiewski M, *et al.*: **The evolution of standards and data management practices in systems biology**. *Mol Syst Biol.* 2015; **11**(12): 851.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Subramanian A, Tamayo P, Mootha VK, *et al.*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A.* 2005; **102**(43): 15545–15550.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Subramanian I, Verma S, Kumar S, *et al.*: **Multi-omics data integration, interpretation, and its application**. *Bioinforma Biol Insights.* 2020; **14**: 1177932219899051.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Szklarczyk D, Gable AL, Nastou KC, *et al.*: **The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets**. *Nucleic Acids Res.* 2021; **49**(D1): D605–D612.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tekle KM, Gundersen S, Klepper K, *et al.*: **Norwegian e-Infrastructure for Life Sciences (NeLS) [version 1; peer review: 2 approved]**. *F1000Res.* 2018; **7**: ELIXIR–968.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- The Gene Ontology Consortium: **The gene ontology resource: 20 years and still going strong**. *Nucleic Acids Res.* 2019; **47**(D1): D330–D338.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Thiele I, Swainston N, Fleming RMT, *et al.*: **A community-driven global reconstruction of human metabolism**. *Nat Biotechnol.* 2013; **31**(5): 419–425.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tiwari K, Kananathan S, Roberts MG, *et al.*: **Reproducibility in systems biology modelling**. *Mol Syst Biol.* 2021; **17**(2): e9982.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Uhlen M, Oksvold P, Fagerberg L, *et al.*: **Towards a knowledge-based human protein atlas**. *Nat Biotechnol.* 2010; **28**(12): 1248–1250.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- van Iersel MP, Pico AR, Kelder T, *et al.*: **The Bridgedb framework: standardized access to gene, protein and metabolite identifier mapping services**. *BMC Bioinformatics.* 2010; **11**: 5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Waltemath D, Golebiewski M, Blinov ML, *et al.*: **The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE)**. *J Integr Bioinforma.* 2020; **17**(2–3): 20200005.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Waltemath D, Karr JR, Bergmann FT, *et al.*: **Toward community standards and software for whole-cell modeling**. *IEEE Trans Biomed Eng.* 2016; **63**(10): 2007–2014.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wehrs M, Tanjore D, Eng T, *et al.*: **Engineering robust production microbes for large-scale cultivation**. *Trends Microbiol.* 2019; **27**(6): 524–537.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wilkinson MD, Dumontier M, Aalbersberg JJJ, *et al.*: **The FAIR guiding principles for scientific data management and stewardship**. *Sci Data.* 2016; **3**: 160018.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Willighagen EL, Mayfield JW, Alvarsson J, *et al.*: **The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching**. *J Cheminform.* 2017; **9**(1): 33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wimalaratne SM, Juty N, Kunze J, *et al.*: **Uniform resolution of compact identifiers for biomedical data**. *Sci Data.* 2018; **5**: 180029.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wittig U, Rey M, Weidemann A, *et al.*: **SABIO-RK: an updated resource for manually curated biochemical reaction kinetics**. *Nucleic Acids Res.* 2018; **46**(D1): D656–D660.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wolkenhauer O: **Systems medicine: integrative, qualitative and computational approaches**. Academic Press, 2020.  
[Reference Source](#)
- Wolstencroft K, Krebs O, Snoep JL, *et al.*: **FAIRDOMHub: a repository and collaboration environment for sharing systems biology research**. *Nucleic Acids Res.* 2017; **45**(D1): D404–D407.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wolstencroft K, Owen S, Krebs O, *et al.*: **SEEK: a systems biology data and model management platform**. *BMC Syst Biol.* 2015; **9**: 33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zanin M, Aitya NAA, Basilio J, *et al.*: **An early stage researcher's primer on systems medicine terminology**. *Netw Syst Med.* 2021; **4**(1): 2–50.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhao P: **Report from the EMA workshop on qualification and reporting of Physiologically Based Pharmacokinetic (PBPK) modeling and simulation**. *CPT Pharmacomet Syst Pharmacol.* 2017; **6**(2): 71–72.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status: ? ? ✓

---

## Version 1

Reviewer Report 23 January 2023

<https://doi.org/10.5256/f1000research.139171.r156728>

© 2023 Androulakis I. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Ioannis Androulakis** 

<sup>1</sup> Biomedical Engineering Department and Chemical & Biochemical Engineering Department, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

<sup>2</sup> Biomedical Engineering Department and Chemical & Biochemical Engineering Department, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

This is a very comprehensive report/account of systems biology activities. I do not think I have any reservations. I realize this is more of a position statement and account of capabilities and possibilities and less of a concrete discussion of specific applications. Therefore, I am assessing the manuscript as such. It does present an overwhelming picture - possibly raising concerns about the applicability - but I also realize that the purpose of this report is to present the current state of the art.

This is admittedly a very ambitious undertaking. In that respect, I would also be curious to get the authors' opinion on whether or how such an integrative approach could eventually extend beyond large groups and integrated teams and be of assistance and value to smaller academic or research groups.

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** quantitative systems biology and pharmacology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response ( ) 17 Apr 2024

**John Hancock**

We thank the reviewer for their supportive comments.

With reference to making this initiative helpful to smaller as well as large groups, we have added two sentences on this topic (lines 326-329)

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 12 January 2023

<https://doi.org/10.5256/f1000research.139171.r156725>

© 2023 Sauro H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Herbert Sauro** 

<sup>1</sup> Department of Bioengineering, University of Washington, Seattle, WA, USA

<sup>2</sup> Department of Bioengineering, University of Washington, Seattle, WA, USA

dos Santos et al, provide a comprehensive summary of the current state of systems biology modeling as well as the challenges they see in the future within the ELIXIR framework. One of the important discussion points is suggestions for short-, mid-, and long-term goals within systems biology. The list of suggestions, however, would not have been out of place if published 20 years ago. This highlights that we still have a long way to go. For example, one of the long-term aims is improved interoperability, something that the community has been attempting to achieve for a long time. Many of the goals have a similar ring to them, though some are more interesting than others. For example, one of the mid-term goals is developing the basis for theoretical and practical multi-scale modeling. I consider this to be a major issue in multi-scale modeling which I feel urgently needs addressing. Another thoughtful goal is to use models to improve the design of experiments, that is, model-driven rather than data-driven approaches.

Overall, the paper presents a good coverage of the efforts being undertaken in Europe and to some extent elsewhere, although some things appear to be obviously missing. For example, there is no mention of SBOL (Synthetic Biology open language) standard, even though synthetic biology is frequently mentioned in the text and is a key aspect in both the article's short-term and mid-

term goals. The biggest gap, however, is that there is very little discussion of machine learning or AI. Having said that, I should say I am firmly a card-carrying mechanistic modeler. Machine learning is a two-edged sword. On the one hand, it's very good at finding patterns in data, but it's also very good at hiding any understanding of why the pattern exists at all. As sentient beings, we don't want to completely give up our intellect to a machine, however, there must be some halfway point where machine learning can inform our mechanistic models, help us determine what data we should collect (model-driven), or supply useful hypotheses to peruse. The ELIXIR program has a unique opportunity to augment its long terms goals by weaving machine learning into our mechanistic view of the world without losing our ability to understand reality. However, this is a basic research effort that will take time to develop and would probably belong to the list of long-term goals.

Another missing component in their long-term goals is fostering tighter collaboration with the USA and other countries in the systems biology field, particularly with joint grant awards. This of course, has, and is happening. Most notably, many of the popular systems biology standards were the result of close collaboration between groups in the US and Europe. This is still very much an active area, but a statement in the list of long-term goals would help emphasize that continuing collaboration is very important to the success of the field.

Finally, I think one of the long-term goals should be to advertise success stories more strongly. Europe, in particular, probably has more success stories in mechanistic modeling than any other trading block. This is particularly the case in understanding metabolism, where groups led by individuals such as Jacky Snoep or Bas Teusink and co-workers have made significant contributions to our understanding of energy metabolism using systems biology approaches.

In summary, the article is interesting. It provides a status report on systems biology but perhaps misses some opportunities for future opportunities.

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Systems Biology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have**



**significant reservations, as outlined above.**

Author Response ( ) 17 Apr 2024

**John Hancock**

- We have added mention of SBOL, particularly where we mention the COMBINE initiative in the "Data and metadata" but also elsewhere in the document
- With reference to AI and machine learning, We thank the reviewer for the suggestion, which has been added into the manuscript (lines 221-222)
- With reference to fostering closer links to the US, this has also been added as a long-term goal (lines 223-224).
- With reference to advertising success stories, we added this as a mid-term goal (line 203)

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 28 November 2022

<https://doi.org/10.5256/f1000research.139171.r155195>

© 2022 Vodovotz Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Yoram Vodovotz**<sup>1</sup> Department of Surgery, University of Pittsburgh, Pittsburgh, PA, USA<sup>2</sup> Department of Surgery, University of Pittsburgh, Pittsburgh, PA, USA

Dos Santos et al present a white paper outlining the overall rationale, current work, and future plans of the ELIXIR consortium. The manuscript is overall logical and well-written. However, there are some points that the authors should consider and that would improve the manuscript overall:

1. Overall, the thrust of the manuscript and ELIXIR as a whole seems to center on the concept that Systems Biology is predicated on the generation of mechanistic mathematical models, with datasets envisioned as being used for calibration/verification/validation of these mechanistic models. While this is certainly a major goal for many in this field, there seems to be little mention of machine learning (i.e. data-driven modeling) as an alternative (or, ideally, an adjunct approach that could synergize with or integrate machine learning with mechanistic modeling), though there is mention of a focus group on machine learning and a short term goal stated as "Understanding how big/smart data meets models meaningfully." It would seem that there is a need to establish workflows and pipelines wherein data are generated, data-driven modeling/machine learning is performed to identify important features in the datasets, and those features are, in some way, included in the resultant mechanistic models. In this way, the current gap between data-driven and mechanistic modeling could be bridged. This is alluded to, but not made explicit, in the section entitled "Using descriptive models and linking them to predictive models" and is a component of Figure 3. While I realize that this white paper encapsulates and presents the

current consensus within the ELIXIR community, a more explicit, high-level overview of how data-driven and mechanistic modeling could interface (e.g. as noted in Table 5 under “short-term goals”) would improve the manuscript. Perhaps the authors need to clarify exactly what the term “smart big data” means, because it is possible that this term refers to what I have mentioned above.

2. Executive Summary, definition of Systems Biology: “thousands” should probably be changed to “myriad” or some other more general term since biological interactions could add up to the millions or more. For example, while there are thousands of genes, there are millions of single-nucleotide polymorphisms that could impact biological systems.
3. Executive Summary: I think the authors presuppose that readers are already familiar with ELIXIR and its myriad activities. However, many readers will not be aware of this group. Thus, there should be a straightforward introduction to ELIXIR and its core mission. In a related issue, the authors should make sure to state that their focus is predominantly on developments happening in Europe since that is where this consortium is located, though the white paper does cite resources that were developed in the U.S.
4. Executive Summary, Tables 1/2: The authors are to be lauded for their extensive listing of prior/existing Systems Biology initiatives. However, Avicenna Alliance is not mentioned (though it is closely related to VPH), and this group has made major strides in driving the adoption of computational modeling and in silico clinical trials.
5. Introduction: in the section entitled “Systems biology underpinning systems medicine,” the reader may be well-served by mentioning Translational Systems Biology and relevant publications from that related field. This could be included also in the section on Quantitative and Systems Pharmacology. A discussion of both of these approaches is of relevance to the white paper’s focus on in silico clinical trials and digital twins. In this regard, the lack of connection between the Systems Medicine and QSP sections is a bit of missed opportunity since at the end of the day drugs are being developed to treat patients, and the disease models that serve as the basis of digital twins are offshoots of the same models used to develop the drugs in the first place (e.g. the goals and objectives listed under Table 4).
6. Minor:
  1. Introduction: “...the study objects of systems biology”; I assume the authors mean “objectives”?

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Partly

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Partly

**Competing Interests:** No competing interests were disclosed.**Reviewer Expertise:** Systems biology; inflammation; computational biology; immunology**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response ( ) 17 Apr 2024

**John Hancock**

1. We thank the reviewer for summarising the perspective on mechanistic models and AI. Mechanistic models are indeed a major goal for the field. For many years, perhaps decades, they have provided a clear-box approach to modelling, allowing the researchers to be in control of the abstractions, and facilitating a transparent mapping to biological processes. AI models, on the other hand, are prone to create difficulties in understanding their inner workings, and much more so in the large language models developed in recent years. As suggested, interfacing the two approaches could be a way to bridge the gap between the explainable and unexplainable modelling approaches. As an infrastructure Community, we will closely follow the latest research developments in this regard. The manuscript has been adjusted by replacing the term “big/smart data” with “big data and AI” (lines 171 and 575) and to include the above expansion on the interfacing between modelling approaches (lines 1220-1229).

2. The suggestion has been applied (line 50).

3. We have added a brief description of ELIXIR as a European infrastructure to this section (lines 63-66)

4. Avicenna Alliance has been added to table 2 (line 1546).

5. We have tried to address this point in the text, especially in the section discussing Industrial Embedding and PBPK models

6. The authors did indeed intend to discuss “study objects”, that is the objects studied by systems biology (line 244)

**Competing Interests:** No competing interests were disclosed.

Reviewer Response 20 Jun 2024

**Yoram Vodovotz**

I thank the authors for addressing my core concerns.

**Competing Interests:** Co-founder, Immunetrics, Inc.; Chief Scientific Advisor, AnunaAI

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**