ISBE Infrastructure for Systems Biology Europe

WP2

Data and Model management

ISBE WP2 report

Deliverable No: D2.4

Implementing Recommendations for ISBE Asset Stewardship

Data and Model Stewardship for the ISBE Infrastructure.

July 2015

Carole Goble, UNIMAN
Katherine Wolstencroft, UNIMAN/UL
Natalie Stanford, UNIMAN
Jacky Snoep, UNIMAN/Stellenbosch
Renate Kania, HITS
Martin Golebiewski, HITS
Wolfgang Mueller, HITS
Sarah Butcher, IC
Nick Juty, EBI
Henning Hermjakob, EBI
Nicholas Le Novere, Babraham/EBI

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

SEVENTH FRAMEWORK PROGRAMME

Project ref. no.	INFRA-2012-2.2.4: 312455
Project title	ISBE – Infrastructure for Systems Biology Europe
Nature of Deliverable	R= Report
Contractual date of delivery	Month 35
Actual date of delivery	Month 35
Deliverable number	D2.4
Deliverable title	Implementing Recommendations for ISBE Asset Stewardship
Dissemination Level	PU
WP relevant to deliverable	WP2
Lead Participant	UNIMAN
Author(s)	Carole Goble (UNIMAN) Katherine Wolstencroft, (UNIMAN/UL) Natalie Stanford (UNIMAN) Jacky Snoep, (UNIMAN/Stellenbosch), Renate Kania (HITS) Martin Golebiewski (HITS) Wolfgang Mueller (HITS), Sarah Butcher (IC), Nicholas Le Novere (Babraham/EBI).
Project coordinator	Richard Kitney
EC Project Officer	Keji-Alex Adunmo

Introduction

The asset stewardship recommendations made for ISBE in D2.3 serve to ensure that ISBE assets can be integrated and exchanged between ISBE centres. They also outline activities that ISBE should undertake in order to increase the compatibility of ISBE services with the broader systems biology community, promoting standardisation and supporting synergies with other ESFRIs.

Stakeholders

Our stakeholder analysis is organised into six user categories and nine sector categories operating across three levels: institutional, national and international.

Users are identified as: researchers that are systems biology specialists or general bioscientists; application users from clinical/healthcare and/or commercial; and end user policy makers and citizens.

Sector stakeholders are identified as: funding agencies; vendors/commercial interests; employer/host institutions; scientific societies/community groups/networks; standards bodies/groups; research infrastructures; training initiatives; resource/service providers; and public and commercial scholarly communication bodies (notably publishers and libraries).

Asset Management Capability Framework

The Asset Management Capability Framework is a tool to: profile the current readiness / capability of ISBE; highlight priority areas for change and investment; and develop roadmaps. This Framework will serve as a systematic device for planning the Interim Phase of ISBE.

We extended an established framework, including the incorporation of the influence of users/sector stakeholders and their case studies and recognition of the Systems Biology method and the related stewardship lifecycle of Systems biology assets. For stewardship to be effective we identified technical, social, cultural and environment aspects of its implementation that must be well managed.

Technical aspects include: how data, models and SOPs should be managed and exchanged within ISBE, and between ISBE and external resources; which formats, identifiers, standards and ontologies should be used, created and maintained for ISBE, and pathways to their adoption; and how interoperability between data and model resources many be achieved.

Social aspects include: how can compliance to the standards recommended by ISBE be encouraged or mandated; how can annotation and standardisation be made more straightforward and rewarding, and less time consuming, for scientists; how data, model and SOP planning and management can become embedded in Systems Biology practice and

publishing; and how practices can lead to greater collaboration and openness for the research results of publicly funded research.

Cultural aspects include: how existing and new Systems Biologists can be educated with respect to data, model and SOP stewardship; how other stakeholders such as funders, librarians and publishers should be engaged with the importance of data and model management; how to drive change in the recognition of data, models and SOPs as first class, citable and creditable research outcomes; and how to establish career paths for data and model stewards.

Environment aspects include: how the community should select specific public resources and services to be ingested and sustained in the ISBE infrastructure; how to establish partnerships with other RIs such as ELIXIR; how to develop and implement business models for resources and services; and how to develop policies, and responses to ethical, legal, and commercial concerns.

Recommendations from D2.3

1. **FAIR publishing**. All assets generated by EU researchers and projects and stewarded by ISBE recognised resources should be published FAIR - Findable, Accessible, Interoperable, Reusable/Reproducible. Data and models in the academic domain should be shared with the community as soon as possible. Linking individual researchers to their data and models, and providing persistent links to them, however, should enable scientists to gain credit for reuse of their datasets and models, encouraging an open, sharing culture. ISBE should establish FAIR guiding principles for the publishing of research data that should inform all decisions relating to ISBE's management of research data, models and SOPs. Implementation of the principles is the responsibility of all ISBE nSBCs and the cSBC.

2. **Stewardship in the service of predictive modelling**. Stewardship in systems biology requires all related research assets from a systems biology investigation (models, data, SOPs, samples, maps etc) to be aggregated and interlinked. The focus of ISBE is stewardship in the service of models. That is: model stewardship and simulation services; and data/SOP stewardship for collecting data for constructing and validating models and supporting the data results of predictive models. Legacy public archives may be transformed when possible, and dedicated archives constructed to suitably support quantitative biology. Stewardship practices focused on Systems Biology distinguishes ISBE from ELIXIR.

3. **Sustained, dedicated, public archives and repositories**. The modelling of biological systems based on integration of diverse data sets will rely on datasets being available that are suitable for integration. ISBE is responsible for the long term stewardship of strategically important research assets (data, SOPs, tools, maps and models). The research community's outcomes should, first and foremost, be placed in these sustained, dedicated, public repositories and catalogued by these sustained, public, dedicated registries. Data, models and SOPs generated by projects supported by the ISBE infrastructure/training, or publicly available and compliant with ISBE best-practice

recommendations, should also be catalogued, archived and published in compliance with ISBE's FAIR principles.

ISBE should seek to (i) *identify and sustain key established dedicated public repositories/registries* for the benefit of the community, seeking partnerships with other RI where appropriate, and *develop and sustain key missing dedicated public resources* where identified by users and stakeholders; (ii) establish, curate and sustain *a Systems Biology Tools and Resources Registry*, leveraging and aggregating pre-existing resources, in particular ELIXIR's registry; and (iii) monitor the usage, performance and quality of such resources against to be established metrics. Open and transparent processes and achievable and appropriate criteria need to be established. Selected, key, investigator-lead resources or assets will need to be migrated to become backed sustainably by nSBCs.

Compliance to the ISBE FAIR Principles will be a criteria for acceptance of a resource into the FAIR Infrastructure.

4. **A sustained Systems Biology Commons.** The modelling of biological systems based on *integration and cross linking* of diverse data sets. A Commons is a community controlled environment that brings together distributed research assets and distributed users/contributors. Systems Biology investigations are inherently integrated, cross-asset, cross-archive, cross-researcher (experimentalist, modeller), and often cross-lab. A Commons enables researchers to catalogue, pool (exchange, share, publish), cross-link, access, and analyse their own and public assets, using their own and third party tools. Benefits include: (i) aggregating repositories with contextual metadata; (ii) overcoming the fragmentation of the asset-specific repositories (iii) hosting experiment-specific, "boutique" datasets; (iv) retaining, and preserving assets of independent researchers; (v) driving compliance of standardisation practices; (vi) making project outcomes available for stakeholders and tracking their usage; and (vii) bridging research practice and research publishing.

   The key part of a Commons is the ***pan-asset, pan-repository*** catalogue that indexes and links the assets associated with a published investigation, which may well be stored in different repositories hosted by different organisations. Thus Commons are gateways to public archives to deposit outcomes, as well as access content, while retaining the connections to the investigation context and cross-links to related assets (models with data, data with SOPs etc). Commons use is governed by established regulations and policies for behaviours, for deposition and metadata standardisation, FAIR use, FAIR reuse and FAIR sharing with appropriate security, privacy and access controls regulated against a minimum set of community-accepted rules.

   ISBE should seek to (i) establish an **EU-wide Systems Biology Commons** that retains and catalogues the assets of Systems Biology projects in Europe; and (ii) monitor the usage, performance and quality of the Commons against to be established metrics.

   Compliance to the ISBE FAIR Principles will be a criteria for acceptance of a resource into the FAIR Infrastructure.

5. **Sustained stewardship services and technical services**. ISBE should provide a *set of services* to support both ISBE stewards and researchers to curate, archive and share research assets, including: data and model management planning; pathways for public publishing; and technical compliance validation of data and models against standards, policies and practices; authenticated and authorised and identified access; and data transfer. ISBE does not govern the science or scientific methodology that at undertaken using its infrastructure. That is the purview of peer review.

   The framework of services and resources must not dictate a single platform or a tightly integrated data infrastructure. Systems Biology is integrative by nature, drawing upon the ecosystem of data and model resources (legacy, emerging and provided by pre-existing or forthcoming Research Infrastructure (RIs)). In order to ensure sustainability, ISBE infrastructure, interoperability and compliance policies must be the minimal required for functionality, and devised in partnership with those RIs. The conventions for data and model services interoperability should be based on minimal "hourglass" approach, a specification of lightweight interfaces, standard protocols and standard formats.

6. **Support projects and researchers with asset management platforms.** For data, models and SOPs generated by projects supported by the ISBE infrastructure/training, ISBE should identify and support platforms that enable researchers, projects, institutions to manage their assets. Platform should to "RARE" research practices (Robust, Accountable, Intelligible, Reproducible) with workflows for "FAIR" Publishing using ISBE public resources.

7. **Support for commercially sensitive and personally sensitive data.** ISBE will support life sciences research, health research and commercial collaborations in these areas. Patient data for clinical or biomedical applications requires secure and sensitive handling. A mixture of open and commercially sensitive data/models and open and commercial services should be catered for. Commercial services may form part of the ISBE data and model framework: from the publishers and publishing services through to commercial data and knowledge bases and modelling tools and underpinning commercial cloud hosting. We anticipate potential financing as a public private partnership and the implications this may have on data visibility – its accessibility and accessibility. The operating conditions that ISMB should support private and proprietary data needs to be defined.

   Clear policies, standard operating procedures and supporting infrastructure are required to ensure that private health care information or commercial assets are kept with secure and restricted access (the "A" in FAIR stands for Accessible, not open). In some cases an Information Security Management System defined by Policies and Standard Operating procedures certified to ISO27001 will be required.

8. **Development and adoption of common practices and standards.** The ISBE data and model management framework focuses on conventions that enable data interoperability and stewardship and compliance against data and metadata standards, policies and practices. We must define, develop and adjust criteria and standards that must be met by data, maps, tools and models; support the accuracy, reliability and quality of data,

models, tools and maps.; and make the re-use of data sets, models, SOPs etc. possible in future projects.

The conventions for data and model metadata descriptions must be founded on community standards for identifiers, formats, checklists and vocabularies, developed through community engagement, to make data, models, and tools re-usable. A knowledge hub and training activities will be needed to disseminate these practices and standards, and technical development to implement them into tools.

ISBE must be an active and engaged advocate for the development and adoption of standards. We recommend a concerted action of the European systems biology infrastructure with the respective ISO committees as ISO/TC 276 with the objective of defining a horizontal framework standard for the data and model patchwork in the field. Such a strategic alliance has to include the corresponding domain-specific grassroots standardization initiatives like COMBINE, FGED, PSI, MSI and others; wider standardisation bodies such as the Research Data Alliance and the Global Alliance; and work with journals and funders to establish practical best-practice usage of community standards for publication.

ISBE should set in motion measures (training, services, and infrastructure) for the making and habitual use of standards for the research assets of Systems Biology, notably data, SOPs and models, and how these are related to each other and to investigations.

9. **Build stewardship capacity and capability.** When ISBE acts as a broker to bring researchers who generate data into contact with researchers who require data, standards-based and model-compliant data generation must be ensured along with data management planning. We will need *stewarding services* support to store and explore the links between data, models, protocols and results from ISBE investigations, showing the systems level details of the experiments, and to understand how separate datasets (e.g. genomics, transcriptomics and proteomics) can be interpreted together, or how they are used for construction or validation of the model, to enable a systems level understanding. *Training and education* is required across the different expertise of ISBE users, stakeholders and stewards, including members of nSBCs, ranging from in-house training to curriculum development for higher education institutions. ISBE must partner with international training initiatives such as GOBLET and Software Carpentry, and national initiatives such as SysMIC.

10. **The recognition of all assets and all stewardship activities**. Data and models must be citable and cited, with credit given to their authors and stewards, and commoditised so that they can be re-used modularly. Stewardship needs to be recognised and rewarded as a first class and habitual activity. Assets need to be recognised and rewarded as first class research outcomes with appropriate credit metrics. Dedicated stewards and Research Data Engineers, and those Research Software Engineers producing stewardship tools, should be recognised with established and rewarding career paths. ISBE should establish partnerships with stakeholders: institutions, funders, publishers, journal editorial boards, learned societies, pressure groups and networks (such as Force11) to advocate for the recognition of all assets and the recognition of the skills of asset

stewards, develop supporting infrastructure and establish practical best-practice usage of community standards for creditable publication.

11. **Sustained funding and business models**. ISBE should seek avenues for sustainable funding for asset stewardship and public resources, and develop a portfolio of business models. transformative 5% tax. example: the Netherlands and NWO, DTL. Funding agencies and grant allocation could also allow funds to go directly to curation and stewardship activities, thereby facilitating the longevity of data and data accessibility in the longer term.

12. **Develop Synergies with other RIs and other partners**. Synergies should be identified across the various RIs in a systematic way; repositories that can provide data of use in ISBE should formalise agreements for data sharing and access, SOPs should be established for curation and annotation of datasets and models, with a clear policy established for responsibility for the data, in terms of where and how it is stored, and on the means it should be accessed by ISBE, and by the systems biologist.

13. **EU, national and community regulations and compliance vigilance**. ISBE must maintain awareness and vigilance with respect to EU and national regulations and compliance mandates. Regulation in ISBE is challenging as national and European regulations are at play. The most notable regulation is European Commission's European Data Protection Regulation, which replaces the previous Data Protection Directive. The aim of the new European Data Protection Regulation is to harmonise the current data protection laws in place across the EU member states. As a "regulation it is directly applicable to all EU member states without a need for national implementing legislation. The regulation on the movement and processing of personal data is much tougher than previously. Other regulations are national or community standards for, Information Security Management Systems (ISO27001).

Proposed ISBE Infrastructure

ISBE will be a distributed infrastructure that provides services and resources to support world-class systems biology research. It will cover 5 strategic areas of services and resources required for producing successful systems biology research (Figure a):

- Training and education
- Modelling
- Community activities
- Standards
- Data, model and SOP management and stewardship

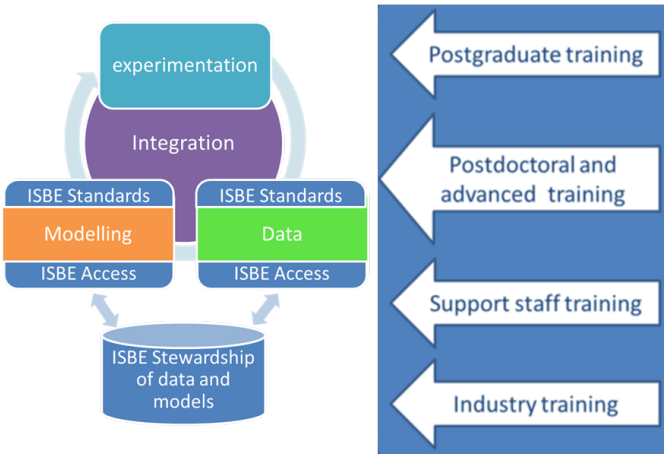

Figure a: Strategic areas of service and resources required for systems biology research.

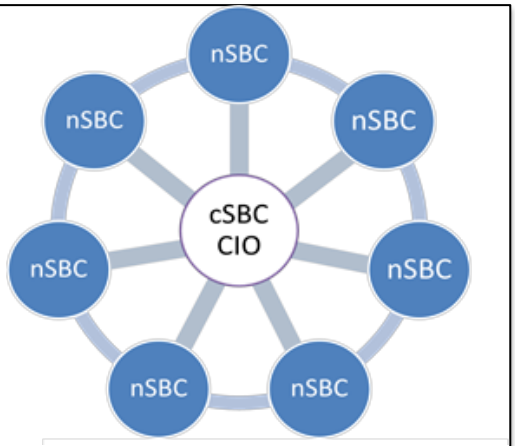

Figure b: ISBE hub and spoke model.

ISBE will be structured according to a hub and spoke model (Figure b). ISBE is represented on the international level by a single central Systems Biology Centre (cSBC) which is responsible for operational strategy (i.e. planning and reviewing present and future services). The cSBC will be connected to the national Systems Biology Centres (nSBCs), which will be responsible for ensuring delivery of services. At the national level the organisation is somewhat flexible depending on how an individual country chooses to implement its nSBCs; a country may consider a central institute to act as a co-ordinator with other institutes as partners, or may choose to implement a centralised body which co-ordinates all institutes as members. Each of the nSBCs will contain component services and resources from any to all of the 5 strategic areas recognised as ISBE service priorities. The portfolio of nSBCs, centrally coordinated by the cSBC, will cover all strategic areas of services and resources.

The ISBE infrastructure is a complex network of physical and virtual resources designed to support a model-centric and data-centric approach to Life Sciences. Tilsley and Coveney present an infrastructure viewpoint that refers to: (i) data repositories, catalogues and libraries, and data services such as LIMS and citation tracking; (ii) software and algorithms such as modelling tools, and data/software management systems; (ii) underpinning "consumables" such as storage, compute and networks; and (iv) cross-cutting services such as access authorisation and authentication. Infrastructure also includes (v) people and their expertise: Systems Biologists who generate and use the data and models, data and model curators, systems administrators and so on.

The distributed, interconnected infrastructure envisaged by ISBE depends on the adoption of best practices, standards, technical infrastructure, and capacity for the management and distribution of data and models, and the management and sustainability of data and model management software. It is easy to overlook the fact that both data and models are entirely dependent on the software used to manage, access, search, run, exchange, regulate, validate them. In 2014 the UK House of Lords[i] went as far as to state that in fact infrastructure was software and that storage/compute facilities were consumables, a sentiment echoed in funding council's roadmaps[ii]. The sustainability and maintenance of data and model management software is thus crucial to ISBE infrastructure.

Provisioning a common framework for the nSBCs and users will enable data and models arising from the ISBE infrastructure to be retained and managed. Adopting a common framework and standards will enable the FAIR exchange of data, models and SOPs between

nSBCs and will allow scientists to (a) support the reproducibility of results; and (b) discover and reuse these data and models for their own research. Adopting standards that are already in use in the wider Life Science community will additionally ensure easier exchange with external resources, such as those from ELIXIR, Euro-Bioimaging and BBMRI.

ISBE aims to provide asset services and resources at two levels:

1. **Specialist public archives** managed for the international community by national or pan-national providers that are: (a) asset-specific datasets such as BioModels, SABIO-RK, Metabolights, BRENDA, JWS Online, COMBINEArchiveWeb etc; (b) public tools such as COPASI for modelling and DMPOnline for data management planning ; (c) catalogues of datasets and tools such as res3data.org, and metadata standards such as Biosharing.org, to support, respectively, the Findability and Interoperability/Reusability of FAIR research outcomes. Providers may be aligned with nSBCs and those nSBCs will contribute those resources/services to the ISBE Infrastructure. Alternatively, they may be part of another RI (e.g. ELIXIR) and their provision to the ISBE infrastructure contributed through MoUs. ISBE will also take advantage of, and partner with, general repository providers such as figshare and data infrastructure providers such as Dropbox.

2. **Project outcomes** with locally deployable platforms and centralised resources to: support inherently integrated, cross-asset, cross-archive Systems Biology investigations; provide a unified **Sys Bio Commons** to the outcomes of European projects (as identified by our Industry Survey); and to locally and/or centrally directly support asset management in the field in research projects, with a pathway for public deposition in the public archives and/or in publisher repositories/companion sites. Examples include the FAIRDOM Initiative's SEEK platform and FAIRDOM Hub[iii].

ISBE Research Infrastructure will be made up of distributed resources and services. Distributed nSBCs will provision a single point of access for data by users and sector stakeholders. The nSBCs implementing this ISBE Infrastructure are expected to: manage public resources; offer a **unified view** over resources generated and used, in the context of the experiments that produced them; and support the stewardship of research assets arising from Systems Biology experiments executed by users of the infrastructure.

The FAIRDOM Project

The FAIRDOM Project[iii] is a 5 year €2.7million programme supported by four European Research Councils (BMBF, NWO, BBSRC and SystemX). It commenced in Oct 2014. Building on previous extensive investments since 2008 in Data and Model Management for Systems Biology, it aims to make the assets of European Systems Biology projects - Data, Operations, Models - Findable, Accessible, Interoperable and Reusable.

The FAIRDOM has four major pillars:

1. **FAIRDOM Software Suite**: An open suite of software for establishing an Asset Management FAIRDOM Platform for Sys Bio projects. The prime software is a cataloguing, metadata-rich online front end (SEEK), supported by data management

backend for local project data management (OpenBIS), a range of community public archives, and a Core Pool of pluggable tools.

2. **FAIRDOMHub:** A Systems Biology Commons for projects, implemented by the Software Suite;
3. **FAIRDOM Facilities**: A network of facilities in national centres for supporting Sys Bio asset stewardship, training and support and contributing to FAIRDOM sustainability;
4. **FAIRDOM Community**: A range of activities to build capacity and capability for asset stewardship, contribute to standards, support tool/resource developers and administrators, and champion the FAIRDOM project and its systems.

FAIRDOM is seeded by the ERANet ERASysAPP and is the companion data and model management service for the projects of that programme, as it was for the SysMO EraNET and the German Virtual Liver Network. It also aims to become a go-to Commons for publishers, funders, investigators and other stakeholders, as well as providing an open platform for independent local project management.

FAIRDOM forms part of the ISBE-Light Infrastructure for the Systems Biology Commons, standards and stewarding services.

[i] http://www.publications.parliament.uk/pa/ld201314/ldselect/ldsctech/76/76.pdf
[ii] http://www.epsrc.ac.uk/SiteCollectionDocuments/ourportfolio/EInfrastructureRoadmap.pdf
[iii] http://www.fair-dom.org, http://www.seek4science.org; http://www.fairdomhub.org

1 FAIR Publishing

> "***FAIR publishing***. *All assets generated by EU researchers and projects and stewarded by ISBE recognised resources should be published FAIR - Findable, Accessible, Interoperable, Reusable/Reproducible. Data and models in the academic domain should be shared with the community as soon as possible. Linking individual researchers to their data and models, and providing persistent links to them, however, should enable scientists to gain credit for reuse of their datasets and models, encouraging an open, sharing culture. ISBE should establish FAIR guiding principles for the publishing of research data that should inform all decisions relating to ISBE's management of research data, models and SOPs. Implementation of the principles is the responsibility of all ISBE nSBCs and the cSBC.* "

1.1 Status

We have seen numerous studies which demonstrate the knowledge loss from the scientific community when research outcomes are not reproducible, with some studies showing as much as $28 billion a year is spent on irreproducible research[4]. Reproducible research requires a shift in focus from scientific output being confined to publications, and a move towards the publication of robust and findable research assets themselves[5]. To this end the open movement FAIRport[6] was established with a vision to join and support communities to enable FAIR data publishing. FAIRport comprises of a wide range of communities including FORCE11[7], ELIXIR[8], BD2K[9], RDA[10], ODEX4all[11], ENPADASI[12], BBMRI-NL[13], and FAIRDOM[14].

FAIR research assets are realised by communication, reconciliation, adoption, and endorsement of community standards for provenance, versioning, identity, citation, description, and dependency. In the systems biology community there are many grass roots communities that work towards this goal including COMBINE[15], PSI[16], FGED[17]. There are also pan-national general initiatives such as RDA[18]. Besides the grassroots community efforts, there are also some established standardization bodies such as W3C[19] and ISO[20] at international or CEN[21] (European Committee for Standardization) and CENELEC[22] (European Committee for Electrotechnical Standardization) at European level, and national bodies (e.g. German DIN[23], British BSI[24], NEN[25] in the NL, and others) that closely collaborated with the European and international standardization bodies. The availability of community-developed standard formats for data and models (e.g.SBML, SBGN, CellML , PSI-MI or MAGE-TAB), controlled vocabularies and domain ontologies for the used terminology (e.g. Gene Ontology GO, Systems Biology Ontology SBO,), and minimum information models for the scope and extent of the metadata, e.g. Minimum Information About a Microarray Experiment (MIAME), Minimum Information About a Proteomics Experiment (MIAPE), Minimal Information Required In the Annotation of biochemical Models (MIRIAM), Minimum

Information About a Simulation Experiment (MIASE), as well as tools and software (e.g. Copasi, JWSonline, CellDesigner, CellML) mean that for systems biology FAIR research assets can be realistically implemented for many data and model types. However, there are still significant barriers with uptake including lack of knowledge about availability and implementation of standards, and which tools, software are databases that should be used to support the implementation. They are however rarely mandated by funding councils and journals currently.

1.2 Interim phase implementation (up to 3 years)

FAIRDOM is a supporting project in the ISBE Light phase, which provides scientists with access to a commons environment that complies with the FAIR publishing guidelines, FAIRDOMHub. A pilot project forms part of the FAIRDOM project, which sees the EraSysAPP EraNET projects supported with a data and model management programme that provides access to pre-planning consultancy regarding data and model management requirements for grants; curation and stewardship services throughout the projects; and a ten-year research asset stewardship agreement.

Further to the development of FAIRDOMHub, and the pilot project, resources will be made available which aid with selecting, understanding, and implementing FAIR compliant research asset management stewardship in the broader community. The broader community will also be supported on a case-by-case basis during the ISBE Light phase, whereby data and model management needs will be discussed and assessed, and suitable plans and costing will be established for individual grants.

1.3 Legal phase implementation

Once ISBE is established as a legal entity, all ISBE nSBCs and the cSBC will be responsible for adopting and enforcing FAIR Data principles for all research assets they steward. In order to facilitate this, a consistent set of curation and stewardship tools and standards, and best practice implementation guidelines will be drawn up and mandated for compliance in each centre. These suites of tools, standards, and guidelines will take into account current community best practice, and sufficient implementation guidelines and training will be provided to ensure the correct implementation.

ISBE will establish a full set guidelines for ISBE for managing the life-cycle - from project planning, to experimentation, reuse and archival/disposal – of research assets available in ISBE. These will be closely aligned with the relevant compliance required by funding councils, and journals within the ISBE nation state, and more broadly within the community. The resources supporting research asset availability will be made sustainable throughout the life-cycle using established business models developed from the experience gained in the Interim Phase.

For the broader community, ISBE best practice will be available as guidelines and training material. However, we will also offer searchable databases for identifying suitable curation and stewardship tools and standards, and community best practice and training facilities. These will allow coverage of much broader research asset management resources, providing flexibility in how the community at large can achieve FAIR research assets.

[4] http://news.sciencemag.org/biology/2015/06/study-claims-28-billion-year-spent-irreproducible-biomedical-research?utm_campaign=email-news-latest&utm_src=email
[5] https://www.force11.org/group/joint-declaration-data-citation-principles-final
[6] http://datafairport.org/
[7] https://www.force11.org/
[8] http://www.elixir-europe.org/
[9] https://datascience.nih.gov/bd2k#sthash.GADVGHZP.dpbs
[10] http://www.dtls.nl/parties-signed-odex4all-project-ready-go/
[11] http://www.dtls.nl/parties-signed-odex4all-project-ready-go/
[12] http://www.healthydietforhealthylife.eu/index.php/joint-actions/enpadasi-partners
[13] http://www.bbmri.nl/en-gb/home
[14] http://fair-dom.org/
[15] http://co.mbine.org/
[16] https://www.systemsbiology.org/hupo-proteomics-standards-initiative-mass-spectrometry-controlled-vocabulary
[17] http://fged.org/
[18] https://rd-alliance.org/
[19] http://www.w3.org/
[20] http://www.iso.org/iso/home.html
[21] https://www.cen.eu/
[22] http://www.cenelec.eu/
[23] http://www.din.de/
[24] http://www.bsigroup.com
[25] https://www.nen.nl/Home-EN.htm

2 Stewardship in the service of predictive modelling

> *"Stewardship in systems biology requires all related research assets from a systems biology investigation (models, data, SOPs, samples, maps etc) to be aggregated and interlinked. The focus of ISBE is stewardship in the service of models. That is: model stewardship and simulation services; and data/SOP stewardship for collecting data for constructing and validating models and supporting the data results of predictive models. Legacy public archives may be transformed when possible, and dedicated archives constructed to suitably support quantitative biology. Stewardship practices focused on Systems Biology distinguishes ISBE from ELIXIR."*

2.1 Status

The integration of data into computational modelling for hypothesis generation and testing is a core component of systems biology research. In order to generate high quality predictable models researchers need access to heterogeneous semantically related data sets, which are collected under physiologically similar conditions. A large number of databases (e.g ArrayExpress, Metabolights and PRIDE) allow storage of, and access to the expanding variety of data types that are used for integration into models. However the still siloed nature of some research topics, and these databases means that cross identifying semantically related datasets is either difficult or impossible.

There has been some advancement towards improving semantic relationships between research assets within the field, by the construction and dissemination of Commons resources, such as SEEK[26]. Commons resources allow the storage and access of data in semantically linked interfaces, where interrelated datasets and models can be identified. These resources rely heavily on the inclusion of suitable metadata, which provides contextual information, and helps users decide the parameters in which the research asset can be validly used. These resources also offer linking of research assets stored in external silo repositories, which allows users a flexibility in the search context based on both data type (e.g. RNA data with RNA data) and functionality type (e.g. cell, tissue, organ), which provides greater utility for silo fields, and integrated research fields.

In order to understand the contextual validity of a given model for a given *in silico* experiment, the provenance of the data used to construct (and validate) the model must be available. This is an area that is currently underdeveloped, with standards largely unable to cope well at describing parameter provenance. Curated databases such as SABIO-RK[27], which provide access to kinetic parameters, and their experimental origins, are helping to improve this. Ideally, provenance information should be part of the required metadata and captured in the model in the same way as biological identifiers and reactions. The Research Object model (RO[28]) and the COMBINE archive with its underlying Open Modeling EXchange format (OMEX)[29] also help to address this problem as they aggregate related research assets (publications, data, models, workflows, simulation settings, etc) into single collections, with

defined relationships, so that they retain their context. Similarly the ISA[30] framework, which is a sub-type of research object, allows the aggregation and semantic contextualisation of experiments with both laboratory and computational components.

2.2 Interim phase implementation (up to 3 years)

ISBE, through the FAIRDOM project will be supporting the development and dissemination of the SEEK Software Platform. SEEK allows the deposition of research assets, or the cataloguing of research assets from silo databases (BioModels, PRIDE etc) to be represented in a semantically interlinked interfaces. Research Objects, which are packages of semantically linked research assets, will soon be integrated into SEEK, so that research project results can be packaged up into self-contained units for dissemination.

The description of biological samples and their treatments is an important aspect of experimental metadata. A number of initiatives already exist in this area, including the EBI BioSamples database and the ISA biosamples model. FAIRDOM will use these initiatives to improve the characterisation, and standard formatting and requirements of Samples (including experimental and computational) so that they can be better formatted and described within repositories and commons.

2.3 Legal phase implementation

ISBE will ensure that interoperability standards of modelling, and associated services for archiving, integration, and re-use keep pace with the data developments, which will primarily be driven by ELIXIR. In addition to this, ISBE must ensure that any additional requirements for standardisation or metadata descriptions that support the integration of data into models are developed alongside ELIXIR.

ISBE also will drive the adoption of standards for research asset aggregation (e.g. Research Object Model and ISA – Investigations, Studies, Assays). All nSBCs will be responsible for adopting these standards in the presentation of their research assets, however stewardship specific nSBCs will drive the coordination of community activities which push forward the development of these aggregation standards so that they remain relevant for use. To ensure the uptake with the wider public ISBE will produce resources to aid uninitiated users to identify and use these aggregation standards within their work.

ISBE will identify and support (national) resources for quantitative biology, in partnership with other research infrastructures where appropriate. Part of this work will be ensuring that resources within the community are suitable for supporting the needs of quantitiative biology, and where this is not the case, work with the community to establish resources that are.

ISBE will work with the community (e.g. COMBINE, expert nSBCs, researchers) in order to establish guidelines of best practice for modelling implementation. These will include appropriate standards for formatting and annotating models to ensure the origin of

parameters within the model can be identified. It will also include best practice data collection for modelling purposes to ensure that the data is valid for use.

[26] http://fair-dom.org/seek
[27] http://sabio.villa-bosch.de/
[28] http://www.researchobject.org/
[29] *BMC Bioinformatics* 2014, **15**:369
[30] http://isatab.sourceforge.net/format.html

3 Sustained, dedicated, public archives and repositories

"The modelling of biological systems based on integration of diverse data sets will rely on datasets being available that are suitable for integration. ISBE is responsible for the long term stewardship of strategically important research assets (data, SOPs, tools, maps and models). The research community's outcomes should, first and foremost, be placed in these sustained, dedicated, public repositories and catalogued by these sustained, public, dedicated registries. Data, models and SOPs generated by projects supported by the ISBE infrastructure/training, or publicly available and compliant with ISBE best-practice recommendations, should also be catalogued, archived and published in compliance with ISBE's FAIR principles.

ISBE should seek to (i) identify and sustain key established dedicated public repositories/registries for the benefit of the community, seeking partnerships with other RI where appropriate, and develop and sustain key missing dedicated public resources where identified by users and stakeholders; (ii) establish, curate and sustain a Systems Biology Tools and Resources Registry, leveraging and aggregating pre-existing resources, in particular ELIXIR's registry; and (iii) monitor the usage, performance and quality of such resources against to be established metrics. Open and transparent processes and achievable and appropriate criteria need to be established. Selected, key, investigator-lead resources or assets will need to be migrated to become backed sustainably by nSBCs. Compliance to the ISBE FAIR Principles will be a criteria for

3.1 Status

Public resources include specialist public archives, tools, and software that aid with many aspects of systems biology research asset handling (storage, formatting, annotation, analysis, simulation, visualisation etc.).

A number of repositories cover experimental reporting, which specialise in the collection of a single data type for example: proteomics (PRIDE[31]), metabolomics (Metabolights[32]), models (BioModels[33], JWSOnline[34]), Samples (BioSamples[35]). Others are curated knowledge bases, where domain specific information is mined from papers, curated into a specific format with a high level of associated metadata, and stored for community use (e.g. SABIO-RK[36], ChEBI[37], KEGG[38]). Experimental reporting databases offer guidelines for possible formats and annotations for data, but often do not make these explicit requirements. Maintainers of the database are not involved in assessing the reproducibility of the published data, therefore the quality of data within these repositories can suffer. Curated

knowledge bases involve selection and curation of the data such that it can reproduce publication findings, and sufficient annotations are present to allow provenance of the research asset itself. This results in higher quality research assets, but also requires significant investment of resources from the repository maintainer.

Tools and software are generated regularly by individual researchers, whole research groups, and coordinated research communities. The reliability of these tools and software vary greatly: some are generated, produced, and published with no mind for sustainment (e.g. when the tools/software have supported a specific research topic); some are generated to support internal group work only, but may be made available for the general public to use with caveats that general support will not be provided; others are sustained group and community efforts to deliver resources that researchers can rely on, with provisions for training and general use, as well as open community development. The latter of these is obviously more conducive to enhancing the resource availability to the community, and encouraging usage. Tellingly, many of these tools and resources that are developed for adoption by the wider community provide a much greater resource and drive for helping improve standardisation and reproducibility of data and models within the community.

A few specialist resources are supported sustainably at the EBI, but even these depend upon grants for continued development. As in many other places, other resources may have been developed as part of a research projects. Once the projects ended, there were very few opportunities to apply for suitable grants to support continued development, KEGG being a prime example. This can lead to many problems for researchers. They can become heavily reliant on a resource or service that suddenly become unsustainable and closes, leaving a huge gap in the research pipeline, and in previous published work no longer being reproducible. It has also been known for researchers to avoid reliance on specific resources out of fear that this may happen. It is clear that we need to move to a situation where key resources are guaranteed in the long term through sustainable business models that do not impede access to the community at large.

3.2 Interim phase implementation (up to 3 years)

ISBE must identify some of the most valuable resources that facilitate systems biology research. These can be used in the legal phase to prioritise nSBC investment into these key resources to support their use by the whole community. Longer term strategy models for this also need to be developed.

With respect to the 'hub and spoke' model proposed in the Business Plan, one could envisage that the various nSBCs would, as part of their function, contain an ISBE resource that either accepts model data, or else is responsible for model annotation and or curation. Hence there would be defined links between the model archives and repositories that are 'spokes' in the ISBE infrastructure; one nSBC may accept models (for instance contains a model repository), which would instigate a process of annotation by a different nSBC (for instance JWS Online), which feeds a further process of curation (for instance through BioModels curation team).

3.3 Legal phase implementation

Strategies for supporting the key resources identified in the interim phase will be established. ISBE priority resources should involve providers being aligned with nSBCs, and the nSBCs establishing with national funding bodies the case for support, and requirements for sustainability. Where the resource is a key resource shared with other RIs (e.g. ELIXIR), ISBE should establish Service Level Agreements, which identify the commitment of each RI to the resource and an appropriate funding model associated with it. It is key that funding models cover the sustainability of the resources existence, maintenance and service/curation costs.

To support the access and use of these resources, ISBE must establish, with ELIXIR, a key resource registry that contains all supported and sustainable resources. This will allow users to identify the resources which will be available over the long term.

[31] http://www.ebi.ac.uk/pride/archive/
[32] http://www.ebi.ac.uk/metabolights/
[33] http://www.ebi.ac.uk/biomodels-main/
[34] http://jjj.mib.ac.uk/
[35] https://www.ebi.ac.uk/biosamples/
[36] http://sabio.villa-bosch.de/
[37] https://www.ebi.ac.uk/chebi/
[38] http://www.genome.jp/kegg/pathway.html

4 A Sustained Systems Biology Commons

"The modelling of biological systems based on integration and cross linking of diverse data sets. A Commons is a community controlled environment that brings together distributed research assets and distributed users/contributors. Systems Biology investigations are inherently integrated, cross-asset, cross-archive, cross-researcher (experimentalist, modeller), and often cross-lab. A Commons enables researchers to catalogue, pool (exchange, share, publish), cross-link, access, and analyse their own and public assets, using their own and third party tools. Benefits include: (i) aggregating repositories with contextual metadata; (ii) overcoming the fragmentation of the asset-specific repositories (iii) hosting experiment-specific, "boutique" datasets; (iv) retaining, and preserving assets of independent researchers; (v) driving compliance of standardisation practices; (vi) making project outcomes available for stakeholders and tracking their usage; and (vii) bridging research practice and research publishing.

The key part of a Commons is the ***pan-asset, pan-repository*** *catalogue that indexes and links the assets associated with a published investigation, which may well be stored in different repositories hosted by different organisations. Thus Commons are gateways to public archives to deposit outcomes, as well as access content, while retaining the connections to the investigation context and cross-links to related assets (models with data, data with SOPs etc). Commons use is governed by established regulations and policies for behaviours, for deposition and metadata standardisation, FAIR use, FAIR reuse and FAIR sharing with appropriate security, privacy and access controls regulated against a minimum set of community-accepted rules.*

ISBE should seek to (i) establish an ***EU-wide Systems Biology Commons*** *that retains and catalogues the assets of Systems Biology projects in Europe; and (ii) monitor the usage, performance and quality of the Commons against to be established metrics. Compliance to the ISBE FAIR Principles will be a criteria for acceptance of a resource into the FAIR Infrastructure."*

4.1 Status

Benefits include: (i) aggregating repositories with contextual metadata; (ii) overcoming the fragmentation of the asset-specific repositories (iii) hosting experiment-specific, "boutique" datasets; (iv) retaining, and preserving assets of independent researchers; (v) driving compliance of standardisation practices; (vi) making project outcomes available for stakeholders and tracking their usage; and (vii) bridging research practice and research publishing.

The key part of a Commons is the ***pan-asset, pan-repository*** catalogue that indexes and links the assets associated with a published investigation, which may well be stored in different repositories hosted by different organisations. Thus Commons are gateways to public archives to deposit outcomes, as well as access content, while retaining the connections to the investigation context and cross-links to related assets (models with data, data with SOPs etc). Commons use is governed by established regulations and policies for behaviours, for deposition and metadata standardisation, FAIR use, FAIR reuse and FAIR sharing with appropriate security, privacy and access controls regulated against a minimum set of community-accepted rules.

4.2 Interim phase implementation (up to 3 years)

The FAIRDOM project is initiating a programme of sustainable support for the FAIRDOMHub Commons. The front-end cataloguing platform SEEK[39] and the back-end platform OpenBIS form the heart of a suite of software for implementing the Commons. These assist with the collection, harvesting, cataloguing, storage, and sharing of data, models, SOPs to the broader community.

ERANets SysMO and EraSysAPP and national projects in Synthetic Biology (UK) and the Virtual Liver Network (Germany) are being used to pilot the Commons for long-term research asset management. Coupled to this, SysMo[40] (the project which SEEK was generated to support), research assets are being stored long-term within SEEK and are fostering collaboration in follow up work.

Agreements will be drafted to ensure that projects that use FAIRDOMHub can have their research asset storage and availability guaranteed for 10 years after the end of the project (in line with funding council agreement). The project will also develop sustainability funding models for these platforms.

4.3 Legal phase implementation

A Commons resource which aggregates all research assets centrally should also be established and hosted by an nSBC, with a sustainability framework developed for its long term security. The usage, performance and quality of the Commons must be measured against metrics to be established.

A Commons will establish a single "go-to" resource which directs to all EU Systems Biology resources for users – making finding other resources much easier.

[39] http://www.seek4science.org
[40] https://www.sysmo-db.org/

5 Sustained stewardship services and technical services

"ISBE should provide a set of services to support both ISBE stewards and researchers to curate, archive and share research assets, including: data and model management planning; pathways for public publishing; and technical compliance validation of data and models against standards, policies and practices; authenticated and authorised and identified access; and data transfer. ISBE does not govern the science or scientific methodology that at undertaken using its infrastructure. That is the purview of peer review.

The framework of services and resources must not dictate a single platform or a tightly integrated data infrastructure. Systems Biology is integrative by nature, drawing upon the ecosystem of data and model resources (legacy, emerging and provided by pre-existing or forthcoming Research Infrastructure (RIs)). In order to ensure sustainability, ISBE infrastructure, interoperability and compliance policies must be the minimal required for functionality, and devised in partnership with those RIs. The conventions for data and model services interoperability should be based on minimal "hourglass" approach, a specification of lightweight interfaces, standard protocols and standard formats."

5.1 Status

There are a large fraction of tools and capacities for creating services already exist or are being built. Tools supporting the project life cycle (e.g. dmptools[41] for generating data management plans), libraries and tools for visualizing, annotating and transforming data (e.g. libSBML[42], SemanticSBML[43], SBGN tools[44] and libraries). Data and model collections and capacities, like e.g. BioModels, BRENDA[45], SABIO-RK, and Commons like the FAIRDOM hub.

Capacities providing services include (i) supporters (ii) curators, (iii) developers. Community Supporters facilitate the use of the services for the users, they outline possibilities, suggest ways to go for optimizing added value and liaise to curators and developers where needed. Curators curate data and models: They structure and enrich data where needed. Enrichment amounts to annotating data to ontologies, restructuring data and models for meeting best practises, as well as checking consistency. (iii) Developers provide the tools used for the above.

5.2 Interim phase implementation (up to 3 years)

Building on the deliverables of ISBE and the work of standardisation (COMBINE, ISO) and curation organisations (DCC, ISB) we will establish best practises guidelines for curation, and community engagement.

Tools and workflows for the construction and annotation of models should be developed and made available either through individual national Systems Biology Centres (nSBCs), or accessed directly from the central Systems Biology Centre (cSBC).

5.3 Legal phase implementation

ISBE will establish resources that allow the use of repositories, standard formats and annotations, controlled vocabularies, and analysis and simulation tooling, without the use requiring to actively alter formats of models or data to analysis and integrate data, or to construct and simulate models.

ISBE will establish and partner with platforms and resources that support data and model management (from data collection to publication and storage), by automatically formatting and annotating data and models with appropriate information with little input by the researcher. Removing the time and knowledge barriers that are often present with data and model management implementation.

[41] https://dmp.cdlib.org/
[42] http://sbml.org/Software/libSBML
[43] http://semanticsbml.org/semanticSBML/simple/index
[44] http://www.sbgn.org/SBGN_Software
[45] http://www.brenda-enzymes.info/

6 Support Projects and researchers with asset management platforms.

> *"For data, models and SOPs generated by projects supported by the ISBE infrastructure/training, ISBE should identify and support platforms that enable researchers, projects, institutions to manage their assets. Platform should to "RARE" research practices (Robust, Accountable, Intelligible, Reproducible) with workflows for "FAIR" Publishing using ISBE public resources."*

6.1 Status

To encourage the publication of FAIR research assets, research asset management must start inside research groups, before research assets are even close to being ready for publication. This entails adequate and appropriate documentation of procedures at the outset of an investigation.

One of the barriers to the support of such research asset management, and consequently downstream the supporting of FAIR research asset publication, is the lack of platforms and tools to facilitate the formatting, annotation and exchange of research assets within the lab. Subsequently, this means that preparing research assets at the point of publication becomes a large job, which can be difficult and time-consuming. If a researcher has left a project, it is not always easy to decipher what they have produced and how, meaning that data can be lost or wasted.

Research groups tend to have a large number of researchers on short-term contracts. For example a post-doctoral researcher will tend to be employed for around 3 years as part of a larger grant. Due to the time-limited nature of contracts, and also researcher progressing up the research career ladder, and also out of academic science into other career avenues, it can pose problems with the exchange of research assets to the PI when a researcher leaves the group. The cross over usually requires the researcher to explain to all of the procedures, and the research asset to other colleagues who may then pick up or pass on the work to another researcher. There is much scope for important information to be lost in translation. Management platforms such as SEEK, which can be used in-house for unpublished assets have helped in some instances to reduce this researcher turnover cost to groups. However it is not a practice employed in all labs. If researchers had more access to these and could use them effectively it would be beneficial to the entire research pipeline, and prevent the loss of a lot of data.

Many research projects have their own in-house solutions, ranging from shared folders, to content management systems and LIMS. Other research projects have nothing implemented. Some universities and institutes provide and recommend specific resources. These however do not encourage the RARE qualities that eventually make a FAIR research asset that can be submitted to an external database. For instance, many internal archives

allow the upload and storage of data in any format, without annotations. This makes retrieval, search and comparison of the data extremely difficult.

6.2 Interim phase implementation (up to 3 years)

In order to improve the ease with which research assets can be stewarded within labs, we are setting up a pipeline of support as part of the FAIRDOM project. Here OpenBIS[46] and SEEK will be integrated together with other useful resources (including JWS online, Sycamore[47], Bives[48]) into a single solution FAIRDOM platform. The FAIRDOM platform will allow full storage, annotation and processing support for research assets from machine collection to final publication. This will be fine-tuned using a pilot project with EraSysAPP projects, and Synthetic Biology Centres within the UK to ensure that the platforms are suitable. This will greatly improve the burden of stewardship, and therefore encourage better research asset management practices within labs.

Key resources to be implemented ISBE wide during the legal phase should be identified.

6.3 Legal phase implementation

ISBE will establish a coordinated development of research asset management platforms that can support research asset management from instrument production to final publication. These should include Electronic Lab Notebooks (ELNs), Laboratory Information Management Systems (LIMS), construction, annotation, simulation, and commons platforms that assist with the varied data ISBE will be responsible for. The adoption of these platforms within nSBCs will be crucial to ensure that ISBE is operating successful RARE and FAIR research asset collection and publication.

ISBE will provide and develop these platforms, in conjunction with other appropriate infrastructures, and associated training, to all researchers. Business models with appropriate support costing for these platforms will be developed so that research groups can assess their needs and identify resources and costing to support their research asset management requirements. This will ensure sustainability of the platforms, research assets, and the development required to retain the resources in a usable condition.

ISBE will coordinate annual foundry workshops to engage tool and software developers within the field. Here ISBE will be a key driver in aiding with identification, and implementation of the communities requirements through the key developers within the community.

Once again Strategic Level Agreements (SLAs) with other research infrastructures relating to resources must be established.

[46] http://www.cisd.ethz.ch/software/openBIS
[47] http://sycamore.eml.org/sycamore/
[48] https://sems.uni-rostock.de/projects/bives/

7 Support for commercially sensitive and personally sensitive data

"ISBE will support life sciences research, health research and commercial collaborations in these areas. Patient data for clinical or biomedical applications requires secure and sensitive handling. A mixture of open and commercially sensitive data/models and open and commercial services should be catered for. Commercial services may form part of the ISBE data and model framework: from the publishers and publishing services through to commercial data and knowledge bases and modelling tools and underpinning commercial cloud hosting. We anticipate potential financing as a public private partnership and the implications this may have on data visibility – its accessibility and accessibility. The operating conditions that ISMB should support private and proprietary data needs to be defined.

Clear policies, standard operating procedures and supporting infrastructure are required to ensure that private health care information or commercial assets are kept with secure and restricted access (the "A" in FAIR stands for Accessible, not open). In some cases an Information Security Management System defined by Policies and Standard Operating procedures certified to ISO27001 will be required"

7.1 Status

Commercially sensitive data.

Commercially sensitive data/models are often produced during collaborations between industry and academia as part of a public private partnership. In these instances, involved researchers/institutes/companies use non-disclosure agreements. Non-disclosure agreements provide a legal framework around which interested parties can agree to share material, knowledge, or information with each other, but not with third parties. The nature of the materials, knowledge, or information must be clearly outlined in the agreement, it must be signed by all parties, and it can be designed to be bilateral (i.e all concerned parties are restricted in the re-use/sharing of information), or unilateral (i.e. only one party within the agreement is restricted with respect to re-use and sharing).

Academic institutes are also encouraging the commercialisation of research findings. Many academic institutes have set up internal centres that provide support and advice to academics on how to protect their intellectual property (e.g. patents), and also how to move towards commercialisation (e.g. Knowledge Transfer Networks). When intellectual property rights are being filed, such as patents, it is crucial that the research assets, and any supporting information is not made publicly available. This a prerequisite of being granted intellectual property rights, as any information considered already in the public domain is no longer patentable.

Personally sensitive data
Privacy aware management of data is one of the key challenges to be met by systems approaches such as systems medicine, synthetic biology, and -at the core- systems biology.

Systems biology projects need to be carefully designed to minimise interfaces between privacy sensitive clinical and less privacy sensitive systems biology data. As legal regulations have an influence on e.g. the question who can send clinical data to whom, how data can be reused, it makes sense to keep this separation of data by their privacy levels. ISBE will mainly focus on data that have few privacy restrictions. It will help its users to keep regulations with minimal overhead.

In particular, human genome data can be perfectly identified, as it is unique and does not change over a lifetime. At the same time, it carries sensitive information, not only about the person who is the DNA donor, but also about their relatives. This, in turn, causes ethical and legal problems of clinical day-to-day work as well as data management, as addressed for example by EURAT (Ethical and Legal Aspects of whole (=Total) genome sequencing) and the Global Alliance 4 Genomics and Health (GA4GH)[49].

At the same time, these problems spawn new fields of research, e.g. research about Genome Privacy, i.e. ways to combine the advances in privacy enhancing techniques with the field of 'Omics analyses.

Along with such technical research, there is research into consent models that respect the patients' privacy as well as making the most of study participation in the interest of both science and patients.

ISBE will monitor and influence the legal frameworks around sharing of science-relevant clinical data. ISBE will focus on less privacy-sensitive data. It will provide service for projects that encompass data with multiple levels of privacy sensitivity, providing data integration and aggregation across privacy levels.

7.2 Interim phase implementation (up to 3 years)

Commercially sensitive data
In the short-term commercially sensitive research assets will be supported on a case-by-case basis. A number of the primary platforms such as the FAIRDOM SEEK Platform are capable of handling research assets for common storage and exchange between specific parties, without the assets needing to be public. This means that FAIR research assets can still be produced and maintained, whilst adhering to non-disclosure agreements.

In some cases an Information Security Management System defined by Policies and Standard Operating procedures certified to ISO27001 will be required to support personal data, with the appropriate AAI management and single sign-on.

Personally sensitive data
In the same way as commercially sensitive research assets will be supported, ISBE will support mixed personally sensitive/insensitive data settings. To this end, it will liaise with infrastructures such as ECRIN-ERIC and BBMRI-ERIC, mainly focusing on the personally non-

sensitive data. Suitable architectures and data flows will be suggested and implemented on a case-by-case basis, leading to data and model best practises for such projects.
As part of this activity, ISBE will sign service level agreements with the concerned research infrastructures.

It will monitor development concerning the national and international legal frameworks, it will monitor and influence the ethical and governance best practises as proposed by the Global Alliance for Health[50] and national initiatives like EURAT.

7.3 Legal phase implementation

Much of this work should be supported through generic requirements of the research asset management plan that is established by each project that will be supported by ISBE research asset management resources. It should include when research assets are expected to be made public after a funding grant has been completed, and the due process for requesting extensions to this period of time. There should be clear definitions of the time-scale that extensions can feasibly be taken, and at what point extra special consideration needs to be sought, and these should be balanced with the requests of the associated funding organisation. Each nSBC will have to establish its own specialist requirements based on the national landscape in which it sits.

ISBE should establish/adopt industry standard non-disclosure agreements that set out the the terms of privacy. These should always aim to have the research assets in the public domain at future time, where appropriate. Again nSBCs need to be mindful of the specific requirements that must be in place for their nation state.

The technical infrastructure developed by ISBE, and the infrastructure ISBE supports through SLAs with other RIs should have capabilities for handling these exceptions to research asset publication. This process should be simple to initiate, and clear communication through online platforms, and through the physical infrastructure should be available.

ISBE should identify any particular weaknesses with catering for commercially sensitive data and see how these could be addressed through already available resources within other RIs. Where no obvious solution exists, the RIs should work together to generate support for the community, with costs/time apportioned against each infrastructures community service requirements (i.e. if it is a core activity of one infrastructure, and minor activity of another, then more resource burden should fall to the infrastructure who operate the service as a core activity).

In the legal phase, ISBE will work together with infrastructures such as BBMRI or ECRIN for supporting projects that support inter-infrastructure projects spanning multiple levels of data sensitivity. It will provide and update tools and documentation: (i) helping to design the project data flow (ii) suggesting the right cooperating infrastructures (iii) defining the interfaces and boundaries between privacy sensitive and insensitive data.

ISBE will continually identify gaps in legal and ethical regulations occurring in projects. These will inform the support of projects. It will monitor new developments in genome privacy and other techniques related to patient privacy.

[49] http://genomicsandhealth.org/
[50] genomicsandhealth.org

8 Development and adoption of common practices and standards

"The ISBE data and model management framework focuses on conventions that enable data interoperability and stewardship and compliance against data and metadata standards, policies and practices. We must define, develop and adjust criteria and standards that must be met by data, maps, tools and models; support the accuracy, reliability and quality of data, models, tools and maps.; and make the re-use of data sets, models, SOPs etc. possible in future projects.

The conventions for data and model metadata descriptions must be founded on community standards for identifiers, formats, checklists and vocabularies, developed through community engagement, to make data, models, and tools re-usable. A knowledge hub and training activities will be needed to disseminate these practices and standards, and technical development to implement them into tools.

ISBE must be an active and engaged advocate for the development and adoption of standards. We recommend a concerted action of the European systems biology infrastructure with the respective ISO committees as ISO/TC 276 with the objective of defining a horizontal framework standard for the data and model patchwork in the field. Such a strategic alliance has to include the corresponding domain-specific grassroots standardization initiatives like COMBINE, FGED, PSI, MSI and others; wider standardisation bodies such as the Research Data Alliance and the Global Alliance; and work with journals and funders to establish practical best-practice usage of community standards for publication.

ISBE should set in motion measures (training, services, and infrastructure) for the making and habitual use of standards for the research assets of Systems Biology, notably data, SOPs and models, and how these are related to each other and to investigations. "

8.1 Status

Novel technologies in systems biology generate heterogeneous and high-dimensional data sets from a wide variety of experimental setups. For the automated downstream processing of the obtained raw data and for the further integration of the data originating from different sources the key requirements are the consistent usage of standardized data

formats, as well as the standardized description of its context, reproducibility, relevance and accuracy. Thus, consistent standards for the generation and acquisition of raw instrument data and the recording of corresponding information are needed, as well as standards for the following data processing and integration steps that include transformation of the data into numeric values, data (pre-)processing (including image processing, parameter determination from the numerical raw data, etc.), data reduction, data storage and quality assurance. The processed data has to be described in its environmental and experimental context, in order to cluster and connect it to related data, e.g. for the setup of simulateable computer models. This process can be compared to assembling a jigsaw puzzle where all interfaces between the pieces have to be well defined and compatible to each other. It is essential to define stable and coherent minimal standards to cover all data processing and integration steps.

Standards are an agreed and consistent way of doing things; they represent highly distilled knowledge of experts within a field or community who know the needs of that field or community with respect to the commodity. In systems biology, the commodities are research assets including data, models, and maps. Systems biology promotes a high degree of exchange of research assets between researchers, owing to the breadth of fields that comprise it. To aid with this exchange it is helpful for the research assets to be organised in a standardised way. Format standards allow the information in a research asset to be structured appropriately, allowing the information to be identified easily by both researcher and computer. In addition to this, the context of the information also needs to be outlined, which allows decisions to be made on correct usage of the information within a given context.

As Systems Biology research and development highly depends on the integration of heterogeneous data from a manifold of technologies for which there exists mandatory standards to format and describe data and its metadata (data describing the data), the systems biology standards community is highly active. This has led to a maturity within the field, particularly for molecular systems biology, which has dominated the early advancement in the field. For physiological systems biology dealing with higher level systems as tissues, organs or even the whole body - which is becoming increasingly achievable - the standards are less mature, but in line with the community development itself. Through adopting accepted community standards from individual research fields such as genomics, transcriptomics, proteomics and metabolomics, most data types originating from the applied technologies in these fields can be formatted in a standardised way. Community standards for data and model formats include the Systems Biology Markup Language (SBML), the Systems Biology Graphical Notation (SBGN), CellML , the Proteomics Standards Initiative Molecular Interaction format (PSI-MI) or the MicroArray Gene Expression Tabular format (MAGE-Tab), among others. The metadata descriptions are supported by specific systems biology controlled vocabularies and ontologies such as SBO (Systems Biology Ontology), Mathematical Modelling Ontology (MAMO) and Kinetic Simulation Algorithm Ontology (KiSAO), or more general life science vocabularies, like GO (Gene Ontology), Ontology for Biomedical Investigations (OBI), Chemical Entities of Biological Interest (CHEBI) or the Cell Ontology (CL). There are also a wealth of minimum information checklists available through MIBBI, hosted through Biosharing (https://www.biosharing.org/) to aid with the required scope of the delivered metadata for a comprehensive description of data and models, e.g. Minimum Information About a Microarray Experiment (MIAME), Minimum

Information About a Proteomics Experiment (MIAPE), Minimal Information Required In the Annotation of biochemical Models (MIRIAM), Minimum Information About a Simulation Experiment (MIASE), and others.

Provenance – the ability to trace the origins/history of information – is a currently under-developed aspect of systems biology standards, and requires significant improvement over the coming years. The Research Object model (RO[51]) and the COMBINE archive with its underlying Open Modeling EXchange format (OMEX)[52] provide a good foundation to address this problem as they aggregate related research assets (publications, data, models, workflows, simulation settings, etc) into single collections, with defined relationships, so that they retain their context. Also initiatives which provide resolvable persistent URIs used to identify data for the life sciences as Identifiers.org (http://identifiers.org)

Overall the Systems Biology standards community is highly self-driven by a cross-section of researchers who focus on standards development either from a model centric view (COMBINE community[53]), or with the scope of a certain domain, as for the Genomic Standards Consortium (GSC), the Functional Genomics Data Society (FGED), the Metabolomics Standards Initiative (MSI) or the Proteomics Standards Initiative (PSI).

Many standards used by Systems Biology are not Sys Bio specific (for example the Gene Ontology). Other RIs such as ELIXIR and GA4GH are driving adoption of interoperable data practices and are important partners for ISBE.

8.2 Interim phase implementation (up to 3 years)

ISBE will retain a strong presence in the grass-roots standards community COMBINE, helping co-develop the "meta standards" that will be used to build a technology-independent framework of minimal requirements and rules for standardised formatting of models and corresponding data, and annotation with metadata. This will also include the adoption of standards for entity relations and corresponding qualifiers such as BioModels.net Qualifiers (http://co.mbine.org/standards/qualifiers), as well as standards for the description of the evidence and quality of the described data and its sources.

There has been some diversion in the ISATab standard for structuring research projects. An ISA hackathon will be arranged to identify where there are differences in the implementation and use of the ISA standard, and how these might be merged back into a single standard used by the community.

There will be work on developing ISO standards by sub-committees of the technical committee for biotechnology standards (ISO/TC 276). A particularly important sub-committee of this with respect to the European systems biology infrastructure is the working group "Data processing and Integration", as it aims to define interfacing and integration standards for data and models with different formats and originating from heterogeneous sources. We recommend a concerted action of ISBE with the respective domain-specific grassroots standardization initiatives like COMBINE, GSC, FGED, PSI, MSI and others, as well as with relevant international ISO committees as this ISO/TC 276 sub-committee with the objective of defining a horizontal framework standard for the data and model patchwork in the field.

ISBE will engage with the ELIXIR Interoperability Platform, which is seeking to standardise and manage identifiers, dataset reporting and API handling across a range of bioscience datasets.

ISBE will push the advancement of standard format and annotation support for multi-scale modelling, where there is a lack of standards, and analysis algorithms.

8.3 Legal phase implementation

Developing and maintaining sustainable resources that allow systems biologists an easy-accessible overview of the relevant community standards, their fields of application and typical use-cases, including information about interfacing and combination options between them. Such a system should integrate access to available mapping tools and registries such as BioSharing.org, as well as access to online services for providing resolvable persistent identifiers (URIs) used to identify data as identifiers.org.

ISBE should also be responsible for ensuring that the activities of COMBINE and HARMONY (the Sys Bio Standards conference) can continue and can be sustained. To assit this, ISBE must partner with major infrastructures working on standards, notably ELIXIR and GA4GH, and international standards bodies such as Research Data Alliance.

As the experimental technologies and modelling techniques are rapidly developing, the community standards have to be constantly adapted and extended. To keep the ISBE infrastructure up-to-date and the supported standards compatible with the implemented infrastructure, persistent and sustainable involvement in the further development and refinement of community standards and relevant official standards is crucial. With this aim, ISBE should seek long-term strategic alliances with the relevant grass-roots standardization initiatives like COMBINE, GSC, FGED, PSI, MSI and others, as well as with the corresponding committees at ISO or CEN/CENELEC, e.g. the committees for biotechnology (ISO/TC 276) or for health informatics (ISO TC 215/CEN TC 251) and their national mirrors committees.

ISBE will adopt and promote standards that allow for the exchange of models and modelling results in reproducible and reusable way. Where there are not suitable standards available, ISBE will engage the community and organise the development of these standards to ensure that modelling experimentation can be as systematic as experimental data.

The most important but the most challenging task will be to ensure the uptake of the standards by the Systems Biology community. To obtain this goal, ISBE has to build capacity for the relevant standards and their application in typical use-cases by hands-on and online tutorials, as well as 1:1 support for the researchers, if necessary. ISBE also has to build capability by providing the tooling and infrastructure necessary to make the adoption of standards feasible and worthwhile.

[51] http://www.researchobject.org/
[52] *BMC Bioinformatics* 2014, **15**:369
[53] http://www.co.mbine.org/

9 Build stewardship capacity and capability

"When ISBE acts as a broker to bring researchers who generate data into contact with researchers who require data, standards-based and model-compliant data generation must be ensured along with data management planning. We will need stewarding services support to store and explore the links between data, models, protocols and results from ISBE investigations, showing the systems level details of the experiments, and to understand how separate datasets (e.g. genomics, transcriptomics and proteomics) can be interpreted together, or how they are used for construction or validation of the model, to enable a systems level understanding. Training and education is required across the different expertise of ISBE users, stakeholders and stewards, including members of nSBCs, ranging from in-house training to curriculum development for higher education institutions. ISBE must partner with international training initiatives such as GOBLET and Software Carpentry, and national initiatives such as SysMIC."

9.1 Status

Stewardship capacity has been growing in systems biology, as it is recognised that good research asset management practices are vital due to the high level of asset exchange and reuse between researchers.

Standards for stewardship

There are a number of standards that are available for formatting data, as well as annotating data, and these are on constant development through grass-roots activities within the community. Some software exists for implementing these standards for model generation, and simulation (e.g. SemanticSBML). There are also platforms like OpenBIS which assist with annotation and formatting of data straight from the instrument.

Annotating data with terms from controlled vocabularies and ontologies can be tricky for casual users who are unfamiliar with their ontological structure and content. However, this is frequently required in order to conform to a particular data standard. RightField and Ontomaton are tools that assist with this process. RightField allows ontology term lists to be embedded into spreadsheet cells. A collection of RightField-enabled spreadsheets for standards-compliant data formatting is available from SEEK (https://seek.sysmo-db.org/help/templates), but soon will also be available from FAIRDOM. Users annotate their data by selecting from simple drop-down lists. Ontomaton allows free text to be tagged with ontology terms in Google spreadsheets and allows users to search for ontology terms for annotation.

Training for Stewardship
The community is also trying to improve stewardship capacity through training to use tools that exist, and bringing together developers as part of foundry workshops. Software and Data Carpentry workshops, for example, teach scientists the basics of versioning and managing their code and data. Once participants have attended a workshop, they can help organise the next and eventually run their own. This method allows best-practice to spread through the research community.

"Bring your own data" workshops are also gaining popularity. Users can bring their own data and models for practical examples of converting these into a standardised format. The goal of these workshops is to generate a better understanding and practicality with tool, software and standard usage for every day work.

Foundry workshops are used to identify developers, and managed resources that support research in the wider community. They bring together the key developers of the resources, along with researchers and standards developers within the community to discuss their work, and near future and long-term priorities for development. By generating a close interaction at the interface between research and specialised resources development, this allows the resources to be developed according to need of the community, making the resources more directly relevant, and more likely to become key resources for the community.

Providing services long-term means maintaining the capacity, as well as the tooling. Maintaining the capacity means, staying up to date with changes in (i) surrounding infrastructures, (ii) surrounding standards, (iii) state of the art in best practises. It also means maintaining highest standards of data stewardship in the presence of personnel churn.)

9.2 Interim phase implementation (up to 3 years)

A systems biology foundry has been established for those who generate software and tools for systems biology related projects. Meetings will be held every year where developers can present their work, and establish further steps for improving tool and software availability to support systems biology stewardship.

Training workshops using the "Bring Your Own Data" format will be held as part of EraSysAPP, where key training will be provided to frontline users. These workshops will also help to inform the standards communities about the direction and needs of the research community, and help direct the development to key areas.

ISBE should develop companion courses for courses such as SysMic [54] which offer comprehensive systems biology training for researchers with diverse backgrounds, and career stages. We should also progress in involvement with training schools and practical courses which are specifically for educating young career researchers, co-developing the courses to support training in data and model management.

9.3 Legal phase implementation

Stewardship capacity will be built throughout each nSBC, using train the trainer workshops. These workshops will ensure that all nSBC service providers are capable of implementing suitable stewardship to the data and models handled in ISBE.

Training programs will be developed to establish stewardship knowledge and practice within the community at large. These will be organised by nSBCs, and primarily cover training in the nSBCs host country.

ISBE cISBE (comprised of all nSBCs) will take responsibility to coordinate within the community of standards, tools, and software, to drive development directions in accordance with the needs of the community. This includes providing more resource to the development, conducting community surveys, and establishing the future needs for ISBE itself in the realm of standards, tools, and software. It should set out white papers which help to guide the community with the requirements for future stewardship practices in order to address the community needs.

ISBE should co-develop tools for the construction and annotation of models, as well as automatic detection of metadata (species, parameters, parameter provenance) on models submitted to ISBE managed data and model portals. These tools should be made available through ISBE cSBC portal, with what their usage is relevant for, and use guides.

ISBE nSBCs may take responsibility for annotation/curation, taking responsibility, for example, for SBML-encoded or CellML-encoded models. To increase the reusability of models, accepted models should undergo a transformation process, which would generate a variety of alternative formats. For example, an SBML encoded model can already be transformed into an SBGN (graphical) format. There are various transformations that can be undertaken for the various formats, but some of these may be 'lossy'.

[54] http://sysmic.ac.uk/home.html

10 The recognition of all assets and all stewardship activities

> *"Data and models must be citable and cited, with credit given to their authors and stewards, and commoditised so that they can be re-used modularly. Stewardship needs to be recognised and rewarded as a first class and habitual activity. Assets need to be recognised and rewarded as first class research outcomes with appropriate credit metrics. Dedicated stewards and Research Data Engineers, and those Research Software Engineers producing stewardship tools, should be recognised with established and rewarding career paths. ISBE should establish partnerships with stakeholders: institutions, funders, publishers, journal editorial boards, learned societies, pressure groups and networks (such as Force11) to advocate for the recognition of all assets and the recognition of the skills of asset stewards, develop supporting infrastructure and establish practical best-practice usage of community standards for creditable publication."*

10.1 Status

In research currently journal articles and the h-index are the main metrics for success. These are used as primary assessors of the impact of individual researchers upon the field. One of the issues with this measurement is that numerous studies have found journal articles to be non-reproducible when put to the test. A main aspect of this irreproducibility is the poor availability of the data, models, analysis/processing tools, and standard operating procedures that comprise the studies.

In addition, the availability of research assets from journals is also important to make available so that the outputs of public investment can be accessed by all interested stakeholders.

This is an aspect of science that high quality research asset management can improve. To this end, we have seen in recent years both funding councils and journals move towards improved requirements for the formatting, annotation, and public archiving of research assets produced from research projects funded by the funding councils, and published by the journals.

All funding councils have policies for published outputs (e.g. journals, conference papers), data (for access and maintenance of electronic resources), timeframes for making research assets accessible, and sharing policies. In fact many now require a full research asset management plan as part of all grants applied for. Many journals also provide guidelines and support (repositories, costs etc) for making research assets available. This is a great move

forward for research asset management as it helps to make clear the conditions of the research funding, whilst also providing help and resources to adhere to the requirements. Despite these advances supplementary data is still very prevalent in publications. Supplementary data is not suitable for archiving data and models because it is generally does not adhere to FAIR principles. There is still much that can be done with regards to specialist tools, repositories, and guidance.

Journals have also been moving towards stronger requirements for the publication of research assets associated with the paper. Most mandate that research assets used within the paper are made publicly available. Few go so far as to mandate the formats (with some proteomics/metabolomics journals as exceptions). Very few offer curation services. Although FEBs journal offers curation of any SBML models published within the journal, and this also aids with the reproducibility of the science presented using the models.

10.2 Interim phase implementation (up to 3 years)

There will a pilot phase with Journals where a similar approach to that with FEBS journal will be implemented, and all SBML submitted models will be subject to curation. During this phase a more sustainable model for implementing this, and also offering extended services (such as coding up of non-standardised models into standardised formats) will be established – ideally with the aim of researchers paying for additional stewardship curation services for their model in a similar way to open access charges.

10.3 Legal phase implementation

ISBE will work with funding councils in each nSBC to establish what requirements would be feasible and advantageous to promote that would improve the perception and uptake of research asset management (nb that longer term plans should be made so potential future requirements can begin to be catered for in tools/software/standards before introduction).

ISBE needs to coordinate meetings between key journals, and key standards/tool developers (e.g. the developers foundry) to establish how to better implement community needs for research asset management.

Need to establish business models for working with journals where more hands on support is needed to aid with curation and reformatting of models at least in the shorter term – the movement on the whole should be towards more self-sufficiency of researchers through high quality support (tutorials/software/tools etc).

11 Sustained funding and business models

> *"ISBE should seek avenues for sustainable funding for asset stewardship and public resources, and develop a portfolio of business models. transformative 5% tax. example: the Netherlands and NWO, DTL. Funding agencies and grant allocation could also allow funds to go directly to curation and stewardship activities, thereby facilitating the longevity of data and data accessibility in the longer term"*

11.1 Status

Sustainability of key resources through carefully thought out business models is fundamental for research infrastructures to add value to the research landscape, as this is the area in which many resources provided by research groups individually suffers.

ELIXIR is one of the more established research infrastructures and operates on a "node" contribution method, whereby resources are contributed by nodes (host country specific) to the ELIXIR (EU) infrastructure. The funding methods sustaining resources are established by the nodes, in a model that suits the landscape in which the node is set (e.g. resources available, funding council policies). For example the Netherlands has set up a public private partnership, DTL[55], which receives 5% of all research grants to support data management requirements in the Netherlands.

Some institutions have infrastructure already set up to support certain research asset management in the medium-term = for instance Heidelberg Institute for Theoretical Studies supports all data submitted to FAIRDOMHub, through ERANet and EraSysAPP funded projects for a minimum of 10 years after the end of the project.

Grants to kick-start sustainability from funding councils are also in play. An example would be FAIRDOM, where the one of the tasks of the funding phase is to establish a sustainability business model to ensure that the resources supported and developed within the grant see long term availability and usability.

Not for profit models are also in place for e.g. APACHE Foundation, tranSMART, Software Carpentry Foundation to name a few. Foundations are particularly popular as they allow a legal entity into which IP and funds can be channelled. The downside of foundations is that they generally levy membership fees. Which as model we would not expect to follow for a research infrastructure, particularly for the service of researchers.

Volunteerism and in Kind contributions have also show to yield sustainable results e.g. open source software, and wikipathways. It has the disadvantage that sustainability cannot be guaranteed to a high certainty. It also reduced the level of control that the providers or authors have over the content, which would not be suitable for a large research infrastructure itself.

11.2 Interim phase implementation (up to 3 years)

Each nSBC needs to assess their portfolio of resources, and identify which resources will be communal with other research infrastructures and which will be ISBE specific. It also needs to identify any key gaps that may present for research asset management in the host nSBC, and through contact with cSBC assess whether these gaps can be filled by other resources from other nSBCs that will form ISBE.

11.3 Legal phase implementation

It is envisioned that ISBE will be comprised of:

- ISBE selected public archives
- ISBE Systems Biology Commons
- ISBE endorsed catalogues and technical support services
- ISBE endorsed software platforms and affiliated tools
- Metadata specifications and templates and ontologies.
- Network community.
- Training programme and materials

All of these resources will be distributed across different nSBCs, as well as shared through SLAs with other research infrastructures. It will be the responsibility of each nSBC to take their list of identified resources and identify, along with funding councils, private investors to identify a funding model that is best suited to supporting these key resources. For example, these could include blanket requirements for all related projects (decided by the nature of a funding call) provide a specific % of the total grant output to a centralised body within the nSBC. Costs to resources could then be apportioned to specific resources based on their value/use by the community, and also in accordance with any SLA agreements. The model should be flexible so that each nSBC can decide on what is appropriate based on the nSBCs needs and resources.

Aside from general sustainability nSBCs also need to decide on business models for costing support to private investors, whereby some resources may be allowed to make profits through investment, whereas others may have strict obligations not to.

The business/funding models must also take into consideration how to apply investment into future resources that are required to support the changing nature of systems biology research. These resources may be co-developed with other research infrastructures or be ISBE specific. The main requirement however, is that there is scope for change and expansion within the portfolio as it develops.

[55] http://www.dtls.nl/

12 Develop synergies with other RIs

> *"Synergies should be identified across the various RIs in a systematic way; repositories that can provide data of use in ISBE should formalise agreements for data sharing and access, SOPs should be established for curation and annotation of datasets and models, with a clear policy established for responsibility for the data, in terms of where and how it is stored, and on the means it should be accessed by ISBE, and by the systems biologist."*

12.1 Status

The research infrastructure landscape has a large number of players, and given systems biology is fairly broad in its coverage of biological fields, there are clear areas where ISBE services and objectives interface with those of other research infrastructures. It is vital that research infrastructures work together to gain the most synergies from these interfaces, and establish common strategies and goals. These strategies and goals could relate to where the main responsibility lies in developing or implementing services at these interfaces, ensuring that all services are complementary and streamlined at these interfaces, and ensuring that there are clear strategies for investment at these interfaces.

The value of these interfaces has been noted by the EC, and support for projects to establish these common interfaces have been funded. Corbel is one such project, aimed to establish a collaborative and sustained framework of shared services between the ESFRI Biological and Medical Research Infrastructures. It is comprised of 11 BMS RIs including ISBE (BBMRI[56], EATRIS[57], ECRIN[58], ELIXIR[59], INFRAFRONTIER[60], INSTRUCT[61], EU-OPENSCREEN[62], EMBRC[63], EURO-BIOIMAGING[64], and MIRRI[65]).

What is clear is that no infrastructure will be able to provide services to each community on its own, and there will be clear funding advantages to combining infrastructures for service provision, much as there are clear advantages for establishing infrastructures in the first instance.

12.2 Interim phase implementation (up to 3 years)

The CORBEL project will be running for through the ISBE interim phase, and aims to create "harmonized accession processes, unified ethical and legal support, joint data management, and coordinated user access to advanced research instruments, facilities and samples" between the 11 BMS RIs.

Links between the services of data management in ELIXIR and ISBE will be established through collaborative talks with key members of the ELIXIR infrastructure. Early meetings to establish common ground have already been set, and from these clear downstream plans will be formulated for implementation during the ISBE legal phase.

12.3 Legal phase implementation

ISBE will provide model centric research asset management – which will mean where appropriate taking elixir data and including appropriate additional metadata and make it available.

ISBE will work with ECRIN, EATRIS, ELIXIR to identify appropriate research asset management guidelines/tools/software for safely managing personally sensitive data.

ISBE will work with PRACE, EUDAT, EGI and GEANT on identifying appropriate underpinning infrastructures for data storage, transfer, compute and security.

[56] http://bbmri-eric.eu/
[57] http://www.eatris.eu/
[58] http://www.ecrin.org/
[59] https://www.elixir-europe.org/
[60] https://www.infrafrontier.eu/
[61] https://www.structuralbiology.eu/
[62] http://www.eu-openscreen.eu/
[63] http://www.embrc.eu/
[64] http://www.eurobioimaging.eu/
[65] http://www.mirri.org/home.html

13 EU, national and community regulations and compliance vigilance.

"ISBE must maintain awareness and vigilance with respect to EU and national regulations and compliance mandates. Regulation in ISBE is challenging as national and European regulations are at play. The most notable regulation is European Commission's European Data Protection Regulation, which replaces the previous Data Protection Directive. The aim of the new European Data Protection Regulation is to harmonise the current data protection laws in place across the EU member states. As a "regulation it is directly applicable to all EU member states without a need for national implementing legislation. The regulation on the movement and processing of personal data is much tougher than previously. Other regulations are national or community standards for, Information Security Management Systems (ISO27001)."

13.1 Status

Part of managing research assets is ensuring that the management plan for each asset is compliant with any regulations, directives or policies that are applicable and/or mandated. On a general level, there will be certain requirements for storing/sharing of research assets that are defined by different funding bodies, and these will be similar for all projects funded, for instance EU funded projects must make all data available from their projects for a minimum of 10 years after the end of the project. Therefore research asset management must cater not just for the term of the project, but for the lifespan of the research asset. These are long-term commitments.

There are also instances where specific international legislation or directives that relate to ethics, standardisation, etc. will need to be taken into consideration when managing research assets. Because of the broad nature of systems biology there are a wealth of potential regulation and compliance issues to contend with. In the clinical domain, research asset management must adhere to certain aspects of Good Clinical Practice guidelines[66], EU regulatory context (Dir 2001/20/EC and 2005/28/EC)[67]. In the Biotechnology domain there are specific international standards that research assets must be compliant with (ISO/TC 276)[68]; as well as health informatics (ISO TC 215[69]/CEN TC 251) to name a few.

Some of the compliance instances for specialist research assets can involve the research asset not being able to leave the nation state – e.g. German clinical data must remain federated in Germany. In these instances, the infrastructure that houses the data must also understand the restrictions. There are also specialist requirements relating to patents, whereby any research assets associated with the patent cannot be publicly published before the patent is granted without compromising the validity of the patent.

13.2 Interim phase implementation (up to 3 years)

A generic research asset management planning resource should be established, which directs researchers to identify the specific requirements of their research assets for management and stewardship within currently available LIMS and cataloguing resources. These should provide a section which allows researchers to state the specific regulations that will apply to their research assets, so it is understood how to cater for these in the resources.

Some infrastructures such as ELIXIR, ECRIN, and BioMedBridges will be more heavily involved in the specialised research asset management compliance (e.g. clinical data, patented data) where specialist infrastructure and handling techniques are required. ISBE should form cross talk to identify which areas of research asset management it would be pertinent to form SLAs over.

13.3 Legal phase implementation

ISBE should establish resources that are capable of handling the bulk of research asset management requirements that are not seen as specialist requirements (beyond being systems biology specific). This means that the infrastructure should be established that are capable of supporting the generic requirements of each nSBC (where generic requirements are outlined by each nSBC). They should also be capable of handling commercially sensitive data, and exceptions in research asset release based on patent applications. It should also be able to hold sensitive data which has only the requirement of geographical federation in storage and availability.

More complex research asset management infrastructure, should be catered for with SLAs devised between key infrastructures such as ELIXIR, ECRIN, BioMedBridges. These will be key for ISBE to keep legal pace with the changing nature of ethics surrounding sensitive data, and specialist requirements for more niche areas of systems biology research.

It is particularly important that ISBE engages with experts in legal and ethical requirements of research asset storage, and ensures that any key ISBE resources hosted by nSBCs can be clear to users on the compliance level they offer, and therefore the their validity of use for projects – including advising on expansion of resources to support more mainstream requirements. It should be made clear the breadth of compliance that ISBE is responsible for, and the breadth of compliance that is the responsibility of the research asset generators/owners.

[66] http://www3.imperial.ac.uk/clinicalresearchgovernanceoffice/researchgovernance/goodclinicalpractice
[67] http://ec.europa.eu/health/human-use/clinical-trials/directive/index_en.htm
[68] http://www.iso.org/iso/home/standards_development/list_of_iso_technical_committees/iso_technical_committee.htm?commid=4514241
[69] http://www.iso.org/iso/iso_technical_committee?commid=54960

Glossary of Terms

RARE	Robust, Accountable, Reproducible, Explained.
FAIR	Findable, Accessible, Interoperable, Reproducible.
ELN	Electronic Lab Notebooks
LIMS	Laboratory Information Management System
SLA	Service Level Agreement
RI	Research Infrastructure
BBMRI	BBMRI the Biobanking and Biomolecular Resources Research Infrastructure (http://bbmri-eric.eu/)
EATRIS	EATRIS The research infrastructure for translational medicine. (http://www.eatris.eu)
ECRIN	ECRIN The European Clinical Research Infrastructure Network (http://www.ecrin.org).
ELIXIR	ELIXIR the pan_European research infrastructure for biological information (http://www.elixir-europe.org).
INFRAFRONTIER	Infrafrontier The infrastructure for mouse disease models and phenotype data (http://www.infrafrontier.eu)
Instruct Integrating Biology	Instruct integrated structural biology unlocking the secrets of life (htt://www.structuralbiology.eu)
EU-OPENSCREEN	EU-OPENSCREEN the European Infrastructure of Open Screening Platforms for Chemical Biology (http://eu-openscreen.de)
EMBRC	EMBRC the European Marine Biological Resource Centre (http://www.embrc.eu)
EURO-BIOIMAGING	Euro-BioImaging the research infrastructure for imaging technologies (http://www.eurobioimaging.eu)
MIRRI	MIRRI the microbial resource research infrastructure (http://www.mirri.org)