Bioschemas Project Plan

Introduction

Bioschemas aims to improve data interoperability in life sciences. It does this by encouraging people in life science to use schema.org markup, so that their websites and services contain consistently structured information. This structured information then makes it easier to discover, collate and analyse distributed data. The main outcome of Bioschemas is a collection of specifications that provide guidelines to facilitate a more consistent adoption of schema.org markup within the life sciences.

Since November 2015 ELIXIR coorganised several meetings with BD2K and Google involving efforts like GOBLET, Biosharing, Pistoia Alliance and Biocaddie to discuss how to use Schema.org for the Life Sciences. As a result we came up with several ideas that have been shaped and prioritised based on ELIXIR requirements to propose this project plan.

Within this project plan we do talk about registries and repositories. To avoid confusion we would like to clarify how the terms are used in this document. "... A registry is a list of items with pointers for where to find the items, like the index on a database table or the card catalog for a library. A repository stores the actual items, like a database table itself or a library's shelves of books ..."[1], "... registries hold references to things and repositories hold the things ..."[2]

Achievements to date

Bioschemas started as a community effort in November 2015. Many ELIXIR members have been involved since then not just in meetings but in pushing forward Bioschemas technical activities. We run several workshops and meetings that helped us to collect feedback, come up with ideas of how to apply schema.org in life sciences, engage the community, get better organised. To date several Bioschemas specifications have been created, like the Events, Training Materials and Tools specification. Use cases like the integration of Bioschemas Events and Training materials have been implemented and adopted by registries like TeSS and iAnn as well as several data providers. This has helped to demonstrate Bioschemas is not useful just to facilitate search engine indexing but facilitate the work of our specialised registries to collect and disseminate information. Use cases like this bring good examples of how we can do something similar for biological data.

[1] "What is the difference between Registry and Repository from SOA ..." 2015. 11 Aug. 2016 <http://stackoverflow.com/questions/2276124/what-is-the-difference-between-registry-and-repository-from-soa-point-of-view>
[2] "Software Engineering - Best Practices: [Misc] Registry vs. Repository." 2008. 11 Aug. 2016 <http://best-practice-software-engineering.blogspot.com/2008/04/misc-registry-vs-repository.html>

Approach

In this project we aim not to spread too thin and the same time be as inclusive as possible. To do so we plan to do most of the work proposed in this project in collaboration during several hands-on workshops. So far the funding is distributed across three ELIXIR nodes (ELIXIR EMBL-EBI, ELIXIR UK and ELIXIR NL) leading the implementation of this project. Several ELIXIR nodes like ELIXIR FI, ELIXIR DE, ELIXIR ES, ELIXIR DK and ELIXIR SE as well as external partners like BBMRI, BD2K and Force11 participate in all the project and will be supported with travel funding to participate in our activities.

The objectives of this project are:

- Develop Bioschemas specifications and demonstrators for prioritised content types (eg. Data repository, Dataset, Sample, Phenotype and Protein annotations)
- Facilitate the discovery and validation of Bioschemas compliant resources
- Support and engage the Bioschemas community (eg. via meetings, hackathons, knowledge dissemination and training).

This project is organised in subprojects aligned to ELIXIR scientific use cases and platform activities.

MaU = Marine metagenomics use case, **PlU** = Plant sciences use case, **RaU** = Rare disease use case, **HuU** = Human data use case, **DaP** = Data platform, **ToP** = Tools platform, **CoP** = Compute platform, **InP** = Interoperability platform, **TrP** = Training platform

Category	ID	Subproject	MaU	PlU	RaU	HuP	DaP	ToP	CoP	InP	TrP
Life sciences Content Types	1	Data repositories	x	x	x	x	x	x	x	x	x
	2	Datasets	x	x	x				x	x	
	3	Beacons				x					
	4	Samples	x	x	x	x				x	
	5	Phenotypes		x	x						
	6	Protein annotations					x				
Discovery and Validation	7	Bioschemas registry								x	
	8	Bioschemas validation								x	
Community Support	9	Project and community coordination								x	x

“Data repositories”, “Datasets”, “Samples”, “Protein annotations” and ”Project and community coordination” are subprojects funded by the ELIXIR Bioschemas implementation study. The ”Beacons” subproject will be delivered by the Beacons implementation study and supported by this implementation study. The “Phenotype” subproject will be supported with travel funding from the ELIXIR Bioschemas implementation study and we aim to push part of the proposed work forward through EXCELERATE activities. Other important subprojects like “Bioschemas registry” and “Bioschemas validation” might be pushed by nodes or other stakeholders interested to engage in this project. Other sources of funding might be made available to support this project and additional activities.

Subproject details

Life sciences Content Types

The objective of the following sub-projects is to engage data providers and registries to define and work towards the adoption of life sciences schema.org content types.

Data repositories

Lead	EMBL-EBI (Henning Hermjakob)
Members	EMBL-EBI, ELIXIR UK
Delivery	Starting from 1 January 2017 for a period of 12 months
Funding	ELIXIR Bioschemas implementation study (7PMs)

Problem

- Most Life sciences data repositories are missing a home page providing information about themselves with consistent structured data that would help search engines and registries to index them.
- Several registries (eg. biosharing, bio.tools, identifiers.org, ...) maintain overlapping efforts to collect certain metadata (eg. title, description, keywords, ...) about “data repositories” (eg. UniProt Knowledgebase, Human Protein Atlas, Protein Data Bank, ...)
- Most of these registries have a manual curation process
- There is lack of consistency between the metadata collected by these registries

Objectives

- Describe data repositories using Bioschemas compliant markup so data repositories can be more easily indexed by search engines and registries.

- Evaluate how registries should collect structured metadata exposed by data repositories to facilitate an automatic or semiautomatic update their records and present more consistent descriptions.
- Explore how to collect structured metadata for some of the metrics proposed by the ELIXIR data platform.

Milestones

1.M1 Identify needs and define technical use cases
    - Describe and justify use cases. Potential use cases are: identifiers, access interfaces, citation, data release reports, metrics, ...

1.M2 Analysis and mapping of metadata already used in existing registries and data repositories.

1.M3 Define minimum information guideline based on mapping results
    - Identify a minimum set of common properties
    - Identify domain specific properties not common but required by specific use cases

1.M4 Test adoption and improve specification with selected data repositories

1.M5 Test how registries could ingest structured data exposed by data repositories

1.M6 Propose any new suggested types or properties to schema.org

Deliverables

1.D1 Bioschemas specification

1.D2 Data repository using Bioschemas compliant markup

1.D3 Data registry using Bioschemas compliant markup

References

- 06- Automating the collection of “data repositories” metadata
- 03- Data release reports with Schema.org
- Data repositories draft specification

Datasets

Lead	ELIXIR UK (Susanna A Sansone)
ELIXIR Members	ELIXIR UK, EMBL-EBI, ELIXIR NL
Delivery	Starting from 1 January 2017 for a period of 12 months
Funding	ELIXIR Bioschemas implementation study (7PMs)

Problem

- Most dataset repositories and registries of dataset do not provide structured data easily crawlable by search engines.
- Registries like DataMed, OMICsDI and BioSamples do automated ingestion of content mainly through APIs but not all the data repositories have a programmatic interface and the existing variety of programmatic interfaces are subject to changes which break integration workflows.

Objectives

- Facilitate the ingestion of datasets metadata from data repositories (databases) into search engines and dataset registries like OMICsDI and DataMed via Bioschemas
- Automate the linking of datasets metadata to samples in dataset registries like Biosamples, and identify cases where samples are missing or metadata is absent.
- Engage and help data providers to test and adopt the exposure of dataset metadata Bioschemas
- Contribute to increase the number of indexed data repositories via Bioschemas.
- Make dataset registries compliant with Bioschemas.

Milestones

2.M1 Test adoption and improve specification with selected data repositories
2.M2 Propose any new suggested types or properties to schema.org

Deliverables

2.D1 Dataset Bioschemas specification based on the schema.org dataset model and the feedback from existing dataset standards in life science: eg. DATS, OMICsDI and W3C HCLS using as a template the existing schema.org dataset type.
2.D2 Data registry using Bioschemas compliant markup
2.D3 Automate the linking of datasets metadata to samples and identify cases where samples are missing or metadata is absent

References

- BD2K DataMed DATS annotated with schema.org
- 06- Automating the collection of “data repositories” metadata
- 08- Schema.org biological dataset via Bioschemas.org
- 13- Porting JATS into Schema.org
- Google science dataset documentation
- Ontology-based Dataset Exploration
- The healthcare and life sciences community profile for dataset descriptions

Beacons

Lead	ELIXIR Hub (Serena Scollen)
ELIXIR Members	ELIXIR FI, ELIXIR ES
Delivery	Starting from 1 January 2017 for a period of 12 months
Funding	ELIXIR Beacon implementation study

Problem

- At the moment the registration of a Beacon service in the Beacon Network is done manually and needs to be updated manually if the beacon service changes

Objectives

- Expose Beacon service metadata in its default landing web page with Bioschemas
- Explore automated ingestion of the beacon service metadata into the Beacon Network

Milestones

3.M1 Analyse how to use or extend Schema.org to describe a Beacon service
3.M2 Adopt Bioschemas at least in one beacon implementation
3.M3 Test adoption and improve specification with selected beacon end points
3.M4 Propose any new suggested types or properties to schema.org

Deliverables

3.D1 Bioschemas specification
3.D2 Implement support of Beacons compliant with Bioschemas via the Beacon Network

References

- https://beacon-network.org

Samples

Lead	ELIXIR EMBL-EBI (Helen Parkinson)
ELIXIR Members	ELIXIR UK, EMBL-EBI, ELIXIR NL
Delivery	Starting from 1 January 2017 for a period of 12 months
Funding	ELIXIR Bioschemas implementation study (7PMs)

Problem

- Information of samples is scattered in multiple and dispersed samples data repositories.
- Not all the sample data repositories have a programmatic interface and the existing variety of programmatic interfaces are diverse and changeable.

Objectives

- Facilitate the ingestion of sample metadata from data repositories (eg. Biobank databases) into registries like the Biosamples, BBMRI Biobank directory or the UKCRC Tissue Directory via Bioschemas.
- Engage and help data providers and developers of BioBank LIMS to test and adopt the exposure of sample metadata via Bioschemas
- Contribute to contextualise information from data sample registries (eg. Biosamples) and biobank sample repositories (eg. NL Biobank) and Biobank Registries (eg. BBMRI Biobank directory)
- Make registries like Biosamples compliant with Bioschemas.

Milestones

4.M1 Analysis and mapping of metadata already used in existing sample registries and defined by existing standards like MIABIS

4.M2 Define minimum information guideline based on the results of the mapping and feedback from registries of biological samples.

- Identify a minimum set of properties common across repositories

4.M3 Test adoption and improve specification with selected data repositories

4.M4 Propose any new suggested types or properties to schema.org

Deliverables

4.D1 Bioschemas specification

4.D2 Data repository using Bioschemas compliant markup

4.D3 Data registry using Bioschemas compliant markup

References

- 10- Discovery and registration of Biobanks resources
- ELIXIR samples club

Phenotypes

Lead	TBC

ELIXIR Members	TBC
Delivery	TBC
Funding	EXCELERATE

Problem

- Information of phenotypes is scattered in multiple and disperse samples data repositories.
- Not all the phenotype data repositories have a programmatic interface and the existing variety of programmatic interfaces are diverse and changeable.

Objectives

- Relay on the metadata description defined by the ELIXIR plant use case
- Automate the ingestion of sample metadata from phenotype data repositories into registries via Bioschemas.
- Engage and help data providers to test and adopt the exposure of phenotype metadata with Schema.org via Bioschemas.
- Make registries like TransPlant compliant with Schema.org via Bioschemas.
- Focus on plant phenotypes but consider a general definition of phenotype taking into account different types of phenotypes. eg. biomedical phenotypes, mouse phenotypes, ...

Milestones

5.M1 Identify needs and define use cases: biomedical phenotypes as well as plant phenotypes
5.M2 Analysis and mapping of metadata already used in existing registries
5.M3 Define minimum information guideline based on the results of the mapping and feedback from data repositories
    - Identify a minimum set of properties common across repositories
    - Identify set of properties required by specific repositories
5.M4 Test adoption and improve specification with selected data repositories
5.M5 Propose any new suggested types or properties to schema.org

Deliverables

5.D1 Bioschemas specification
5.D2 Data repository using Bioschemas compliant markup
5.D3 Data registry using Bioschemas compliant markup

References

- MIAPPE
- PhenoPacket

Protein annotations

Lead	ELIXIR EMBL-EBI (Maria Martin)
ELIXIR Members	EMBL-EBI
Delivery	Starting from 1 January 2017 for a period of 12 months
Funding	ELIXIR Bioschemas implementation study (6PMs)

Problem:

- In schema.org we cannot find life science types (eg. protein, gene, biological pathway) except those types that overlap with healthcare and medicine domains defined by the health schema.org extension (eg. drug, artery) [ref].
- In previous meetings we discussed the benefits of of Schema.org with several data providers but we also came with a list of concerns that need to be evaluated to be able to encourage data providers to adopt Bioschemas.

Objectives

- Test the adoption of schema.org around a protein use case involving protein resources.
- Evaluating the issues and benefits about how to work with schema.org and Bioschemas

Milestones

6.M1 Identify needs and define use case
6.M2 Analysis and mapping of metadata already used in selected protein data repositories and standards used for protein annotations
6.M3 Define minimum information guideline based on the results of the mapping and feedback from protein data repositories
6.M4 Test adoption by data repositories
6.M5 Report evaluating the issues and benefits, including feedback and guidance about how to work with schema.org and Bioschemas
6.M6 Propose any new suggested types or properties to schema.org

Deliverables

6.D1 Bioschemas specification including a draft schema.org data model for protein annotations.
6.D2 Data repository using Bioschemas compliant markup
6.D3 Create proof of concept client integrating annotation from several resources

References

- 07- Piloting a schema.org protein type

Discovery and Validation

The objective of this workstream is to facilitate the discoverability and quality of Bioschemas compatible resources.

Bioschemas registry

Lead	TBC
ELIXIR Members	TBC
Delivery	TBC
Funding	TBC

Problem

- There is no service that help users to find the sites that are compliant with a schema.org type.
- Even if we know which websites provide a type of content we need to know the url path where to find such information to make more efficient crawling.

Objectives

- Provide functionality so providers or consumers can describe which websites and in which url paths we can find specific content types marked up with Schema.org and Bioschemas.

Milestones

7.M1 Identify requirements and define use cases based on the crawling needs of registries mentioned in this proposal

Deliverables

7.D1 Create functionality to access programmatic sites compliant with Bioschemas

References

- 16- Validation and visualisation of schema.org/bioschemas compliant web sites

Bioschemas validation

Lead	TBC

ELIXIR Members	TBC
Delivery	TBC
Funding	TBC

Problem

- Though search engines provide validation of the schema.org structured data provided in a page it does not make an analysis of the content of a site and do not validate important features in Bioschemas like compliance with content guidelines, vocabularies or cardinality.

Objectives

- Provide a GUI to validate Bioschemas compliant websites and Bioschemas compliant sites
- Validate data repositories adopting Bioschemas

Milestones

8.M1 Identify requirements

Deliverables

8.D1 Create validation tools including a GUI

References

- 16- Validation and visualisation of schema.org/bioschemas compliant web sites
- Validata: RDF Validator using Shape Expressions

Community support

The objective of this workstream is to support this project and the involvement of the Bioschemas community.

Project and community coordination

Lead	ELIXIR UK (Carole Goble)
ELIXIR Members	ELIXIR UK (Susanna A Sansone), ELIXIR EMBL-EBI (Henning Hermjakob, Helen Parkinson and Maria Martin)
Delivery	Starting from 1 January 2017 for a period of 12 months
Funding	ELIXIR Bioschemas implementation study (6PMs)

Problem

- This project includes many stakeholders and several workstreams. For this project to be successful it will require good communication and coordination, not just among partners but also with the Bioschemas community.

Objectives

In collaboration with the Groups:

- Provide support, facilitate communication and engage not just project partners but the community including Bioschemas and other related efforts like Force11 and BD2K
- Produce and enhance documentation to facilitate the adoption of Bioschemas from a technical perspective
- Provide project coordination
- Make sure there is alignment among workstreams as well as alignment among specifications
    - Make sure there is a common set of minimum properties that facilitates contextualisation
- Coordinate with the ELIXIR interoperability platform and external partners

Milestones

In collaboration with the Groups:

9.M1 Arrange regular meeting calls for the project and Bioschemas
9.M2 Update Bioschemas website, manage github site.
    9.Ma Bioschemas group pages maintained by each group.
9.M3 Facilitate the delivery of tasks
9.M4 Facilitate the provision of training, training materials and documentation
9.M5 Promote Bioschemas and project activities
9.M6 Organise and drive open hands on meetings
    9.Ma Kick-off meeting - Plan
    9.Mb Hands-on workshop - Agreement
    9.Mc Hands-on workshop - Adoption

Deliverables

In collaboration with the Groups:

9.D1 Map and set common properties across specifications
9.D2 Project report

Timeline

Activities planned for 2017

ID	Subproject	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
1	Data repositories	1.M1		1.M2 1.M3	1.D1				1.M4 1.D2	1.M5	1.D3	1.M6	
2	Datasets				2.D1				2.M1		2.D2 2.D3	2.M2	
3	Beacons												
4	Samples			4.M1 4.M2	4.D1				4.M3 4.D2		4.D3	4.M4	
5	Phenotypes												
6	Protein annotations	6.M1		6.M2 6.M3	6.D1				6.M4 6.D2		6.M5 6.D3	6.M6	
7	Bioschemas registry												
8	Bioschemas validation												
9	Project and community coordination	9.M6a 9.M1 9.M2		9.M6b 9.M3	9.D1				9.M6c 9.M4	9.M5			9.D2
	Stages	**Planning**		**Agreement**		**Adoption**				**Application**			

Meetings

Date	Meeting	Type
6-8 March	Bioschemas ELIXIR implementation study: Planning meeting	F2F (9.M6a)

Stakeholders

The table below includes the list of stakeholders we would like to engage in this project. It also includes the type of role they can have in this project, the umbrella organisation they represent, the subproject they would be involved in, the main stakeholder representative, the organisation behind each representative and the assignment of resources. We could not provide Person Months (PMs) for all the relevant stakeholders, however we have assigned to many of them Travel Units (TUs) to

cover their expenses and engage them in our meetings and workshops. Since most of the partners are EMBL-EBI partners we decided to host all the workshops at the Wellcome Genome Campus. This way we will be able to count with the participation of more stakeholders.

Steps to claim back travel expenses

- One “Travel Unit” should cover the travel expenses for one person to attend one meeting, as indicated in the table below.
- Each travel unit is maximum of **500 EUR**, unless another budget is agreed in writing with ELIXIR Hub prior to the meeting. Please consult with Rafael Jimenez especially if you are coming from outside Europe.
- If you’re planning to attend a Bioschemas meeting and wishing to claim travel expenses, you must notify the project leaders Carole Goble and Rafael Jimenez. In practise, there will be a sign-up sheet for each meeting to collect this information. You must ensure the signup sheet is up-to-date with your information.
- Travel expenses can only be claimed after the travel has taken place using the ELIXIR travel expense claim form.
- The form must be duly filled as per the instructions, signed and sent based on instructions to the ELIXIR Hub with the original travel receipts attached. Please note: The form needs to be signed (no copies/PDF) and sent with original receipts, copy of signature will not be accepted.
- The form must be sent no later than 6 weeks after the meeting has taken place.
- **The travel expenses cannot be claimed unless the above conditions are met.**
- Note that for partners from University of Manchester and Oxford e-Research Centre from ELIXIR UK and for ELIXIR NL partners, travel units are included in the funds given to the Node so no additional expenses may be claimed from the Hub.
- Note that for ELIXIR FI and ELIXIR SP, travel funds are included in the budget of the 2017 ELIXIR Beacon project.

***PMs** = Person per Month, **TUs** = Travel Unit* *Participation confirmed,* *Unable to participate* *Not confirmed yet*

Stakeholder	Type	Partner	Subproject	Representative	Organisation	TUs	PMs	Workshops		
								1	2	3
University of Manchester	Coordination	ELIXIR UK	9	**Carole Goble** Niall Beard	University of Manchester	3	6	x	x	x
Biosamples	Registry	ELIXIR EMBL	1, 2, 4	**Helen Parkinson** Tony Burdett Simon Jupp	EMBL-EBI		4	x	x	x
PDBe	Data repository	ELIXIR EMBL	1, 2, 6, 7, 8	**Sameer Velankar** Saqib Mir	EMBL-EBI		4	x	x	x

Pfam	Data repository	ELIXIR EMBL	1, 2, 6, 7, 8	**Rob Finn** Aurélien Luciani	EMBL-EBI		4	x	x	x
DATS	Standard	ELIXIR UK	1, 2	**Alejandra Gonzalez-Beltran** Susanna-Assunta Sansone Philippe Rocca-serra	Oxford e-Research Centre	3	2	x	x	x
Identifiers.org	Registry	ELIXIR EMBL	1, 7, 8	**Henning Hermjakob** Nick Juty Sarala Wimalaratne	EMBL-EBI		2	x	x	x
BioSharing	Registry	ELIXIR UK	1	**Peter Mcquilton** Alejandra Gonzalez Philippe Rocca-serra Susanna-Assunta Sansone	Oxford e-Research Centre	3	2	x	x	x
OLS	Registry	ELIXIR EMBL	1, 4	**Simon Jupp** Helen Parkinson Tony Burdett	EMBL-EBI		2	x	x	x
Molgenis	Registry	ELIXIR NL	1, 2, 7, 8	**Morris Swertz**	University of Groningen	3	2	x	x	x
ISA-Tab	Standard	ELIXIR UK	4	**Philippe Rocca-serra** Alejandra Gonzalez Susanna-Assunta Sansone	Oxford e-Research Centre	3	1	x	x	x
UniProt	Data repository	ELIXIR EMBL	1, 2, 6, 7, 8	**Maria Martin** Leyla Garcia Xavier Watkins	EMBL-EBI		4	x	x	x
BBMRI-ERIC Directory	Registry	BBMRI	1, 4	**Morris Swertz** **Petr Holub** David van Enckevort	BBMRI-ERIC					
BBMRI-ERIC Data harmonization toolset (MDR..)	Service/registry	BBMRI	1,2,4,5	**Kaisa Silander** Sebastian Mate **Petr Holub**	BBMRI-ERIC					
BBMRI-ERIC Locator	Registry	BBMRI	2,4,5	**Frank Ückert** Rumyana Proyonova Gianluigi Zanetti **Petr Holub**	BBMRI-ERIC					
BIBBOX	Service	BBMRI	1,2,3,4,5	**Heimo Müller** Robert Reihs **Petr Holub**	BBMRI-ERIC					

Bio.Tools	Registry	ELIXIR DK	1	**Jon Ison**	Technical University of Denmark	3		x	x	x
OMICsDI	Registry	ELIXIR EMBL	1, 2	**Henning Hermjakob** Yasset Perez	EMBL-EBI					
transPLANT	Registry	ELIXIR EMBL	1, 5	**Paul Kersey** Dan Bolser	EMBL-EBI					
EGA	Data repository	ELIXIR SP	1, 2, 3, 7, 8	**Jordi Rambla** Dylan Spalding	CRG EMBL-EBI					
Brassica Information Portal	Data repository	ELIXIR UK	5	**John Hancock** Carlos Horro	Earlham Institute	3		x	x	x
COPaKB	Data repository	BD2K	1, 2, 6, 7, 8	**Peipei Ping**	UCLA	1				
HPA	Data repository	ELIXIR SE	1, 2, 6, 7, 8	**Kalle von Feilitzen**	SciLifeLab	3				
BRENDA	Data repository	ELIXIR DE	1, 2, 6, 7, 8	**Antje Chang**	Braunschweig University of Technology					
Beacon Network	Registry	ELIXIR FI	1, 3	**Ilkka Lappalainen**	CSC - IT Center for Science					
BioCatalogue	Registry	ELIXIR UK	1	**Carole Goble** Niall Beard	University of Manchester	2				
DataMed	Registry	BD2K	2	**Jeffrey Grethe** Ian Fore	USCD	1		x		
EZID/Name2thing	Registry	BD2K	1	**John Kunze**	University of California	1				
UKCRC Tissue Directory	Registry	BBMRI	1, 4	**Philip Quinlan**	The University of Nottingham	3				
EMBL-EBI search	Registry	ELIXIR EMBL	1, 6	**Rodrigo Lopez**	EMBL-EBI					
IMPC	Registry	ELIXIR EMBL	1, 5	**Helen Parkinson**	EMBL-EBI					
TeSS	Registry	ELIXIR UK	1	**Niall Beard**	University of Manchester	3				
MIABIS	Standard	BBMRI	4	**Roxana Merino Martinez** Jan-Eric Litton Petr Holub	Karolinska Institute	2				
MIAPPE	Standard	ELIXIR EMBL	5	**Paul Kersey**	EMBL-EBI					
PhenoPackets	Standard	BD2K	5							
W3C HCLS dataset	Standard	W3C	2	**Alasdair J G Gray**	Heriot Watt University	1				
OMICsDI XML	Standard	BD2K	2	**Henning Hermjakob**	EMBL-EBI					

				Yasset Perez						
SampleTab	Standard	ELIXIR EMBL	4	**Helen Parkinson** Tony Burdett Simon Jupp	EMBL-EBI					
Force11 data citation group	Standard	FORCE11	1	Tim Clark	Harvard University	1				
Biodbcore	Standard	Biosharing	1	Susanna-Assunta Sansone	Oxford e-Research Centre					
DCAT	Standard	DCAT	1							
BBMRI-NL biobank catalogue	Registry	BBMRI	1, 4	Morris Swertz	University of Groningen					
RD-connect sample catalogue	Registry	BBMRI	1, 4	Morris Swertz	University of Groningen					
CHD7.org	Registry	BBMRI	1, 4	Morris Swertz	University of Groningen					
DEB-Central.org	Registry	BBMRI	1, 4	Morris Swertz	University of Groningen					
Gene3D	Data repository	ELIXIR UK	1, 6	**Christine Orengo** Ian Sillitoe	UCL	3				
Schema.org	Standard	Schema.org	2	**Dan Brickley** Natasha Noy	Google					
Intermine	Data repository	ELIXIR UK	2	**Gos Micklem** Justin Clark-Casey	University of Cambridge					
		ELIXIR NL	2	**Michel Dumontier**	Maastricht University					
		ELIXIR Hub	9	**Norman Morrison**	ELIXIR interoperability platform					
		ELIXIR Hub	3	**Serena Scollen**	ELIXIR genomics data and translational data					
		ELIXIR BE	4	**Frederik Coppens**						
Datacite	Registry	Datacite	1,2	**Martin Fenner** Kristian Garza	Datacite					
EU-SOL BreeDB database	Data repository		5	**Richard Finkers**	Wageningen University and Research					
		ELIXIR CH		**Jürgen Haas**	SIB					
		ELIXIR SP		**Josep Lluis Gelpi**	BSC					
						42	**33**			