



Challenges and solutions for life science data in Europe -  
building bioinformatics capacity in the Czech Republic via the  
ELIXIR Infrastructure

Jiri Vondrasek

Athens, Greece, 18 May 2018



[www.elixir-czech.cz](http://www.elixir-czech.cz)

# ELIXIR CZ: A Node of European infrastructure for biological information

Data infrastructure for Europe life-science research:



Data



Interoperability



Tools

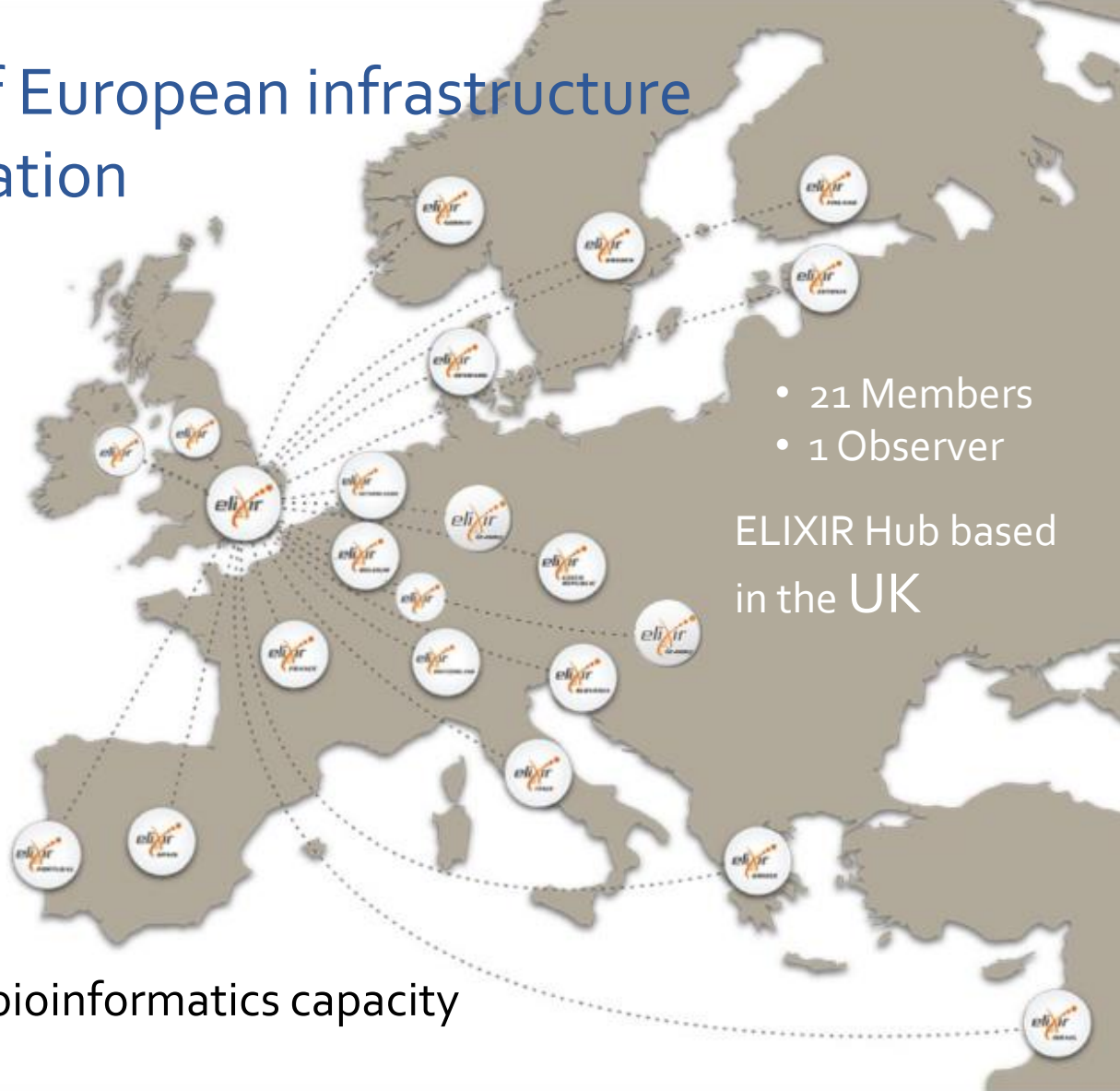


Compute



Training

- ELIXIR Nodes build local bioinformatics capacity throughout Europe
- Over 180 institutes involved in ELIXIR Nodes



- 21 Members
- 1 Observer

ELIXIR Hub based in the UK



[www.elixir-europe.org](http://www.elixir-europe.org)

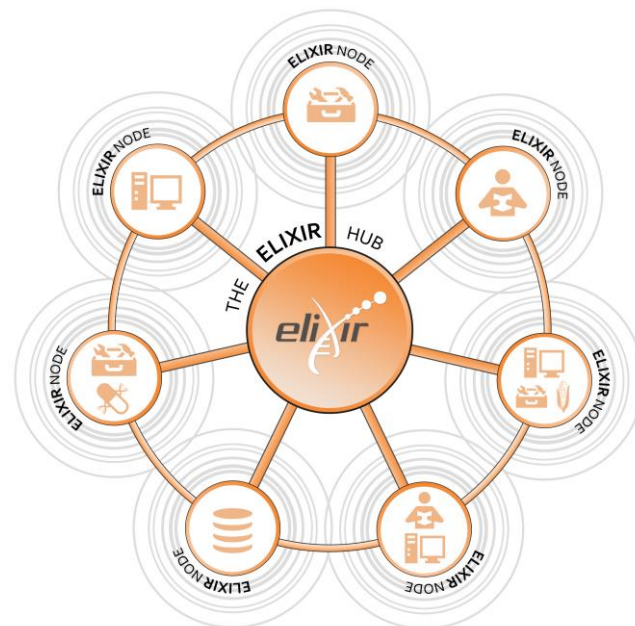


[@ELIXIREurope](https://twitter.com/ELIXIREurope)



# ELIXIR CZ Milestones

- **2012:** Memorandum of Understanding - Czech Republic to join ELIXIR.
- June 2013: Memorandum of Collaboration among institutions in the Czech Republic.
- December 2013: Ratification of ELIXIR EU Consortium Agreement.
- 2014: ELIXIR CZ Node Application.
- March 2015: ELIXIR CZ Consortium Agreement (amended October 2015).
- December 2015: ELIXIR EU Collaboration Agreement signed by the CZ Node.
- **January 2016:** ELIXIR-CZ infrastructure project funding by the Ministry of education, youth and sports (MEYS).



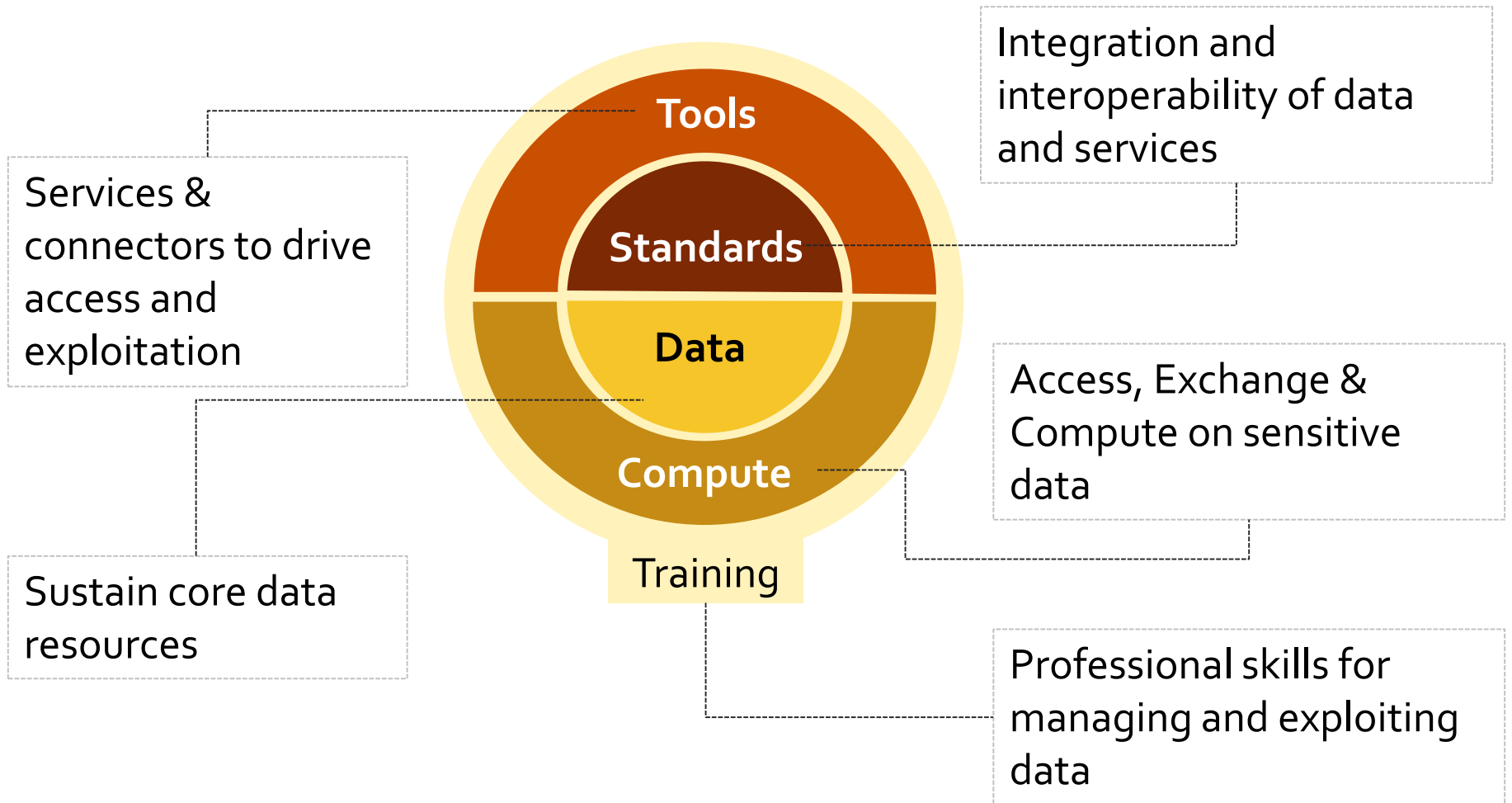
# ELIXIR CZ Governance

- 14 consortium partners
- **ELIXIR CZ Board** - one representative per consortium partner, meeting 3 times per year
- **Node committee** - one representative per consortium partner, teleconference every 2 weeks
- **Head of the ELIXIR CZ Node** - member of ELIXIR Heads of Nodes committee
- **ELIXIR CZ Scientific Advisory Board** - advice on scientific strategy
- Key strategies/guidelines adopted by ELIXIR CZ to manage the infrastructure (Communication strategy, Standardised approach to user count, ...)



[www.elixir-czech.cz/about-elixir-cz/organisation-structure](http://www.elixir-czech.cz/about-elixir-cz/organisation-structure)

# Platforms of ELIXIR CZ Infrastructure



# ELIXIR CZ Services



## Data deposition:

BEGDB, HERVdb, IRESITE, ...



## Bioinformatics tools:

MOLE, RepeatExplorer, ...



## Training:

Data and tools related workshops, Conferences, eLearning



## Compute:

Secure data transfer, cloud computing, AAI



## Data management:

Genome annotation  
Data Stewardship



## Data Interoperability:

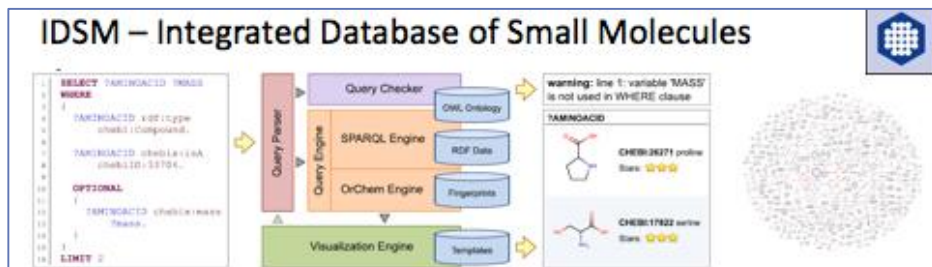
FAIR principles, Ontology services



## ELIXIR CZ Named Services

- Quality indicator of the Service or Tool
- Sustainability & Commitment to Service
- Link to ELIXIR Training programme

# ELIXIR CZ Named Service



## IOCB RDF Platform: PubChem data

The screenshot shows the IOCB RDF Platform interface. It displays a table of PubChem data with columns for 'PubChem ID', 'Molecular Weight', 'Molecular Weight', 'Molecular Weight', and 'Molecular Weight'. The interface also includes a search bar and a 'Results' section with a list of entries.

## PatternQuery

detection of structural fragments in biomacromolecules

Diagram A shows a protein structure with a highlighted fragment. Diagram B shows a protein structure with a highlighted fragment. Diagram C shows a protein structure with a highlighted fragment.

`Rings(5 * ['C'] + ['O']). ConnectedResidues(0). AmbientResidues(4)`

<http://ncbr.muni.cz/PatternQuery>

PatternQuery language – Python based language for description of structural patterns / fragments

PatternQuery web server – application for their detection in the entire Protein Data Bank ~ max 30minutes / query

Incorporated into the Protein Data Bank Europe as the Coordinate Server:

<http://www.ebi.ac.uk/pdbe/coordinates/>

## (Multi)SETTER – SEcondary sTRucture-based TERTiary structure superposition tools

<http://siret.ms.mff.cuni.cz/setter>

The diagram shows the workflow of (Multi)SETTER. It starts with 'Input structures' (represented by a protein structure), followed by 'Superposition' (represented by a protein structure with a red line), and finally 'Tree generation' (represented by a tree diagram).

## MOLEonline 2.0 – channels and pores

The screenshot shows the MOLEonline 2.0 interface. It displays a protein structure with a highlighted channel or pore. The interface includes a 'Job review #2017' section, a 'Channels and pores' section, and a 'Settings' section.

2j0d  
CYP  
3A4

<http://mole.upol.cz>

or download from <http://mole.chemi.muni.cz>

## SETTER

structure superposition

MultiSETTER  
multiple structure superposition

The diagrams show protein structures being superposed. The first diagram shows a protein structure with a red line. The second diagram shows a protein structure with a red line. The third diagram shows a protein structure with a red line.





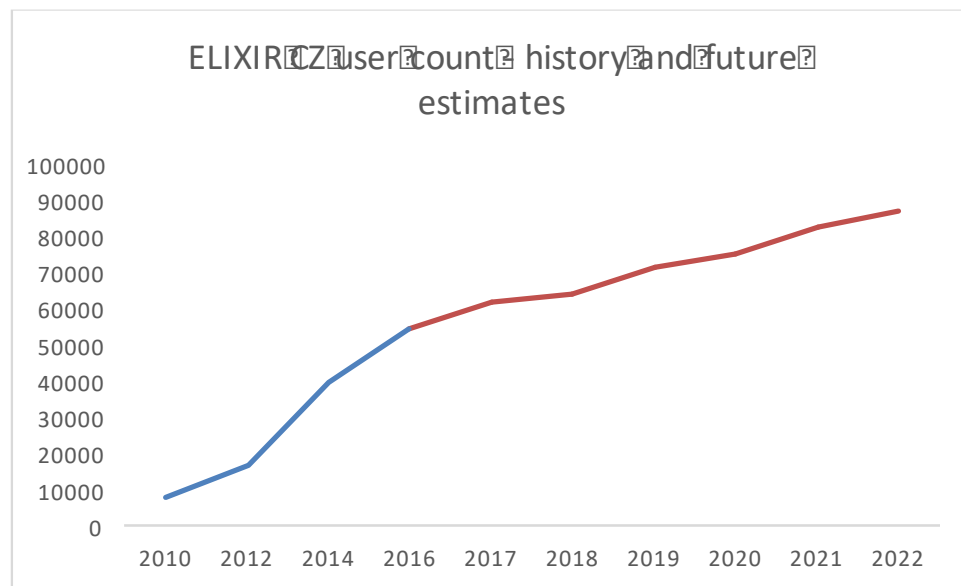
# ELIXIR CZ Data and Tools

## ELIXIR CZ Data and Tools services:

- 9 databases of life science data
- 18 tools focused on the processing and analysis of life science data
- 54735 users in 2016

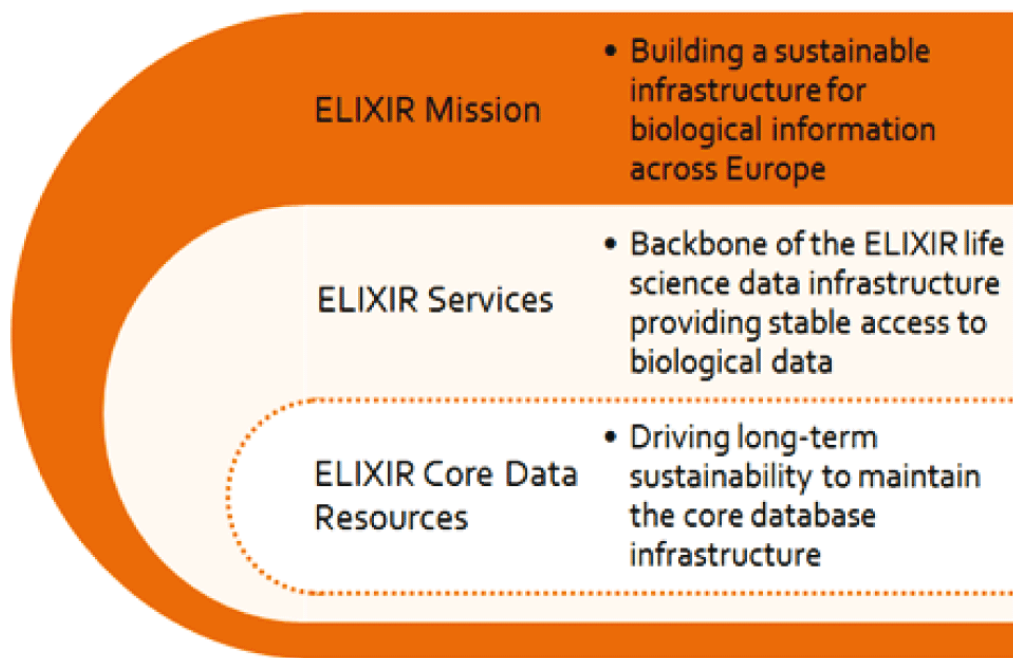
## Most used ELIXIR CZ services in 2016:

- **BEGDB** (The Benchmark Energy & Geometry Database): 4550 users
- **HERVdb** (Identification and description of endogenous retroviruses): 4962 users
- **IRESITE** (Internal Ribosome Entry Site): 18864 users
- **MOLE** (Identification of channels and pores in proteins with automatic start point detection): 2149 users
- **RepeatExplorer** (A computational pipeline for genome-wide characterization of repetitive elements from NGS): 16371 users





# The place of ELIXIR CZ Services and ELIXIR Core Data Resources within ELIXIR's mission.



# ELIXIR CZ Activity in Data Platform

Mission: Towards ELIXIR Core Data Resources - Small Molecules Database

## Project tasks

- in-house implementation
- integrate different data sources
- allow powerful database querying
- support similarity and substructure search

# Semantic web database

- well suitable for integration of sources
- data are expressed as simple statements
  - subject – predicate – object .
  - penicillin – is an – antibiotic .
  - penicillin – has formula –  $C_9H_{11}N_2O_4SR$  .
- some data are already available in RDF form
- exist many standard RDF ontologies
- query language is easier than SQL

# OrChem: tool for structure search

- originally developed by EMBL-EBI
  - for Oracle database
  - tested on 3.5 million compounds
- modified by our group
  - ported for PostgreSQL database
  - very optimized
  - used for 95 million compounds

# PubChem RDF database

- provides information on biological activities of small molecules
- contains:
  - substances: 211.004.779
  - compounds: 95.599.173
  - bioassays: 285.628
  - triplets: 3.338.304.378

<http://cloud3b.cerit-sc.cz:8080/chemweb/>

Query

Result

Details

Properties

About



any

owl:Class

### Query Wizard ...

Search compounds

Method: Substructure search

Query type: MOL/SDF file

Query: Marvin 12270700542D

13	13	0	0	0	0	999	V2000		
-0.7145	-0.4125	0.0000	C	0	0	0	0	0	0
0.0000	-0.8250	0.0000	C	0	0	0	0	0	0
-0.7145	0.4125	0.0000	C	0	0	0	0	0	0
0.7145	0.4125	0.0000	C	0	0	0	0	0	0
0.0000	0.8250	0.0000	C	0	0	0	0	0	0
0.7145	-0.4125	0.0000	C	0	0	0	0	0	0
1.4289	0.8250	0.0000	C	0	0	0	0	0	0

Upload file

Result limit:

Tautomers: yes

Search bioassays

Keywords: 'cancer\*' or 'carcino\*' or 'tumor\*'

Statuses:  active  probe  
 inactive  unspecified  
 inconclusive

Results

Selections:  compound  bioassay  
 score  status

Order by: (none)

Finish Cancel

- Class
- Class

range: [rdfs:Class](#) (Class)

[owl:unionOf](#) (unionOf)  
domain: [rdfs:Class](#) (Class)  
range: [rdf:List](#) (List)

[owl:intersectionOf](#) (intersectionOf)  
domain: [rdfs:Class](#) (Class)  
range: [rdf:List](#) (List)

[owl:oneOf](#) (oneOf)

## Query

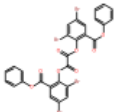
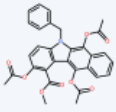
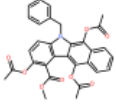
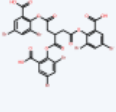
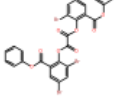
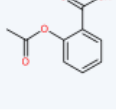
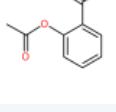
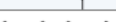


```

1 SELECT DISTINCT
2   ?COMPOUND ?BIOASSAY
3 WHERE
4 {
5   ?COMPOUND
6   orchem:substructureSearch
7   h [
8     orchem:query '''
9     Marvin
10    12270700542D
11    13 13 0 0 0
12    0          999 V2000
13    -0.7145   -0.4125
14    0.0000 C  0 0 0 0
15    0 0 0 0 0 0 0 0
16    0.0000   -0.8250
17    0.0000 C  0 0 0 0
18    0 0 0 0 0 0 0 0
19    -0.7145   0.4125
20    0.0000 C  0 0 0 0
21    0 0 0 0 0 0 0 0
22    0.7145    0.4125
23    0.0000 C  0 0 0 0
24    0 0 0 0 0 0 0 0
25    0.0000    0.8250
26    0.0000 C  0 0 0 0
27    0 0 0 0 0 0 0 0
28    0.7145   -0.4125
29    0.0000 C  0 0 0 0
30    0 0 0 0 0 0 0 0
31    1.4289    0.8250
32    0.0000 C  0 0 0 0
33    0 0 0 0 0 0 0 0
34    2.1433    0.4126
35    0.0000 0  0 0 0 0
36    0 0 0 0 0 0 0 0
37    1.4289    1.6501
38    0.0000 0  0 0 0 0
39    0 0 0 0 0 0 0 0
40    -0.0002   1.6500
41    0.0000 0  0 0 0 0
42    0 0 0 0 0 0 0 0
43    -0.7145   2.0625
44    0.0000 C  0 0 0 0
45    0 0 0 0 0 0 0 0
46    -1.4290   1.6500
47    0.0000 C  0 0 0 0
48    0 0 0 0 0 0 0 0

```

## Result

COMPOUND	BIOASSAY
 <b>CID383326</b>	<b>AID1</b> NCI human tumor cell line growth inhibition assay. Data for the NCI-H23 Non-Small Cell Lung cell line
 <b>CID404315</b>	<b>AID13</b> NCI human tumor cell line growth inhibition assay. Data for the HOP-92 Non-Small Cell Lung cell line
 <b>CID404315</b>	<b>AID15</b> NCI human tumor cell line growth inhibition assay. Data for the NCI-H522 Non-Small Cell Lung cell line
 <b>CID375451</b>	<b>AID51</b> NCI human tumor cell line growth inhibition assay. Data for the XF 498 Central Nervous System cell line
 <b>CID383326</b>	<b>AID125</b> NCI human tumor cell line growth inhibition assay. Data for the HL-60(TB) Leukemia cell line
 <b>CID2244</b>	<b>AID162014</b> Inhibitory concentration in DMSO with purified ovine Prostaglandin G/H synthase 1 (COX-1).
 <b>CID2244</b>	<b>AID612350</b> Inhibition of ovine COX1 at 50 uM incubated for 10 mins by enzyme immunoassay
 <b>CID639952</b>	Inhibition of TNF-alpha-induced

## Details Properties About

owl:Class

**Class**  
owl:Class

**Superclass Hierarchy**  
owl:Class (Class)  
rdfs:Class (Class)  
rdfs:Resource (Resource)

**Subclass Hierarchy**  
owl:Class (Class)  
owl:Restriction (Restriction)

**Properties**

- owl:complementOf (complementOf)  
domain: owl:Class (Class)  
range: owl:Class (Class)
- owl:disjointWith (disjointWith)  
domain: owl:Class (Class)  
range: owl:Class (Class)
- owl:disjointUnionOf (disjointUnionOf)  
domain: owl:Class (Class)  
range: rdf:List (List)
- owl:hasKey (hasKey)  
domain: owl:Class (Class)  
range: rdf:List (List)
- owl:equivalentClass (equivalentClass)  
domain: rdfs:Class (Class)  
range: rdfs:Class (Class)
- owl:unionOf (unionOf)  
domain: rdfs:Class (Class)  
range: rdf:List (List)
- owl:intersectionOf (intersectionOf)  
domain: rdfs:Class (Class)  
range: rdf:List (List)
- owl:oneOf (oneOf)



## Query

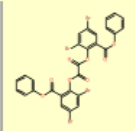
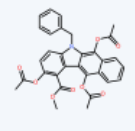
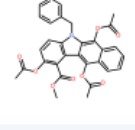
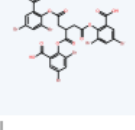
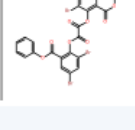
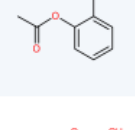
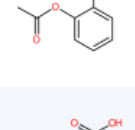


```

1 SELECT DISTINCT
2 ?COMPOUND ?BIOASSAY
3 WHERE
4 {
5   ?COMPOUND
6   orchem:substructureSearch
7   h [
8     orchem:query '''
9     Marvin
10    12270700542D
11    13 13 0 0 0
12    0          999 V2000
13    -0.7145   -0.4125
14    0.0000 C  0 0 0 0 0
15    0 0 0 0 0 0 0 0 0
16    0.0000   -0.8250
17    0.0000 C  0 0 0 0 0
18    0 0 0 0 0 0 0 0 0
19    -0.7145   0.4125
20    0.0000 C  0 0 0 0 0
21    0 0 0 0 0 0 0 0 0
22    0.7145    0.4125
23    0.0000 C  0 0 0 0 0
24    0 0 0 0 0 0 0 0 0
25    0.0000    0.8250
26    0.0000 C  0 0 0 0 0
27    0 0 0 0 0 0 0 0 0
28    0.7145   -0.4125
29    0.0000 C  0 0 0 0 0
30    0 0 0 0 0 0 0 0 0
31    1.4289    0.8250
32    0.0000 C  0 0 0 0 0
33    0 0 0 0 0 0 0 0 0
34    2.1433    0.4126
35    0.0000 0  0 0 0 0 0
36    0 0 0 0 0 0 0 0 0
37    1.4289    1.6501
38    0.0000 0  0 0 0 0 0
39    0 0 0 0 0 0 0 0 0
40    -0.0002   1.6500
41    0.0000 0  0 0 0 0 0
42    0 0 0 0 0 0 0 0 0
43    -0.7145   2.0625
44    0.0000 C  0 0 0 0 0
45    0 0 0 0 0 0 0 0 0
46    -1.4290   1.6500
47    0.0000 C  0 0 0 0 0
48    0 0 0 0 0 0 0 0 0

```

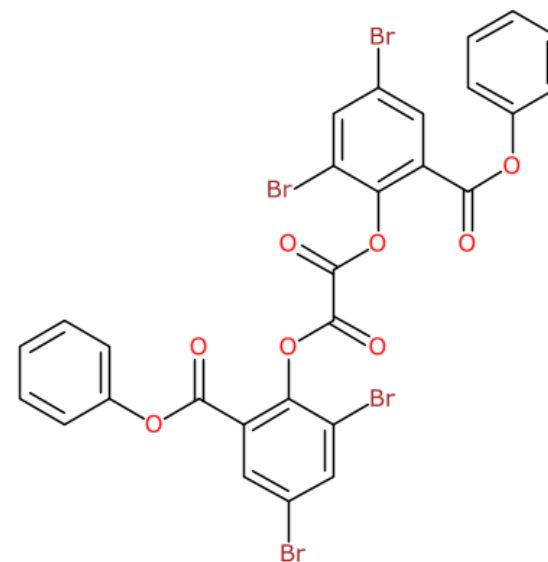
## Result

COMPOUND	BIOASSAY
 <b>CID383326</b>	<b>AID1</b> NCI human tumor cell line growth inhibition assay. Data for the NCI-H23 Non-Small Cell Lung cell line
 <b>CID404315</b>	<b>AID13</b> NCI human tumor cell line growth inhibition assay. Data for the HOP-92 Non-Small Cell Lung cell line
 <b>CID404315</b>	<b>AID15</b> NCI human tumor cell line growth inhibition assay. Data for the NCI-H522 Non-Small Cell Lung cell line
 <b>CID375451</b>	<b>AID51</b> NCI human tumor cell line growth inhibition assay. Data for the XF 498 Central Nervous System cell line
 <b>CID383326</b>	<b>AID125</b> NCI human tumor cell line growth inhibition assay. Data for the HL-60(TB) Leukemia cell line
 <b>CID2244</b>	<b>AID162014</b> Inhibitory concentration in DMSO with purified ovine Prostaglandin G/H synthase 1 (COX-1).
 <b>CID2244</b>	<b>AID612350</b> Inhibition of ovine COX1 at 50 uM incubated for 10 mins by enzyme immunoassay
<b>CID639952</b>	Inhibition of TNF-alpha-induced

## Details Properties About

compound:CID383326

## compound:CID383326 [\(open in pubchem\)](#)



### Synonyms

nsc-671435  
 bis(2,4-dibromo-6-phenoxy-carbonyl-phenyl) oxalate  
 ZINC95551258  
 AC1L8L5J  
 nsc671435  
 CTK516211  
 bis(2,4-dibromo-6-phenoxy-carbonylphenyl) oxalate  
 bis(2,4-dibromo-6-(phenoxy-carbonyl)phenyl) oxalate  
 NCI60\_025283  
 ChEMBL1985534

## Query

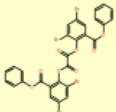
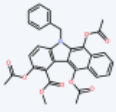
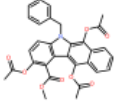
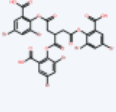
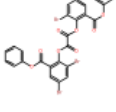
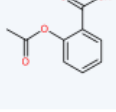
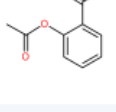
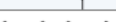


```

1 SELECT DISTINCT
2 ?COMPOUND ?BIOASSAY
3 WHERE
4 {
5   ?COMPOUND
6   orchem:substructureSearch
7   h [
8     orchem:query '''
9     Marvin
10    12270700542D
11    13 13 0 0 0
12    0          999 V2000
13    -0.7145   -0.4125
14    0.0000 C  0 0 0 0
15    0 0 0 0 0 0 0 0
16    0.0000   -0.8250
17    0.0000 C  0 0 0 0
18    0 0 0 0 0 0 0 0
19    -0.7145   0.4125
20    0.0000 C  0 0 0 0
21    0 0 0 0 0 0 0 0
22    0.7145    0.4125
23    0.0000 C  0 0 0 0
24    0 0 0 0 0 0 0 0
25    0.0000    0.8250
26    0.0000 C  0 0 0 0
27    0 0 0 0 0 0 0 0
28    0.7145   -0.4125
29    0.0000 C  0 0 0 0
30    0 0 0 0 0 0 0 0
31    1.4289    0.8250
32    0.0000 C  0 0 0 0
33    0 0 0 0 0 0 0 0
34    2.1433    0.4126
35    0.0000 0  0 0 0 0
36    0 0 0 0 0 0 0 0
37    1.4289    1.6501
38    0.0000 0  0 0 0 0
39    0 0 0 0 0 0 0 0
40    -0.0002   1.6500
41    0.0000 0  0 0 0 0
42    0 0 0 0 0 0 0 0
43    -0.7145   2.0625
44    0.0000 C  0 0 0 0
45    0 0 0 0 0 0 0 0
46    -1.4290   1.6500
47    0.0000 C  0 0 0 0
48    0 0 0 0 0 0 0 0

```

## Result

COMPOUND	BIOASSAY
 <b>CID383326</b>	<b>AID1</b> NCI human tumor cell line growth inhibition assay. Data for the NCI-H23 Non-Small Cell Lung cell line
 <b>CID404315</b>	<b>AID13</b> NCI human tumor cell line growth inhibition assay. Data for the HOP-92 Non-Small Cell Lung cell line
 <b>CID404315</b>	<b>AID15</b> NCI human tumor cell line growth inhibition assay. Data for the NCI-H522 Non-Small Cell Lung cell line
 <b>CID375451</b>	<b>AID51</b> NCI human tumor cell line growth inhibition assay. Data for the XF 498 Central Nervous System cell line
 <b>CID383326</b>	<b>AID125</b> NCI human tumor cell line growth inhibition assay. Data for the HL-60(TB) Leukemia cell line
 <b>CID2244</b>	<b>AID162014</b> Inhibitory concentration in DMSO with purified ovine Prostaglandin G/H synthase 1 (COX-1).
 <b>CID2244</b>	<b>AID612350</b> Inhibition of ovine COX1 at 50 uM incubated for 10 mins by enzyme immunoassay
 <b>AID639952</b>	Inhibition of TNF-alpha-induced

## Details Properties About

bioassay:AID1

### bioassay:AID1 [\(open in pubchem\)](#)

NCI human tumor cell line growth inhibition assay. Data for the NCI-H23 Non-Small Cell Lung cell line

#### Description

Growth inhibition of the NCI-H23 human Non-Small Cell Lung tumor cell line is measured as a screen for anti-cancer activity. Cells are grown in 96 well plates and exposed to the test compound for 48 hours. Compounds are tested at 5 different concentrations and three endpoints are estimated from this dose response curve: GI50, concentration required for 50% inhibition of growth, TGI, the concentration requires for complete inhibition of growth, and LC50, the concentration required for 50% reduction in cell number. These estimates are done by simple linear interpolation between the concentrations that surround the appropriate level. If a compound doesn't cause inhibition to the appropriate level, the endpoint is set to the highest concentration tested.

#### Comment

These data are a subset of the data from the NCI human tumor cell line screen. Compounds are identified by the NCI NSC number. In the NCI numbering system, NCI-H23 is panel number 1, cell number 1. Basically compounds with LogGI50 (unit M) less than -6 were considered as active. Activity score was based on increasing values of -LogGI50.

# ELIXIR CZ Activity in Interoperability and Data Management

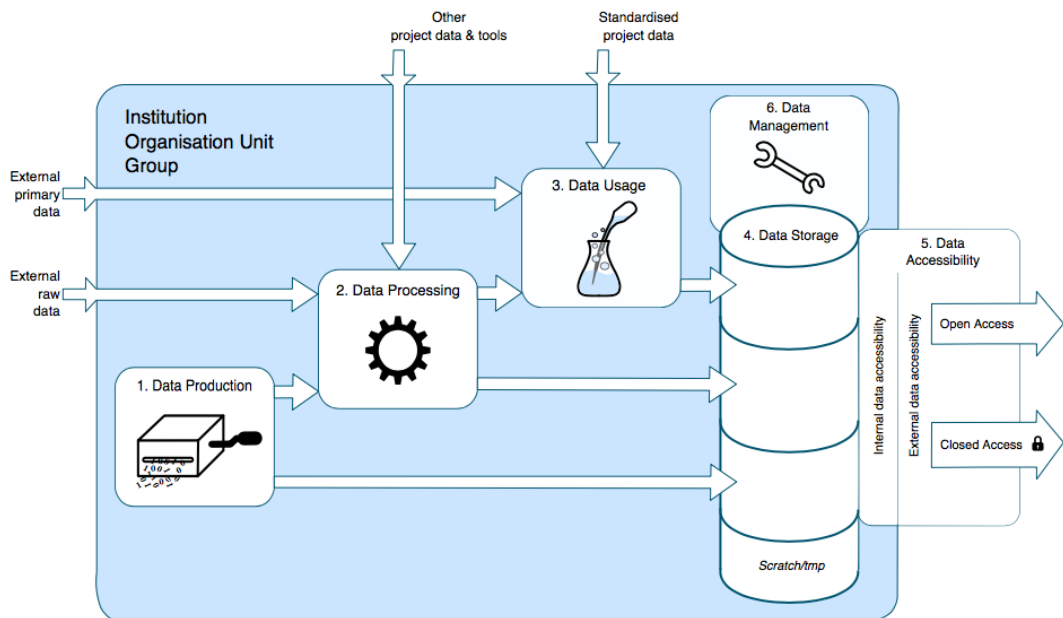
## The Data Management Planning Portal



To be able to submit your data, please apply for a respondent key @ robert.pergl@fit.cvut.cz

Vision Action steps Lifecycle Data Roles Managerial Questionnaire Technical Questionnaire

The questionnaire is founded on the data lifecycle. Click each stage to learn the details:



<http://elixir-cz.fairdata.solutions>



## Context:

- „Data Management Plans (DMPs) are a **key element** of good data management.“
- Horizon 2020: „A **DMP is required** for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt out are **still encouraged to submit a DMP** on a voluntary basis.“
- European Open Science Cloud: Recommendations of the High Level Expert Group



European  
Commission

Horizon 2020  
European Union funding  
for Research & Innovation



Data Stewardship  $\supset$  Data Management

## Data Stewardship

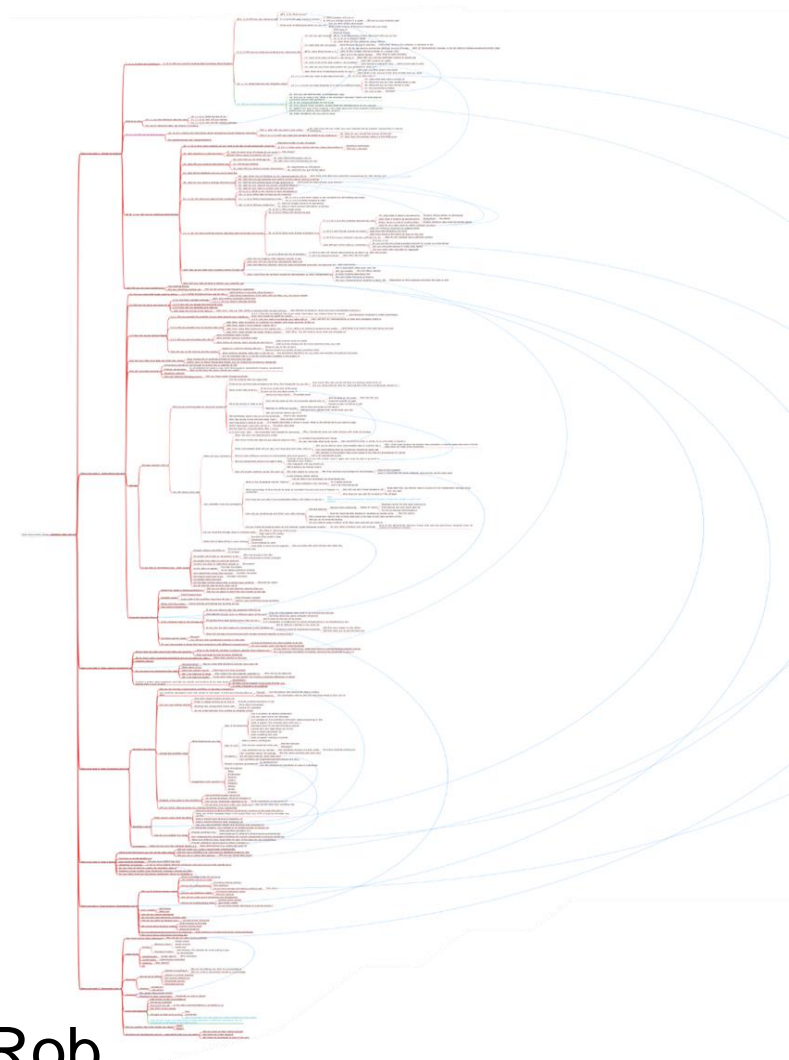
Plan

Management

Long-Term  
Preservation

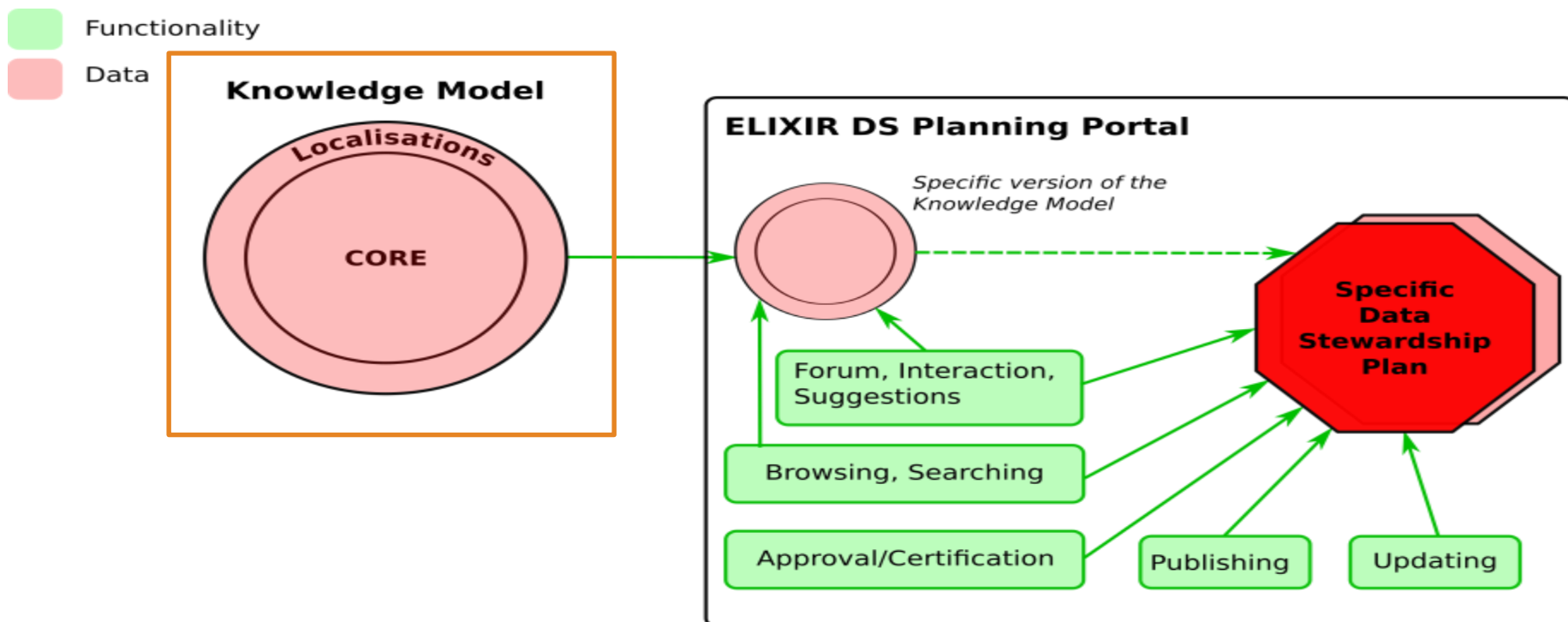
# Challenge

- Scientists struggle to prepare and maintain good DS plans.
- The knowledge is vast.
- Their business is not to acquire it.
- They need effective help.



Data Stewardship Mind Map by Rob Hooft (DTL, ELIXIR-NL)

# Data Stewardship Planning Portal Architecture



<https://f1000research.com/posters/5-2420>



# DS Wizard

The first piece: A prototype for interactive browsing of DS-KM

- Guiding through all the questions

The screenshot shows the DS Wizard web application in a browser window. The browser tab is titled "DS Wizard" and the address bar shows "dmp.fairdata.solutions". The page header includes the "elixir" logo and the text "DATA STEWARDSHIP WIZARD v0.2, KM: JAN 19, 2017". Below the header is a navigation menu with buttons for "About", "Design of experiment" (which is highlighted), "Data design and planning", "Data Capture/Measurement", "Data processing and curation", "Data integration", "Data interpretation", and "Information and insight".

The main content area features an "Introduction" paragraph: "Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others." To the right of this text is a yellow callout box with a lightbulb icon containing the text: "Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?"

The form contains several questions with radio button options:

- Question 1: "Is there any pre-existing data?" with options "No" and "Yes". The "Yes" option is selected.
- Question 2: "Will you be using any pre-existing data (including other people's data)?" with options "No" and "Yes". The "Yes" option is selected.
- Section: "What reference data will you use? List all the items below." with a sub-question "Item" and an input field.
- Question 3: "Do you know where and how is it available?" with options "No" and "Yes". The "No" option is selected.
- Question 4: "Do you know in what format the reference data is available?" with options "I can directly use it" and "I need to convert it before using". The "I can directly use it" option is selected.

<https://dsw.fairdata.solutions>

# DS Wizard

The wizard is intended as the basis for helping with creating data management plans.

The screenshot shows the DS Wizard web application in a browser window. The browser tab is titled "DS Wizard" and the address bar shows "dmp.fairdata.solutions". The page header includes the "elixir" logo and the text "DATA STEWARDSHIP WIZARD v0.2, KM: JAN 19, 2017". Below the header is a navigation menu with buttons for "About", "Design of experiment" (which is highlighted), "Data design and planning", "Data Capture/Measurement", "Data processing and curation", "Data integration", "Data interpretation", and "Information and insight".

The main content area features an "Introduction" paragraph: "Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others." Below this is a question: "Is there any pre-existing data?" with radio buttons for "No" and "Yes", where "Yes" is selected.

A yellow callout box on the right contains a lightbulb icon and the text: "Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?"

Below the first question is another question: "Will you be using any pre-existing data (including other people's data)?" with radio buttons for "No" and "Yes", where "Yes" is selected. This question is followed by a section titled "What reference data will you use? List all the items below." which contains a table with columns for "Item", "Do you know where and how is it available?", and "Do you know in what format the reference data is available?". The "Item" column has a text input field. The "Do you know where and how is it available?" column has radio buttons for "No" and "Yes", with "No" selected. The "Do you know in what format the reference data is available?" column has radio buttons for "I can directly use it" and "I need to convert it before using", with "I can directly use it" selected.

<https://dsw.fairdata.solutions>

# DS Wizard

Future: all the remaining functionality 😊

The screenshot shows the DS Wizard web application in a browser window. The browser tab is titled "DS Wizard" and the address bar shows "dmp.fairdata.solutions". The page header includes the elixir logo and the text "DATA STEWARDSHIP WIZARD v0.2, KM: JAN 19, 2017". Below the header is a navigation menu with buttons for "About", "Design of experiment" (which is highlighted), "Data design and planning", "Data Capture/Measurement", "Data processing and curation", "Data integration", "Data interpretation", and "Information and insight".

The main content area features an "Introduction" paragraph: "Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others."

Below the introduction is a question: "Is there any pre-existing data?" with radio button options for "No" (checked) and "Yes".

Another question is: "Will you be using any pre-existing data (including other people's data)?" with radio button options for "No" and "Yes" (checked).

Below this is a section titled "What reference data will you use? List all the items below." containing a table with columns for "Item", "Do you know where and how is it available?", and "Do you know in what format the reference data is available?".

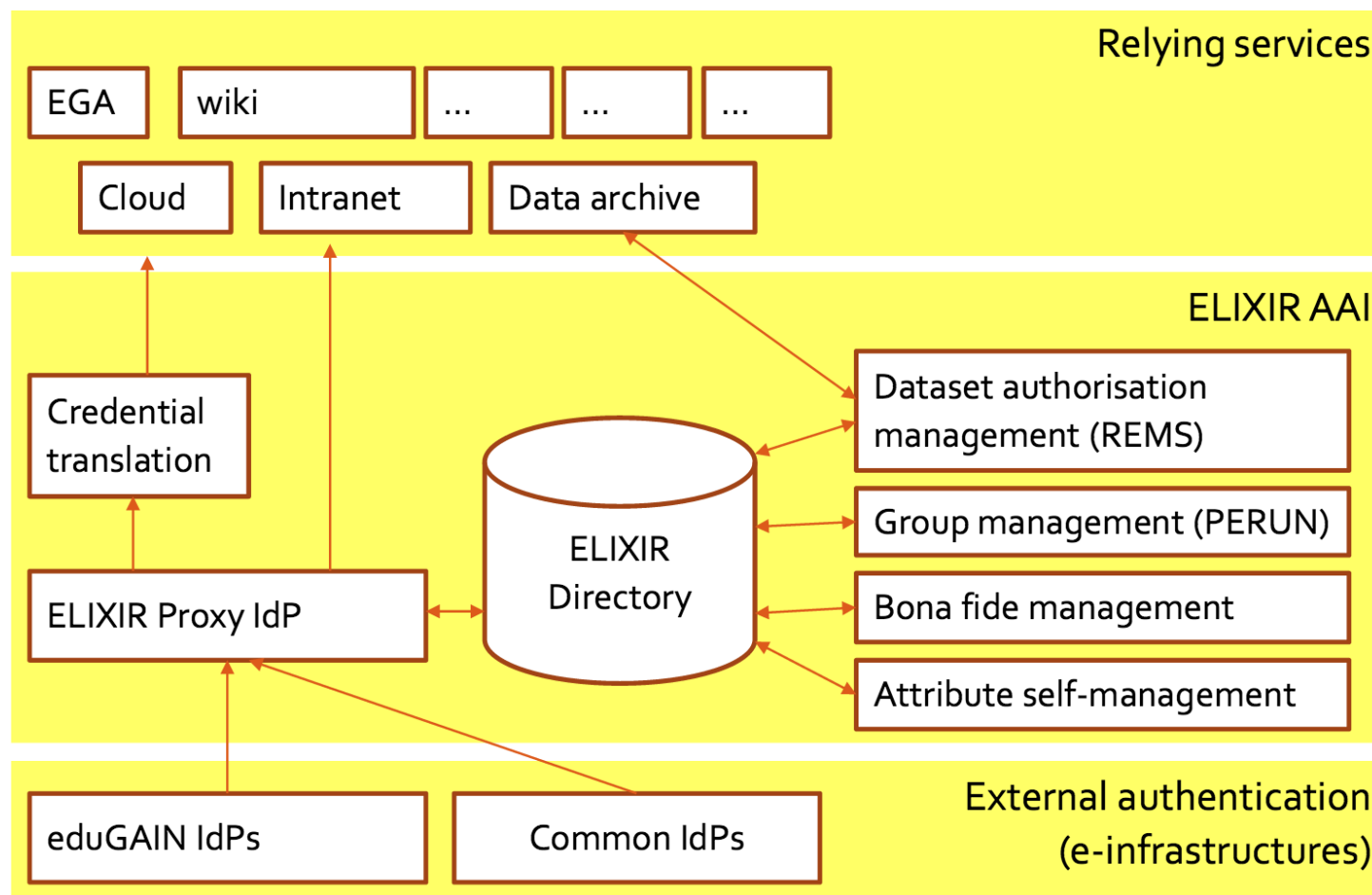
Item	Do you know where and how is it available?	Do you know in what format the reference data is available?
	<input type="radio"/> No <input type="radio"/> Yes	<input type="radio"/> I can directly use it <input type="radio"/> I need to convert it before using

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

<https://dsw.fairdata.solutions>



# ELIXIR AAI (Authorisation and Authentication Infrastructure) - contribution of CZ to Compute Platform



<https://www.elixir-europe.org/services/compute/aa1>

# ELIXIR CZ Outlook

- ELIXIR CZ strategic direction definition - inline with the 2019 – 2023 ELIXIR scientific programme and national community priorities.
  - Potential new areas for development:
    - Proteomics
    - Medical community engagement (Rare diseases)
- Continue addressing challenges related to infrastructure capacity, data management, F.A.I.R. principles.
- Keep delivering training to the community.



# ELIXIR Node Capacity Building Programme

## CZ engagement in ELIXIR Excelebrate WP10

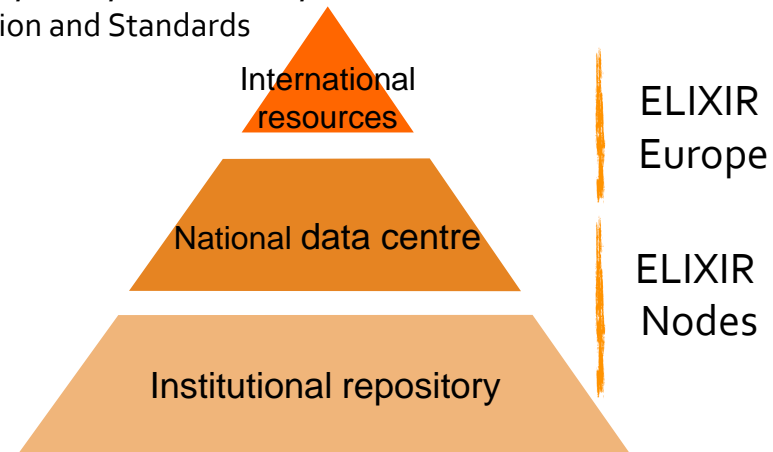
- Create a ELIXIR wide *Data Nodes Network* supporting data submitters and users
- Create an ELIXIR-wide *Genome Annotation Network* to scale and sustain data quality
  - Workshops on genome annotation
- Capacity building programme in newly formed ELIXIR Nodes
  - Accessing EU Structural Funds



# Data Nodes Network

- Aim to **establish guidelines and good practices** to facilitate efficient data collection into core data resources and making them FAIRified.
- A first step is to facilitate **data management/ stewardship**, where common European efforts have resulted in creation of **templates for data management plans**.

Usage , Value, Permanence,  
Curation and Standards



Data stewardship action team

To be able to submit your data, please apply for a respondent key @ robert.pergl@ft.cvut.cz

Vision Action steps Lifecycle Data Roles Managerial Questionnaire Technical Questionnaire

About 0.General Info 1.Production 2.Processing 3.Usage 4.Storage 5.Accessibility 6.Management 7.Roles Finish

Do you produce raw data?  Yes\*  No

**Type of data**

(Estimated) volume of raw data produced inhouse in 2015

Genomics\*

Volume  MB  GB  TB  PB

Cost for year 2015  thousand EUR

Rough estimation of FTEs + investments + consumables

Proteomics\*

Volume  MB  GB  TB  PB

Cost for year 2015  thousand EUR

Rough estimation of FTEs + investments + consumables

Others\*

Raw data production volume  TB

Raw data production cost  thousand EUR

- We have initiated the creation of a data nodes network, **applied on the use case** based on local EGA (cf. WP9). Workshop and Designathon arranged in 2016.
- Plan for architectural alignment is validated and simple API interoperability test has been conducted.
- Need for **local Genome-phenome data archives** is growing every year because of exponential data generation of human data that in many cases cannot leave the country unless access is granted via data access committees.

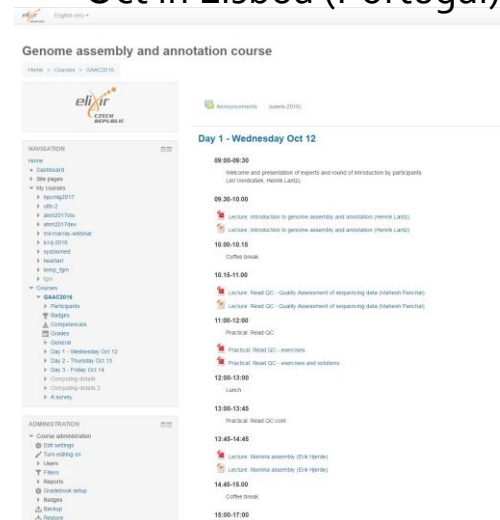




# Genome Assembly and Annotation

- **Advanced Workshops**
  - Preparatory in Uppsala Oct 2015
  - Barcelona April 2016
  - The Hague Sept 2016
- **Best Practices Document**
  - A printed publication and an online wiki
  - Publication to be submitted to Faculty1000 mid-2017.
- **European “Advanced Help Desk”**
  - Pilot project where the Swedish Node is providing high-level expert support to a project from the Czech Republic with genome assembly and annotation.

- **Advanced Genome Assembly and Annotation Courses**
  - First hosted at IOCB in Prague
  - Teachers from 4 nodes (SE, NO, FR, BE)
  - 24 participants, all from Czech Republic
  - 3 days; 2 days assembly + 1 day annotation
  - Lectures and practical computer exercises
  - eLearning platform (EeLP) was used to host presentations, instructions, and survey – provided by ELIXIR-SI
  - **Two more courses are planned in 2017**
    - June in Ljubljana (Slovenia)
    - Oct in Lisboa (Portugal)



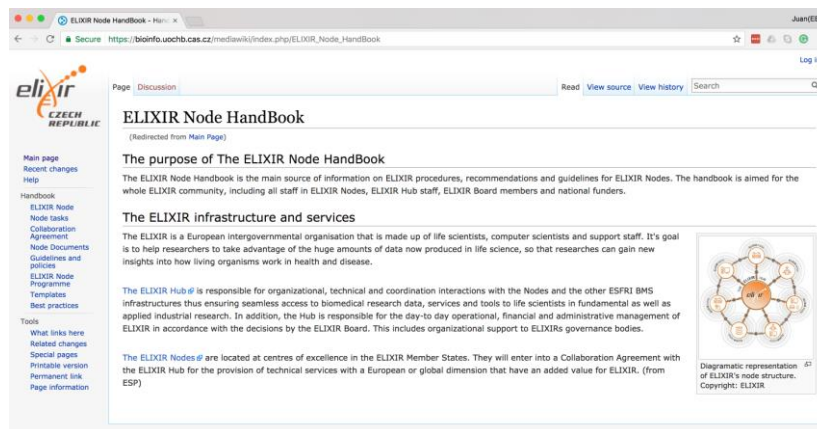
The screenshot shows the ELIXIR eLearning platform interface for the 'Genome assembly and annotation course'. The page is titled 'Genome assembly and annotation course' and features a navigation menu on the left with options like 'Home', 'Dashboard', 'My classes', and 'Courses'. The main content area displays a schedule for 'Day 1 - Wednesday Oct 12' with a timeline of activities from 09:00-09:30 to 15:00-17:00. The activities include a welcome presentation, lectures on genome assembly and annotation, a coffee break, and practical exercises on quality assessment of sequencing data.



# ELIXIR Node Handbook

- Provide **management knowledge transfer among Nodes** to create a set of well-balanced, well-functioning and compatible Nodes.
- **Capacity Building and Node development workshop** in November 2015 in Brussels
- We also started to work on **identification of expertise and competences in ELIXIR Nodes**.
- One of the practical outcomes is the **Node Handbook** – an effective tool for solution of major problems on Node structure and governance by strengthening the local node.
- The handbook is presented in wiki form, it is **interlinked with Handbook of Operations** provided by ELIXIR Hub and it also contains best practices from individual nodes.

[https://www.elixir-czech.cz/node-handbook/index.php/ELIXIR\\_Node\\_HandBook](https://www.elixir-czech.cz/node-handbook/index.php/ELIXIR_Node_HandBook)



The screenshot shows a web browser window displaying the ELIXIR Node Handbook page. The page title is "ELIXIR Node HandBook" and it is redirected from the main page. The content includes sections on the purpose of the handbook, ELIXIR infrastructure and services, and the ELIXIR Hub's role. A diagrammatic representation of the ELIXIR node structure is shown on the right.

# EU Structural Funds

- ELIXIR EU Structural Funds Group
- Exchanges best-practice in accessing EU Structural Funds for Research Infrastructures
  - Workshops in Brno, 8–9 October 2015
  - ELIXIR All-hands meeting in Barcelona, March 2016
  - ELIXIR All-hands meeting in Rome, 22 March 2017
- **2 Grants Awarded in CZ, €3.900.000**
- **1 Proposal Submitted Interreg Cross Border Program Belgium-Netherlands**
- **Calls open in Greece, Estonia and Slovenia**





## Using structural funds for funding ELIXIR nodes

# H2020 and SF: Different fundamentals

	FP7/H2020	Structural Funds
<b>Main focus</b>	Scientific excellence	Economic and regional development
<b>EC department responsible</b>	DG Research	DG Regional Policy
<b>Time scale</b>	2 year work programmes	Calls depending on management authorities, n+2/3 rules

Learning new vocabulary, e.g.:

- Structural Funds (SF) now European Structural and Investment Funds (ESIF)
- Regional Innovation Strategy, now Smart Specialisation Strategy

# H2020 and SF: Different fundamentals

	FP7/H2020	Structural Funds
<b>Management/who decides</b>	Centralised by EC	Decentralised, over 450 Operational Programmes
<b>Funding priorities</b>	Given by EC	General priorities agreed between EC and Member states/regions, details defined by regions
<b>Rules</b>	Single set of rules for all aspects	General rules and principles common, detailed by managing authorities
<b>Co-funding</b>	100 % funding for direct costs + contribution to indirect costs (generally)	General co-funding requirement, i.e. EU pays 85 %, the rest by member state or beneficiary

# Seal of Excellence

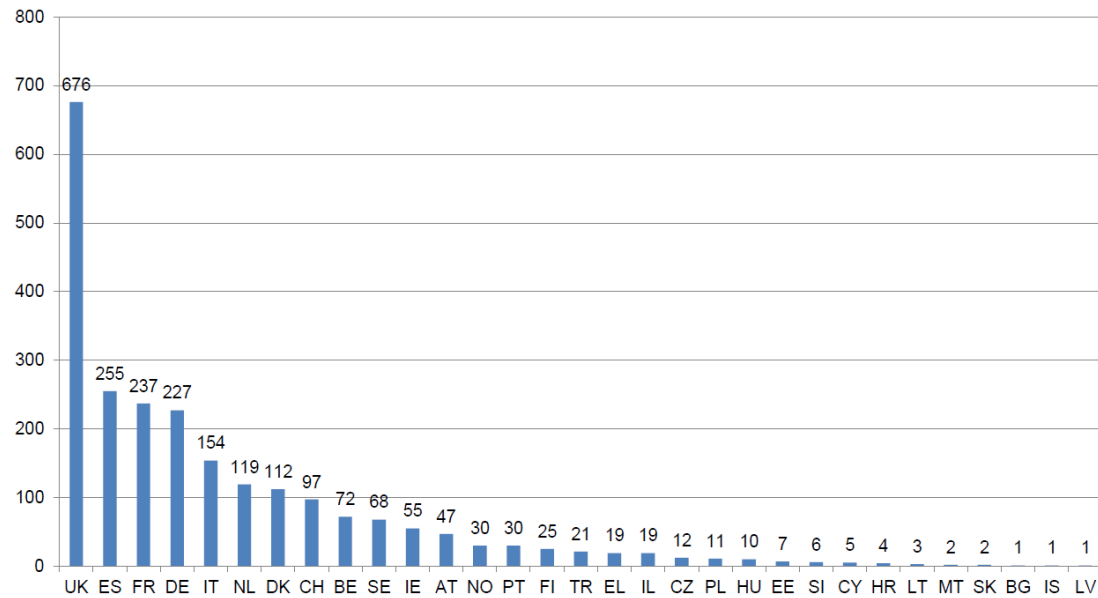
*Initiative of European Commission. The Seal is given to excellent projects (above threshold) but due to limited H2020 budget not funded. Projects shall be funded on national level (national funding, structural funds)*

- **Since March 2016** awarded to Marie-Curie (MSCA) – from 2016 call 2 328 applicants who scored 85% or more and weren't on the funded list - Commitment from CY+CZ, Interest from IT, RO, SI, HR, PL

ELIXIR action: Lobby at the national level – include Seal of Excellence in ESIF or national funding

## MSCA IF Call 2016

2 328 Seals of excellence to be awarded





# Smart Specialisation Strategy

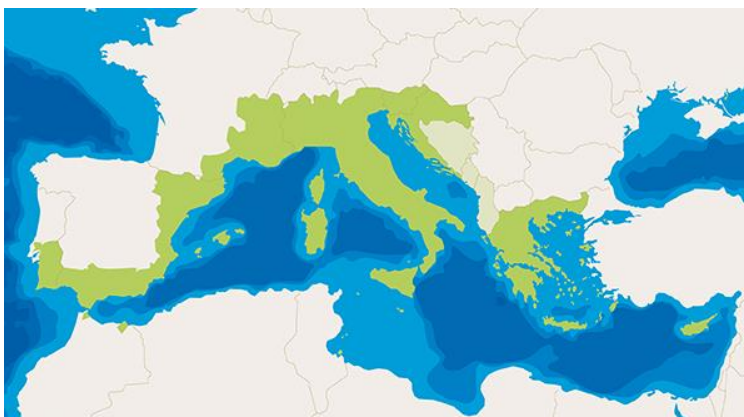
Strategy every region has to develop in order to be able to draw structural funds for research and innovation. From a business perspective, not a research perspective.

Also relates to ELIXIR areas of action:

- Biotechnologies
- ICT
- Agribusiness

# Interreg programmes

- Possibility to submit projects of cooperation between several ELIXIR nodes in a given region
- Regional programmes, e.g. MED, South-East Europe
- Cross border, e.g. Greece-Cyprus, Greece-Italy, ...



# Czech situation 1

Relevant operational programme: Research, Development and Education

3 relevant calls for research (with major funding):

- Excellent research teams – support to establish new teams around experienced researchers from abroad
- Excellent research – broadly open to all R+D institutions and scientific fields
- Research infrastructures – dedicated to research infrastructures

# Czech situation 2

## Research infrastructures call

- Dedicated call to research infrastructures
- Open to all infrastructures listed on Czech roadmap of research infrastructures
- Main aim: cover investment costs 2016-2019, soft funding for „in-house“ research included
- Synergic to support of operational costs from CZ state budget
- One project per infrastructure
- ELIXIR CZ got awarded a project: cca 3 mil EUR

# Czech situation 2

## Research infrastructures call

- ELIXIR CZ got awarded a project – February 2017:
  - Budget: cca 3 mil EUR - 2 mil for hardware, 1 mil for in-house research
  - Duration: 5/2017-4/2020
- Participation of selected ELIXIR CZ partners, based on consortium consensus

## RIAT-CZ project

- Cooperation between CEITEC and Vienna Biocenter Core Facilities
- Funding: Interreg CZ-AT, budget: 1,3 mil EUR
- Timeline: 9/2016-8/2019
- Specific aims:
  - *Improved knowledge on managing RI and expertise in providing services*
  - *Innovation of services through joint use of cross-border complementary equipment*
  - *Enabling access to RI for external users*

<http://riat.ceitec.cz>



## 2021-2027 structural funds period

- European Commission presented proposal for EU budget for this period, incl. structural funds
- For structural funds about 5 % budget decrease is envisaged
- Proposals for legislation on structural funds expected in June 2018
- Time to say on national level that structural funds shall be invested also to research



**Thank you**



[www.elixir-czech.cz](http://www.elixir-czech.cz)