

**Report of the DTL Partner Advisory Committee meeting
November 23rd 2017, 14.00 - 17.00 h**

Meeting location: DTL, Jaarbeurs Innovation Mile, Jaarbeursplein 6, 3521 AL Utrecht

Attendees: Formal representatives of DTL Partners and representation of DTL Board

Meeting documents: <https://www.dtls.nl/about/organisation/pac/documents-pac-meetings/>

1. Opening by Ronald Stolk

a. Short introduction of new partners (Annex 1)

DTL counts 55 partners per January 1st, 2018.

De Research Manager en RSRCH BV joined as new DTL partners per July 1st 2017.

AMC, VUMC and Equalis will join as new DTL partners per January 1st 2018.

b. Report of the 5th meeting of the DTL Partner Advisory Committee (Annex 2)

There are no comments on the report of the previous PAC meeting.

2. DTL update by Ruben Kok

Ruben Kok gives an outline of DTL activities. He talks about the current status of Health-RI and GO FAIR, courses, workshops, focus meeting, hackathons, BYOD's and bi-monthly programmers meetings and international community workshops arranged by the DTL team. Mateusz Kuzak has recently joined the DTL team. He is an expert in software and data carpentry.

A snapshot of 2017 shows the Enabling Technologies Hotel call 2017 which received 160 proposals. A 5th call is being planned. The budgets NWO reserves are for PPP partnerships. DTL will try to secure an option for fully academic projects as well and envisions to move NWO and ZonMw to make this approach part of regular calls. If PAC members have suggestions for future calls, please get in touch with Merlijn van Rijswijk.

Assisted by KPMG, a business plan for Health-RI has been prepared, zooming in on a national infrastructure. A Health-RI Conference will be held on December 8th.

FAIR Data in Open Science has been supported by the G7 in Turin (Sept 2017). October 26, the European Commission published a declaration on the European Open Science Cloud (EOSC), with a key role for the FAIR approach.

Our Ministry of Education, Culture and Science (OCW) decided to fund the GO FAIR **International** Support and Coordination Office in Leiden, led by Barend Mons. Germany and France have decided to support the GO FAIR approach as well, with teams of experts in both countries. The Netherlands still needs a national GO FAIR office. Barend Mons suggests this should be embedded in DTL for the life sciences fields.

August DTL Board strategy session

During a DTL Board strategy session in August, the Board brainstormed about the future course of DTL. The Board sees the data challenges as core to the DTL activities, with Open Science and societal challenges as strong drivers. The FAIR data approach is an important topic, with implementations such as the Personal Health Train (PHT) and Farm Data Train (FDT). Important to work out – and document – tangible examples that demonstrate the added value of these approaches. DTL should drive the improvement of data quality as well, where there is an important role for the associated hotel groups as well as for the training programme. The collaboration with BioSB is strengthened. ELIXIR-NL should connect the other infrastructures across the life sciences at the data-level.

In 2018, the implementation of FAIR data trains will be a key topic. A manifesto is prepared to establish a PHT Public Private Partnership. If you would like to join this PPP, please sign the manifesto!

Ruben announces the second DTL Conference to be organized in 2018, centered around challenges and solutions across the data-intensive life sciences. Partners will be involved.

ELIXIR-NL update by Jaap Heringa

Currently, 21 countries have joined ELIXIR. There is a flexible hub and nodes structure, with a node in each country. Jaap Heringa is Head of Node ELIXIR-NL, with Rob Hooft as Technical Coordinator and Celia van Gelder as the Training Coordinator. ELIXIR-NL is represented in the ELIXIR Board by Bea Pauw (NWO) and Ruben Kok (co-chair of the ELIXIR Board).

ELIXIR's activities are divided into five areas called 'Platforms'. These are Data, Tools, Interoperability, Compute and Training. Regarding the Data and Tools platforms, Jaap Heringa asks the partners to please report tools or data repositories that should be added.

3. International and industrial perspectives on data in Open Science

- a. The Swiss perspective on data infrastructure and ELIXIR Core Data Resources
By dr Christine Durinx of SIB Swiss Institute of Bioinformatics

Introduction

Ronald Stolk introduces Dr Christine Durinx. She has a Pharmacy degree and a PhD in Pharmaceutical Sciences from the University of Antwerp, Belgium. She is the Associate Director of SIB Swiss Institute of Bioinformatics since 2014. At SIB, Christine is responsible for the Corporate Communications and Training departments, the Director's Office, and the Legal & Technology Transfer Office. Christine is also co-lead of the ELIXIR Data platform. In this context, she works on indicators for life science databases and funding models to improve the long-term sustainability of the data science infrastructure. Before joining SIB, Christine worked in the pharmaceutical industry for over 10 years.

History and current status of SIB

Christine Durinx describes the laborious period Swissprot (database and tools) has gone through in the late '90 due to financial challenges. SIB was created in 1998 to work on long

term data management with a healthy infrastructure as a mission. In 2017, SIB counts 65 groups, 800 scientists, 19 partner institutions and 90 ongoing collaborations with industry, and a diverse portfolio of activities. Currently, SIB offers access to 150 databases and software tools, which are all openly available through the SIB platform.

To assure sustainability, 12 core resources selected by the board are funded long term, two of which have been selected as ELIXIR core data resources: 1) UniProtKB / Swiss-Prot and 2) String database (protein-protein interactions). SIB has a state of the art bioinformatics infrastructure but is also providing research support and training in the fields of life and health sciences. SIB also organises courses and scientific conferences to bring researchers together. The Swiss Ministry of Science directly funds SIB.

The ELIXIR Data Platform

SIB as the Swiss node has been very active in ELIXIR to build a sustainable infrastructure for biological information across Europe. SIB sees ELIXIR grow into the European expertise network, jointly offering data services and access to ELIXIR data resources. Today, the life sciences strongly rely on such data resources. Long-term sustainability of these resources needs more attention, which was underlined by the outcome of a recent inventory: 66% of the databases have no funding for over 1 year. A large part of all databases disappears.

An international approach is being prepared by ELIXIR to ensure the long-term sustainability of essential (core) data resources. To be able to do this, objective indicators are needed to identify these essential databases, databases with a fundamental importance to the LS community that have a generic value and high-level usage and service levels.

A carefully chosen set of indicators has been captured that support the decision making: 1] (service) quality, 2] legal/funding/infra, 3] community, 4] scientific focus, and 5] societal impact (e.g. translated in patents). Indicators can also be mapped to the FAIR Principles.

A first selection of ELIXIR Core Data Resources has been published in Summer 2017. In course of time, new data resources may be added to this list after international review.

b. Pharma goes FAIR

By Dr. Herman van Vlijmen of Janssen Pharmaceuticals

Introduction

Ronald Stolk introduces Dr. Herman van Vlijmen. He graduated with a Master's degree in Bio- Pharmaceutical Sciences at Leiden University in The Netherlands and a PhD degree in Physical Chemistry at Harvard University. He worked 9 years at the biotech company Biogen in the Boston area, ultimately as Senior Scientist, in the computational design of small molecule drugs and protein therapeutics. In 2005 he joined Tibotec, a Johnson and Johnson company focusing on infectious diseases, as Director of Computational Drug Design. He is now Head of Computational Chemistry in the Discovery Sciences organization at Janssen, Pharmaceutical companies of Johnson & Johnson, located in Belgium. Since 2008 he is also

Adjunct Professor of Computational Drug Discovery at Leiden University. Herman has more than 60 publications and is inventor on 8 patents.

Importance of FAIR

Many databases are still in silos with poor accessibility and/or findability data and an absent or incomplete use of nomenclature standards. The amount and diversity of scientific data is growing fast, and the most valuable analysis involves data from different domains and technologies. Machine learning and data mining requires unambiguous computer readable data.

FAIR and Janssen

Scientists at Janssen rarely used all data they have access to, it was difficult to access multiple databases and there is a lack of awareness of databases. There is little experience with definition of cross domain analysis. Data from multiple domains and sources (private, public, commercial) is needed for the best possible analysis and here the FAIR principles are very important. Currently, internal Janssen databases do talk to each other. Internal efforts in discovery are done by exploring the use of Euretos' Knowledge Platforms. Herman pleases to build an ecosystem of FAIR data resources.

Open Phacts (OPS)

The Open Phacts project, funded by the Innovative medicines Initiative (IMI), was an early example of the FAIR approach and relevant databases were selected based on complex scientific questions. It was clear that a lot of data will never be published anywhere else, are only sitting in patents and/or are not searchable. The OPS consortium funded project to extract gene/disease information across approximately 4 million patents, which yielded 260 million annotations. A broad set of use cases can be addressed using such a linked data system.

IMI2 - 12th call for proposals

A new IMI call was launched in July 2017 for the FAIRification of IMI and EFPIA data, and candidate consortia are currently being evaluated. The project will likely start in Q3 2018 with a duration of 3 years and a budget of a max. of EUR 4M€. Pharma partners are Janssen, Astra Zeneca, Bayer, Boehringer Ingelheim, Eli Lilly, Glaxo-SK and Novartis.

A summary of the FAIRification proposal:

- Select data sets and databases from finished and ongoing IMI projects, based on scientific value of making this data accessible and interoperable and based on the complexity of making the data available.
- Select databases at individual EFPIA companies where the selection is based on the value for companies and where there is consolidation to a limited set of data domains.
- FAIRify these data sets to enable the sustainable use of the data in answering research questions.

Conclusions

Pharma has a strong interest and need for implementing the FAIR data principles. Expertise is usually lagging due to complex legacy data systems and limited IT resources. Scientific acceptance for FAIRification in pharma requires strong use case examples. Collaboration with academia and the SME community catalyzes expertise and acceptance of FAIR data. Metadata is crucial to make data FAIR, which requires a change in the business processes and buy in of the management layer. People are obsessed with their own project but don't care about the data afterwards, and that has to change. Every new dataset adds value to the total. Where to start or stop making data more FAIR is to decide time by time, sometimes generating a new dataset is faster and cheaper.

4. DTL Exchange: Open Science in perspective for DTL partners **Break-out discussions, Plenary feedback and actions**

Group 1 - What are the key take home messages for your organisation based on the presentations

- Join forces, we all have similar problems
- Everybody goes FAIR
- It's not only about databases. Adding semantics. Also about tools, improve tooling.
- Publishing data - making data accessible
- Adding metadata takes a lot of effort
- We need tools to automatically create metadata - help with semantic modelling
- Reproducibility crisis: experiments are hard to reproduce
- There is a strong need for showcases - convincing examples
- End user/end use case - drive investment and need for FAIRification
- SIB does a lot of manual curation and annotation of data.
- Give credits for contributing data

Threats:

- Confusing people, burdening people with extra work
- Public understanding stands in the way.

Group 2 - What are the opportunities & Threats?

- **What are the implications of Open Science for your organization**
- **What resources /services do you have on offer**
- **What threats do you perceive?**

Main focus on threats and their implication for the DTL strategy and DTL mission:
Role for DTL to advocate next generation solutions and connect experts. Scouting of solutions and examples and pushing them into market. Implementation must be done by DTL partners. DTL in Health RI: focus on resources of data for research.

Reflecting on presentations: competence centres were seen as important. DTL can assist in creating a network of in the house competence centres.

Group 3 - What are the implications for the DTL work programme 2018 and beyond

Low hanging fruit:

- Train, educate PhD students on how to deal with data
- Start outside the Health care data (patient information)
- Use criteria of core data resources and see if we can get the databases adhere to those criteria (treat it as a spectrum)
- Carrot and stick: training, credits, move funding bodies to accept funding of data stewardship, etc (carrots) vs requirements of university boards or funders (sticks)
- Metadata: FAIRification should be done straight away (FAIR at the source). Waiting makes it harder.
- Accepted by society? We need showcases, PR. Drive investment and need for FAIRification.
- Annotating: done by a large group – to obtain scalability.
- Take care of your core resources!! Data stewardship!

Group 4 –

Take-aways:

- Shocked by fleeting nature of data
- FAIR is not necessarily Open
- Internal data -> metadata may be open, “findable”!
- Every data type requires specific approach
- Making your own data FAIR/open relies on other connected data to be FAIR > e.g. climate environment

Threats:

- Standards are not complete, risk in choosing.
- Diverging; global acceptance / standards / implementations
- Difficult to interpret data without the accompanying paper
 - Opportunity: user annotations (even in company setting)
- Changing mindset takes time

Opportunity:

- Extracting and adding value to data becomes more easy

2018:

- Work on mindset
- Supporting tooling required
- Database solutions?
- Further awareness of data openness value for citations
- FAIR data certifications/stamps
- FAIR company certifications/stamps
- Threat: sustainability through project money

5. Closure and drinks