

**Report of the 8th DTL Partner Advisory Committee meeting
June 15th 2018, 13.00 - 17.00 h**

13.00 - 14.30 h, pre-PAC meeting: building a FAIR service providers consortium

14.45 - 17.00 h, PAC meeting plus drinks

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/>

Agenda

1. Pre-PAC meeting: building a FAIR service providers consortium [13.00 – 14.30]

Open to all DTL partners, special invite for DTL's SME partners!

Ruben Kok opens the meeting, welcomes the 25+ participants and explains the context: The DTL community has been instrumental in developing and advocating the FAIR data approach which now has been globally adopted. The time has now come to implement FAIR in many ways. This session is about the development of a FAIR Service Providers Consortium. DTL wants to make sure its partners are well informed and especially welcomes partners with an interest in providing FAIR services to join this consortium.

Albert Mons, provides a historic overview of the FAIR developments, starting at DTL, and summarises the vision on the global "Internet of FAIR Data and Services", as meanwhile guided by the GO FAIR organization. The GO FAIR approach aligns fully with the planned implementation of the European Open Science Cloud, which will be strongly FAIR-based. Science funders globally will also make data stewardship fundable, and an estimated 5% of research budgets overall is expected to be reserved for this.

GO FAIR Office

The GO FAIR (Global Open FAIR) International Support and Coordination Office is headed in Leiden and lead by Germany, France and The Netherlands and with support of 13 European member states. The main tasks of the GO FAIR Office is to guide the Implementation of the Internet of FAIR Data and Services with a focus on Europe at first, support of Implementation Networks and FAIR certification. The FAIR Data Principles are *not* a standard but guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets.

Information about the FAIR Data METRICS, a framework to measure FAIRness, can be found at <http://Fairmetrics.org>.

In the industry and the academic world, there is a growing need for programmes to educate a vast number (possibly over 1.5 million) data stewardship professionals to fill up the current gap of expertise. Data steward is an emerging profession that guards the way we take care of our data in the long run. Large industry such as Bayer are now preparing for this

transformation within their company, which they see as a means to significantly reduce loss of R&D investments due to loss of/lack of reusable data (currently 50% or higher).

In this emerging and huge market, 800 million for the academic market in the Netherlands alone and 62 billion in Europe, there are currently only a few small-size companies with sufficient knowledge to train these FAIR data stewards, make data FAIR, create FAIR tooling, etc.

[*Wouter Franke* (Zorginstituut NL) resonates what has been said and predicts a great need for professional FAIR services also in the health care field. For instance, there is a clear lack of professional tooling. Current tools are academic-made, not yet suitable for large-scale use in the care field.]

GO FAIR Foundation

February 2018, the Stichting GO FAIR (GO FAIR Foundation) has been established. The foundation takes up tasks the GO FAIR office cannot do, such as developing certification and training.

FAIR Service Providers Consortium

In alignment with GO FAIR, a consortium has been started in which parties are brought together to speak with one voice; the FAIR Service Providers Consortium. Current partners are: the GO FAIR Foundation and several DTL company partners (Micelio, Euretos, Phortos, The Hyve and Castor), plus the companies Mobiquity and Purple Polar Bear. The consortium is expected to grow and DTL partners are explicitly invited to join. What binds together the consortium partners is that all organisations are there to support the GO FAIR implementation approach. There are rules of engagement in which parties agree to adhere to the FAIR Principles and prevent vendor lock in. A vision document, a readiness program, certification rules, these are all in development at the GO FAIR Foundation.

Consortium participants

Kees van Bochove (The Hyve) has joined the consortium because the market starts to ask for FAIR services. The danger is that parties develop their own solutions. It is important to align and join up, speak with one voice. The FAIR metrics are therefore crucial. From a commercial perspective, the NL has some steps ahead, but we need to work hard to keep that position.

Annick van Arkel of Purple Polar Bear (PPB) has built a first FAIR Metrics Evaluator, together with colleagues. It has been used at Bio-IT World recently, and was one of the first applications. She has been amazed by the FAIR concept and PPB is interested in developing concrete tooling around it.

Gerbrand Ruiter (Mobiquity) is asked to tell more about the company and the use of FAIR. Mobiquity helps their customers to develop applications in health care. A lot of data is collected through these applications. There is a strong belief that the GO FAIR approach is a smart step forward to standardizing this data. Medical devices have to apply to the GDPR rules, and the Medical Device Regulation forces to check post-market if the application

delivers what it was intended. This legislation started last year but will be controlled by 2020. The use of apps can change, FAIR data can be very useful in this regard.

Albert Mons opens the discussion and invites the audience to respond.

Walter Pirovano (Baseclear) is experiencing difficulties submitting data to databases. It also is hard to reproduce what has been published before. He asks Albert Mons how GO FAIR will be working together with Scientific journals and large institutions. Albert Mons explains that journals typically join an Implementation Network (IN) and institutions can join the OPEDAS (*Other PEople's Data and Services*) network. Now all organizations are doing their own thing, the purpose of the IN's is to de-silo them.

Olaf Lodbrök of Elsevier points out that GO FAIR should also collect standards that help to adhere to the FAIR Principles. Standardization will be paramount for fields such as healthcare and to guide companies in their development cycles.

Marco Roos (LUMC) is the initiator of the GO FAIR Rare Disease Implementation Network (IN). He explains that his group will “fight the fights” in rare disease network because that is not the task of the GO FAIR office. While setting up an IN, GO FAIR forces you to think about questions which you cannot find answers to now and how to test it and/or set it up. A strategy, endorsement and decisions on how to deal with old data is also crucial. E.g. it might be more logic to start with new datasets instead of making the old ones FAIR.

Albert Mons confirms it is a community approach in which the GO FAIR Office is organizing those ideas and giving them back to the community. The results of IN's are open for the community. Manifesto's are published on the GO FAIR website.

An anecdote by *Albert Mons*: At Utrecht University it took 4 people three months to create a semantic model for datasets. A BYOD event was organized, and the same dataset was used. After 3 days, the model and triple store was ready and gave better results. That's a nice anecdote of the advantages of going FAIR and working together.

Several attendees express their interest in joining the FAIR Service Providers Consortium. *Ruben Kok* thanks Albert Mons for organizing the session and closes the pre-PAC meeting.

Anyone interested in taking part in the service providers consortium, please contact Albert. Email: Albert.mons@phortosconsultants.com, Skype: albertmons.

[short break]

2. Opening Partner Advisory Committee meeting

Ronald Stolk, the current PAC Chair, was unfortunately not able to join today. He has asked Jaap Heringa to serve as Chair of today's PAC meeting. Jaap Heringa opens the meeting.

Introduction of new partners

New partner organisations are shortly introduced by their representatives.

AMC/VUmc: Prof. Arjen Brussaard explains that the Amsterdam Medical Centre (AMC) together with the VUmc has recently merged into the Amsterdam UMC. The organization produces an enormous amount of data and would like to make use of the expertise basis in the DTL community to improve on data solutions, taking into account that some datasets are not easily shared in this field.

Equalis: Piet Stam introduces Equalis as a strategy company. Equalis connects hospitals and health care providers with insurance companies, and develops business cases for insurance companies and health providers together based upon data analysis to demonstrate advantages, of certain measures in the healthcare domain.

Aridhia is UK company that develops digital research environments (DRE) for individuals, teams and institutions. In the NL the Aridhia products are part of the DRE by RadboudUMC.

LabServant: Han Bakker explains that LabServant distills data from chaotic research processes and helps researchers to organize their lab process and capture standardized procedures.

Jaap Heringa warmly welcomes the new partners to DTL.
He informs the PAC that *RIVM* has confirmed its partnership today as of the 1st of July.

Report of the 7th meeting of the DTL Partner Advisory Committee (Annex 2)

There were no comments on the report.

3. Short DTL update by Ruben Kok

Ruben Kok also welcomes all PAC members to this 7th DTL Partner Advisory Committee meeting and gives a short update on developments in DTL.

Enabling Technology Hotel call 2017

The recent 4th ETH call has yielded 60 selected projects. Some figures: 1,8M EUR (total 4,5M for all calls), success rate of 38%, private partners in 10-15% of the awarded projects. The majority of projects are on novel data generation, and several data analytics proposals were rejected. This may reflect a bias in the selection process, which is being discussed with ZonMw. DTL has been asked to organize FAIR data training for selected projects/hotels to increase the awareness of the FAIR data principles, make participating projects more FAIR, and improve quality of data management plans. The next call is expected early 2019. For

2018/2019 ZonMw envisages co-funding for ETH projects in other NWO programs. For the longer term the DTL SAC has advised to make the ETH approach an integral part of regular funding schemes. This is being discussed with ZonMw and NWO.

Collaboration BioSB

The extensive collaboration with BioSB has been further formalized. BioSB binds the bioinformatics and systems biology communities. Its education programme is complementary to what is done in DTL, also in relation to ELIXIR-NL, which made teaming up between DTL and BioSB a logical next step. See later presentations by Celia van Gelder and Natal van Riel. DTL now provides organizational support to BioSB, both for its yearly conference as well as for its course programme

How to fund a national data services infrastructure

The NL-Roadmap 2017 for Large Scale Research Infrastructure has selected BSL3, NEMI, NPEC and X-Omics in the Life Sciences field. Not selected are: ELIXIR-NL, MCCA, MRUM, NIEBA-fw, NL-Bioimaging-AM, UNLOCK.

The strategic boundaries and conditions formulated by the Commission of Van Duijn mentions making available research data according to the FAIR principles. Also, facilities within the life sciences are advised to join DTL and coordinate their data approach with ELIXIR-NL. Unfortunately, ELIXIR-NL itself has not been funded, which causes big challenges on how to setup and finance this data service infrastructure. A bridge funding of 500 k for 2 years has been granted for 1st coordination steps. There is a risk of falling outside the regular infrastructure funding schemes as ELIXIR's data services infrastructure is not so much about hardware but largely 'peopleware', software and standards). ELIXIR-NL started the dialogue with NWO. Karel Luyben, Chair of the DTL Board, has been appointed Coordinator Open Science.

DTL Communities @ Work 2018, Oct. 2nd - [Save the date!](#)

All PAC members are invited to join the DTL Communities@Work 2018 conference, that will be held in Utrecht (Jaarbeurs) on October 2, 2018. The meeting will host interactive workshops organized by partner communities on topics such as: Personal Health Train, Farm Data Train, Modelling the Human Metabolism, Machine Learning in Life Sciences, European Million Genome project, FAIR data management, etc. With this conference, DTL offers an open platform to interconnect with other professionals and improve your research and innovation by building shared data solutions together. There is a reduced entrance fee for DTL partners.

Ruben calls for representatives of DTL company partners to advise in regard to this conference. Please contact Ruben Kok (ruben.kok@dtls.nl) if you would like to play a role.

4. Information management and data stewardship

Information Management at Zorginstituut Nederland (ZiN) by Wouter Franke

The main task of the National Health Care Institute (ZiN) is to ensure a basic package of insurance in the NL. ZiN advises the Ministry of Health what to include in the basic package of health care. Maintaining the health care quality is also done by ZiN through the Health Care Quality institute. Overall, the effectiveness of care is important, including financing, quality and accessibility.

For describing standards and managing data sets, ZiN already works together with people from the GO FAIR office with the end goal of describing standards in a FAIR way, using ontologies and FAIR data points and based on Linked Data.

FAIR gives big data a boost and can connect data sets, although they don't talk the same language. It is quite technical but that is logical; data should not only be human readable but also computer readable. There is a lack of professional tooling to work in a FAIR way, most tools are developed within an academic environment or are not yet finalised. For ZiN, becoming FAIR was not that hard because in long term healthcare there is already a lot of effort to make data standardized.

ZiN is very positive about the Personal Health Train approach as part of the Internet of FAIR Data and Services (IFDS). It offers opportunities for Big Data applications. The PHT makes "Privacy by Design" possible where you can analyze data on different locations but only see the conclusion and not the data itself. Confidence (e.g. about the origin of the train) is necessary and need additional agreements. ZiN rather talks about FAIR Data Services instead of the PHT as metaphor.

The way ZiN looks at the FAIR data future: ZiN collects and owns research data, quality data and other data sets and has an interest in making these data sets FAIR. ZiN is also director of Information Exchange and sees FAIR as the longterm healthcare modernizer. Data providers are requested to and informed on how to deliver their data in a FAIR way.

Capacity building in data stewardship: DTL's approach Rob Hooft (DTL)

Data management involves data during a project while data stewardship involved data for eternity. Data stewardship starts before the actual start of the project and involved any operation on digital data in that project which is much more than the 5% that is often mentioned.

DTL's definition of Data Stewardship: *Responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data."*

Getting to the R (reusable) in FAIR, you need to go through F (findable), A (accessible) and I (interoperable) first. Interoperability is a challenge and needs to grow. To make data re-usable, you use more metadata than you 'normally' would in your project. Therefore, planning is crucial.

A design of experiment *wizard* has been co-developed by DTL to make people aware of what it entails to treat data well, and if potentially useful, make them available for re-use by others. To support this process, we need a low barrier approach, like a 'data desk' within an organization, involving the IT department, library and experts in different fields. DTL connects a number of such "data desk" initiatives across the partner organisations.

In the NFU data4LS programme, the UMC's are already coordinating on Data Desks, standards and training. This network brings up all problems data stewards face. Members come together ca. 5 times a year to discuss best practices, etc. DTL also has set up a Data Stewards Interest Group. Rob invites the DTL partner organisations to join this interest group and/or mailing list to prevent re-inventing the wheel.

PAC Discussion

Ruben Kok marks that the presentations illustrate that FAIR-based data stewardship and analysis are extending far beyond academic LS research, and he is glad to see that FAIR starts to be picked up in the care field as well. Implementing this will be important in every day-to-day data practice.

Arjen Brussaard (VUmc) is interested to learn how to execute the FAIR principles, how far can you go in accessibility. At VUmc a lot of data is not open source and there may be IP involved. *Michel Dumontier* (Maastricht University; one of the authors of the FAIR Principles and FAIR metrics paper) explains that FAIR is not equivalent to Open: data experts would like to know that certain data exist, but this does not necessarily mean that these datasets should be open to the world. An access protocol should be developed, in which the accessibility of data to other parties is explained. *Albert Mons* (Phortos) confirms the message conveyed by Rob Hooft that it is important to have growing teams of data stewards on board of your organisation, people who are data experts and know how to deal with making data FAIR.

5. Training and education in data-intensive life sciences

DTL Learning: training initiatives in perspective: DTL's approach *Celia van Gelder* (DTL)

DTL Learning brings together Dutch and international experts to provide data related training for the life science community in order to face today's data challenges. Modern LS calls for advanced data and technological expertise.

Within the DTL Learning community, training expertise is connected, bridges are built between disciplines and application domains, data-related training gaps are identified, courses and trainings are developed together where best practices and expertise is shared. Furthermore, DTL Learning is aligned with both ELIXIR-NL and the ELIXIR-EU programmes.

Interest groups and other community activities (newsletters, mailing groups, etc.) have been setup to accommodate the community building within the different groups, e.g. on FAIR Data Stewardship, Galaxy or NGS. FAIR data treatment and training of researchers and data experts are core to DTL Learning.

ELIXIR-NL training and DTL Learning go hand in hand. The ELIXIR Training portal TeSS gives access to training materials and events and aggregates information from 48 content providers across Europe. All Dutch courses posted in the DTL portal are automatically put into TeSS for international exposure.

Furthermore, there is a collaboration with HBO through Domain Applied Science (DAS), which unites all 14 HBO programs in the applied sciences. DTL is currently exploring membership of HBO Schools with DTL.

As mentioned before by Ruben Kok, DTL and BioSB joined forces by establishing a strong nationwide public private community of LS experts in data stewardship, infrastructure and analytics.

BioSB: Advanced training in bioinformatics and computational systems biology *Natal van Riel* (TU/e; v-chair BioSB Educom)

The BioSB Research School brings together the expert groups in Dutch bioinformatics and computational/systems biology and offers courses with the BioSB 'seal of approval'. A set of advanced 'core' courses is actively developed and maintained, and additional 'specific' courses are evaluated and advertised (www.biosb.nl).

Core courses are in-depth 5-day courses, organised once per 2 years including assignments after the course, possibly resulting in 3EC points. Specific courses can have all kinds of formats.

The course overview for core courses: *Managing and integrating information* (LUMC/LIACS), *Optimization techniques* (UvA, WUR), *Algorithms for biological networks* (WUR, UMCU), *Comparative genomics* (Radboudumc, CMBI), *Pattern recognition* (AMC, TUD, NKI), *Protein structures* (CMBI, NKI), *Quantitative and predictive modelling* (WUR, TU/e).

Specific courses 2018: *Transcriptomics & single cell RNA-seq* (VIB, ELIXIR-NL, ELIXIR-BE), *NGS data analysis* (EMC, LUMC), *Genome re-sequencing* (EMC, Ned. Vereniging voor Klinische Chemie) and *Life science with industry* (Lorentz Centre).

Course fees are set at 400-600-900 euro (excl. 21 VAT) for 5-day course for PhD students, academics and industrial attendees. For a shorter or longer course, the fee is adapted accordingly. BioSB membership of 25 euro/year is included in the fee to attend the yearly conference.

Additional courses that are currently considered are: Metabolomics/Proteomics, High-dimensional statistics, Algorithms for genomics, Systems medicine, FAIR data and semantic interoperability, Constraint-based modelling and Metagenomics.

Ruben Kok wants to know how we can keep track of the needs for training and education in the field, in academia and industry. *Celia van Gelder*: we have several activities ongoing to measure these needs, for example as part of a recently Interreg-funded project directed to

capacity building in the LSH sector. *Natal van Riel*: BioSB wants to keep up with the needs of the community. It therefore welcomes the input of the DTL partners.

Michel Dumontier (UM) mentions it would be useful to look at incorporating FAIR in all courses and he would be happy to discuss this further.

Merlijn van Rijswijk (Netherlands Metabolomics Centre): Part of the X-Omics Roadmap funding will be used to setup new courses for cross-omics data analysis, including genomics and metabolomics, etc. so this offers a good opportunity to work together. Celia van Gelder is already in contact with Alain van Gool (Radboudumc) in that respect.

6. Other PAC issues, conclusions

Ruben Kok: Ronald Stolk has been PAC chair for 5 years and will step down. There is a vacancy for a Chair as well as a Vice Chair of the Partner Advisory Committee. Next to chairing the PAC meetings, the Chair is invited to join the DTL Board meetings to be informed and to inform the Board at PAC level. Ruben Kok will send a request for candidates, on behalf of the DTL Board.

7. Closure and drinks

Jaap Heringa closes the PAC meeting at 17u00 and invites everyone for drinks.