Report 12th meeting of the DTL Partner Advisory Committee

May 28th 2020: 14.00 – 16.30 h, Online meeting.

Participants:
Peter Hilbers (TU/e; Chair)
Astrid Engelen-Visser (IOS Press)
Dimitris Koureas (Naturalis)
Cees Hof (DANS-KNAW)
Patrick Kemmeren (Princes Maxima Centrum)
Jasmin K. Böhmer on behalf of Jeroen de Ridder (UMC Utrecht, CMM)
Henk-Jan van den Ham on behalf of Jos Lunenberg (ENPICOM B.V.)
Steven van Helden (PPSC)
Maurice Bouwhuis (SURF)
Marielle Gallegos Ruiz (Roche)
Tess Korthout (The Hyve)
Dick de Ridder (WUR)
Han Bakker (Lab Servant)
Karel Luyben (TU Delft)
Renger Jellema (DSM)
Robert Hall (WUR)
Rick van Nuland (Lygature)
Barend Mons (GO FAIR)
Eric de Vries (Digeketen)
Maarten Bakker (Innopay, EZK-Data Sharing Coalition)
Aletta Kraneveld (Utrecht University)
Barbara de Ridder

Edwin Cuppen (Hartwig Medical Foundation)
Hans Niendieker (Ivido)
Inez Joung (RIVM)
Jantine Dirksen (Quaero Systems)
Jeroen Belien (VUmc)
John Schmitz (TU Delft)
Olaf Lodbrock (Elsevier)
Niklas Blomberg (ELIXIR)
Wiro Niessen (ErasmusMC - Health-RI)
Robert Veen (ErasmusMC)
Solon Pissis (CWI)
Andra Waagmeester (Micelio)
Ronald Stolk (UMC & University Groningen)
Harry van Haaften (The Hyve)
Frits van Merode (Maastricht UMC+)
Alain van Gool (Radboudumc)
Michel Dumontier (Maastricht University)

Team DTL: Ruben Kok, Frederike Schmitz, Femke Francissen, Mijke Jetten, Celia van Gelder, Merlijn van Rijswijk, Erik Schultes (GO FAIR/DTL)

Agenda, annexes and slides:

The Agenda, annexes and slides from the meeting can be found here.

Table of Content

1. Short update by Ruben Kok (DTL)

Recapitulating discussions and conclusions from the last PAC meeting

Ruben initiated the meeting by summarizing important discussion points from the last PAC meeting on 29-10-2019, in which partners discussed the term 'Integrated Life Sciences' and the role DTL should take. Integrated Life Sciences, according to DTL's partners is characterized by a) 'being able to re-use and combine data', by b) suitable infrastructure that supports research, by c) sophisticated data analysis, d) professional expertise and learning communities, as well as e) knowledge organisation systems, which facilitate information bundling and dissemination.

WHAT IS INTEGRATED LIFE SCIENCES?

BEING ABLE TO RE-USE & COMBINE DATA
Heterogeneous data sets, data standards & interoperability, remove silo walls, combine retrospective data & analysis

INFRASTRUCTURE:
Facilitating the actual research work through data infrastructure (appropriate datasets, terminologies, identifiers, tools, services)

KNOWLEDGE ORGANISATION SYSTEMS
Information bundling & dissemination, making knowledge accessible and usable

DATA ANALYSIS
Accessible data, software, algorithms, computation

PROFESSIONAL EXPERTISE & LEARNING COMMUNITY
Findability of expertise, interdisciplinary collaborations, knowledge brokering, capacity building, community support

What should DTL as an organisation focus on while advancing Integrated Life Sciences?

In breakout sessions the partners discussed what DTL as an organisation should focus on and concluded that DTL should **form a collective voice** for academic and industry partners, also by connecting to other national and international initiatives. In addition, DTL should also focus on '**raising awareness around data sharing'** and '**providing training to increase data literacy'**. DTL's network recognizes the importance of professionalising data stewardship and DTL's efforts to achieve this.

Several partners have raised the need of **'providing appropriate oversight'** related to resources, facilities, tools, standards and expertise. DTL contributes to this by closely working with Health-RI and ELIXIR to bundle and showcase tools and knowledge.
In active discussions amongst the partners it also became clear that DTL should help its partners to **advance AI and data analytics**. DTL recognizes the need to **align recent developments in data analytics with that of data stewardship and data infrastructure into a comprehensive agenda towards 'Digital Life Sciences'**. Together with DTL's partner network of bioinformaticians and computational life scientists (assembled in the research school BioSB) DTL is in a unique position to help close the gap between professional data handling and data analysis.

As DTL is well **connected to international initiatives and organisations in digital life sciences,** such as ELIXIR, GO FAIR, RDA, GOBLET, EOSC and active in international communities, it can pass on insights and summarize recent developments in the field and communicate this information to its partners. This can close the gap in knowledge that local projects generate, to external initiatives and further advances the field as a whole.

WHAT SHOULD DTL FOCUS ON?

Form a collective voice in digital life sciences

Raise awareness around data sharing and provide training to increase data literacy in LS

Help strengthen international connections

Provide overviews of resources, facilities, tools, standards, expertise & recent advancements

Help partners advance in AI and data analytics
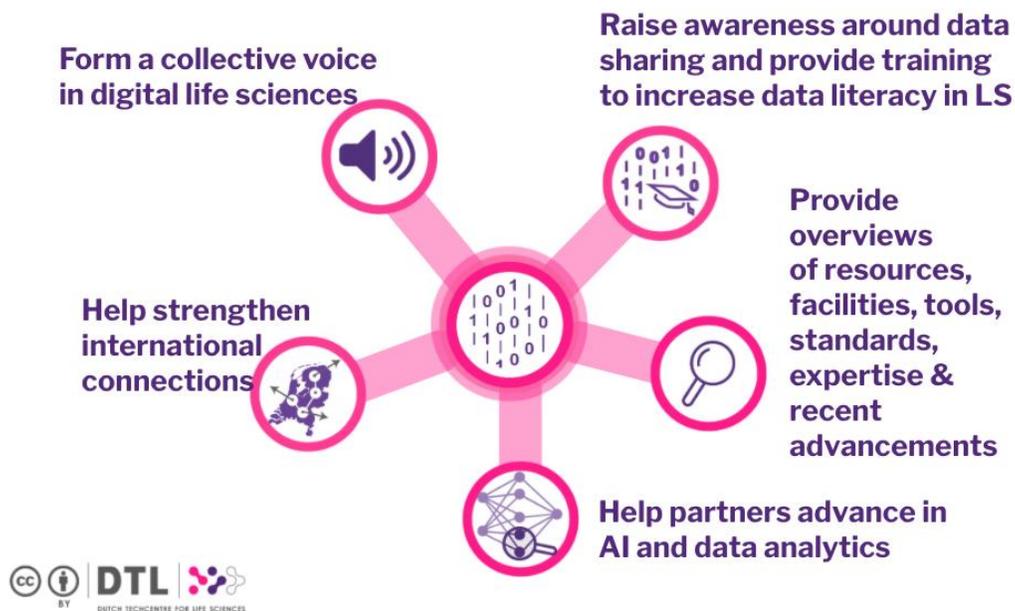
Recent updates from the DTL network/field

Ruben also gave a short presentation on recent developments in the field.

- In the **5th call of the Enabling Technology Hotels (ETH)** 42 projects were awarded with a maximum of 30.000 Euro per project.
- **Health-RI** submitted a proposal to realise a national federated data infrastructure for health research and innovation. As part of a larger consortium of proposals, the "Biotech cluster", Health-RI is requesting funding from the **Hoekstra-Wiebes Innovation Fund (Ministry of Economic Affairs)**. Another proposal for this route of funding was submitted for a **'Data sharing coalition'** in alignment with the FAIR implementation programme of the National Programme Open Science.
- **Building a Data Competence Center (DCC) network:** NWO has opened up the long-expected call for the first round of funding for local DCCs (Deadline 17 Sept 2020). Already, the preparatory process itself has led to the establishment of DCCs in the academic organisations. DTL and ELIXIR-NL, together with Health-RI will make sure to align with these DCCs, to form a network of experts specialized in digital life sciences. The vision would be to establish a thematic DCC for the broader life sciences field.
- **News from ELIXIR**: Several Dutch partner organisations are involved in the emerging ELIXIR communities 'Food & Nutrition'. In addition, ELIXIR-NL launched a national Interoperability Interest Group, which will work closely with the Interoperability platform of ELIXIR, co-led by Chris Evelo of Maastricht University.
- **'Professionalising data stewardship: competences, training and education'** is a follow-up project from previous efforts (LCRDM, ZonMw) and will further advance data stewardship as a profession. This project is part of the **National Platform Open Science (NPOS)**, designated as the NPOS F project. **Mijke Jetten introduced herself as the new DTL's community manager data stewardship**. Mijke is actively involved in the NPOS F project and further aims to advance and expand DTL's Data Stewardship Interest Group.

NPOS F project

**Professionalising data stewardship: competences, training and education**

Background
- Challenges: lack of sufficient, high-quality expertise; lack of consensus on what a data steward is/does; lack of education and certification
- Previous projects: Data stewardship on the map (LCRDM, '19; https://doi.org/10.5281/zenodo.2669149); Towards FAIR data steward as profession for the lifesciences (ZonMw, '19; https://zenodo.org/communities/nl-ds-pd-ls)

Objectives
- National coordination on competences and learning outcomes
- Formal agreement by stakeholders and willingness to implement it
- Enable developers of educational programs to develop curricula

Project
- Work in LCRDM task group (up to August), deliverables and recommendations (September)
- Communication via NPOS website
- More information via DTL's Mijke Jetten & Celia van Gelder
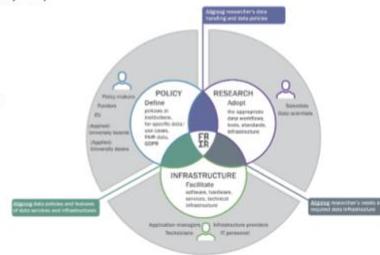
An update on the COVID-19-developments in and around the DTL partnership.
DTL contributes to the overview of covid-19 data initiatives in Health-RI. Erik Schultes introduced himself and his role in the VODAN initiative (GO FAIR initiative) and noticed that data stewardship and FAIR implementation are now being picked up by many organisations, among which funder ZonMw, as essential steps of urgent COVID-19 research projects, and to stimulate rapid data reuse.

The agenda of the PAC was created in such a way that COVID-19 initiatives from DTL partners could be discussed in more detail:

- How FAIR can you go? Modelling COVID-19-induced clinical phenotypes to rationalise drug repurposing. Barend Mons (LUMC, GO FAIR)
- Health-RI response to support national clinical COVID-19 research. Wiro Niessen (ErasmusMC, Health-RI)
- Connecting data, tools and people across Europe: ELIXIR's response to the COVID-19 pandemic. Niklas Blomberg (ELIXIR)

Questions/remarks/ answers raised during the presentation:
- "How is the inventory of resources connected to international initiatives?" Answer: The inventory of resources has two aspects: 1] what are Dutch initiatives on Covid-19, and where do we have close involvement of DTL partners, and 2] what international initiatives are ongoing. For the first we work with Health-RI (overview on Health-RI

website), for the second we point to several international overviews, e.g. of [ELIXIR's response to Covid-19](#).

2. Presentation: How FAIR can you go? Modelling COVID-19-induced clinical phenotypes to rationalise drug repurposing. Barend Mons (LUMC, GO FAIR)

Barend Mons introduced the global [VODAN network](#) of about 150-200 people who are actively involved in creating a FAIR-based Virus Outbreak DAta Network. He mentioned that distributed learning (such as the [Personal Health Train](#)) could assist in the COVID-19 crisis, because not all countries are willing to share personal, confidential data. It is therefore difficult to get access to the data. This by itself is a big challenge. He proposed that FAIR Data Points installed at individual institutional repositories will be used as a machine-interoperable solution, add-on to local infrastructure, to share FAIR metadata across the world and make datasets findable.

Barend further summarized recent COVID-19 related work: publications were checked and their (meta)data was extracted and made FAIR. Often, if not always, the data is not FAIR at all. If data from research would be FAIR at the source we would have had more insights into the pathology of the disease and transmission pattern of the virus by now. He provided an overview of the various COVID-19 data, disease models and related drugs included, highlighting how they are connected, and how the knowledge graphs that result from this, based upon FAIR data, have been used to rationalise repurposing of existing drugs for Covid-19 patients with severe symptoms. He summarized that it took his team about 3 months of diving into the real world patient data, as these were difficult to obtain. He said that if this data would have been available earlier (FAIR at the source) it would have been more easy to perform these analyses based upon already existing knowledge and data. A lot of lifes could have potentially been saved.

Pointing at the Digital Life Sciences approach highlighted by Ruben earlier, Barend proposed to work towards so-called data-rationalisation 'hotels' as part of DTL's Enabling Technology Hotels programme, that combine specialised data generation and data FAIRification (FAIR@source) with data analytics including established knowledge made accessible by infrastructures such as ELIXIR.

3. Presentation: Health-RI response to support national clinical COVID-19 research. Wiro Niessen (ErasmusMC, Health-RI)

Wiro Niessen mentioned that AI is very promising, but it also raises challenges for the health domain. There is an enormous opportunity to organize data-driven science in the health space and artificial intelligence during the COVID-19 crisis. However, the bottlenecks are the lack of standardisation, small sample sizes and the challenge to connect diverse patient data.

To tackle such obstacles with health data, Health RI was initiated. DTL and Health-RI are strongly connected. Health-RI has three action lines to support data-driven science: 1) mobilising the shared voice of stakeholders in the health field, 2) community-led development of a national health data infrastructure, and 3) a national portal and helpdesk for health-data-related services and guidelines. Health-RI is being set up to provide an overview of initiatives, services and tools. It also connects to related national and international initiatives.

To adequately respond to the COVD-19 crisis Health-RI has started to establish a national observational database for COVID-19 data. Researchers with a research question should then be able to access data from Dutch hospitals. To achieve this, issues such as 'who can have access', 'how to acknowledge contributors' and alike have to be solved. Different data sets, such as samples, images, clinal cata and post mortem data from different partners in the Netherlands have to be linked to become available. An example is the Dutch Society of Radiology: all image data should be accessible. Wiro pointed out that it is important to link initiatives and to bridge gaps between researchers of various disciplines. Furthermore, it is also essential to keep data apart for training to test AI algorithms. Different data is needed for

training and for validations, otherwise it is hard to judge whether tools (based on these algorithms) are robust or not.

<div style="border:1px solid black; padding:10px;">

Questions/remarks/ answers raised during the presentation:

- "Does Health RI also connect to international initiatives?" Answer: Health-RI connects the Dutch nodes in several European infrastructures, such as ELIXIR and BBMRI.
- Comment: Niklas Blomberg (director ELIXIR) expresses his 'outside view' that the Netherlands is leading with it's Health-RI concept, in the way it links between domains.
- Comment: Monetary investment is urgently needed and is starting to take shape. Due to the COVID-19 crisis people now realize the missed opportunities and the need to act. It has become clear that data FAIRification is really important, in addition to having access to computing power.

</div>

4. Presentation: Connecting data, tools and people across Europe: ELIXIR's response to the COVID-19 pandemic. Niklas Blomberg (ELIXIR)

'In order to democratize science, everyone should have access to infrastructure', Niklas Blomberg said. ELIXIR helps to be a broker between European countries at the level of life sciences data. Therefore ELIXIR is organized in nodes, represented by European countries. DTL acts as the Dutch node (ELIXIR-NL). ELIXIR assists with connecting data resources and then offering them to researchers. It responded to the COVID-19 crisis by 1) creating federating COVID-19 data platforms which are connected, by 2) fostering good data management practices (incl. data stewards) and by 3) providing open sources tools. In the COVID-19 crisis it has become apparent that it is of utmost importance to use the same data standards for clinical data in all countries. To achieve this, one also has to build trust. Via ELIXIR, distributed national nodes now openly work together and share their work, building on the level of trust created between the nodes. ELIXIR can thus work with various nodes and make sure that national initiatives talk to each other. COVID-19 demonstrates the need for stronger European collaborations (and this could be a basis to work together for other diseases). ELIXIR now collaborates with EBI to establish the European COVID-19 data portal. ELIXIR's projects such as the CONVERGE project also further strengthen and connect European data management networks.

Questions/remarks/ answers raised during the presentation:

- A question about metadata was raised. Niklas mentioned that metadata gives insight into how to access, use, work with the data. Today, there is lots of data out there, but little provenance of the data. This diminishes the value of data. We need to pay more attention to this.
- Comment: Barend Mons agreed to this and mentioned that creating FAIR (meta-)data is now mandatory in the Dutch ZonMw/ NWO COVID-19 grants.

5. PAC Exchange: What COVID-19 initiatives are you involved in?

A list of COVID-19 initiatives partners are involved in:		
Name / organisation	**Initiative**	**Link/reference**
Rob Hooft (DTL)	RDA COVID-19 Recommendations	https://doi.org/10.15497/rda00046
Ingrid Dillo (DANS-KNAW)	RDA COVID-19 Working Group Role: co-chair	https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0
Slava Tykhonov (DANS-KNAW)	CoronaWhy Role: Lead / involved in Common Research and Data Infrastructure	https://www.coronawhy.org
Dimitris Koureas / DiSSCo	DiSSCo-CETAF Task Force on COVID-19 / Biodiversity evidence on zoonotic diseases	https://cetaf.org/covid19-taf-communities-taking-action
Celia van Gelder (DTL)	VODAN IN Training Cluster co-lead	https://www.go-fair.org/implementation-networks/overview/vodan/vodan-in-clusters/#Training
Erik Schultes	Regarding domain-specific metadata, the M4M Working Group is developing an approach to foster convergence rather than divergence. M4M's are running now in VODAN Africa project and will likely be deployed in the ZonMW COVID program	https://bit.ly/M4Mworkinggroup
Erik Schultes	FAIR Implementation Profiles: Community defined implementation of FAIR Principles (for example, metadata schema created in M4M workshops are components of the FIP)	https://bit.ly/FIPworkinggroup
Erik Schultes	VODAN Africa	https://www.vodan-totafrica.info

Erik Schultes	DSCC IN with overlap in the VODAN data stewardship cluster	https://www.go-fair.org/implementation-networks/overview/dscc/
Mijke Jetten (DTL)	Health-RI Covid data programme	https://www.health-ri.nl/initiatives/dutch-covid-19-data-support-programme

6. Presentation: Scouting and optimising the Dutch data landscape for open science. Melle de Vries (KNAW, NPOS)

Melle de Vries is project manager of the NPOS E project. He introduced the three lines of the NPOS programme: 1) Open access, 2) FAIR data and 3) Citizen science. Together with the NPOS F project (data stewardship: competences, training, education; project co-leads: Margreet Bloemers (ZonMw), Margriet Miedema (LCRDM), with strong support by Mijke Jetten and Celia van Gelder (DTL), the NPOS E project falls in the line of the FAIR data programme.

Melle explained how the landscaping project team has mapped the different actors in the Dutch data landscape, by 1) analysing various national and international reports, by 2) panel meetings by 3) visiting various research organisations and by 4) exchanging knowledge with other countries. What became apparent from this analysis is that the landscape is very fragmented. This is generally not a bad thing but on a national level more coordination is needed.

Coordination should be done on three levels: 1) data services, 2) expertise and 3) international collaborations. Melle recommended to keep supporting bottom up initiatives, yet strive for more connection. There should be a substantial investment made in quantity and quality of data stewardship. This also connects to the outcomes of NPOS project F.
He mentioned that most universities have already started to create Data Competence Centres (DCC's, mentioned earlier) and that there will be a second call for thematic competencies centres, in which DTL could really take a role.

Questions/remarks/ answers raised during the presentation:

- "Within the DCC call not all research performing organisations seem to be eligible for co-funding (e.g. musea). Also in the national data landscape description significant

biodiversity-related investments (at national/int. level) seem to be missing (e.g. NLBIF / DiSSCo). Is that correct?" Answer: This is correct and based on NWO's choice. But DCCs will be needed in all small/large organizations and connections need to be made.

- "How and who ensures the quality/quantity of data stewards? Who could train them?" Answer: this is really depending on the needs of the disciplines. This is why Melle suggested to structure this bottom up.
  - Comment 1: different disciplines have different needs. The data is different and today there is more awareness of the advantages of diversity. Therefore people are often willing to combine and share their knowledge. Now, this needs to be more and better organized and there should be more role models. Compliments to Mijke & Celia on work on Data Stewardship in DTL because it got a lot of attention, also internationally.
  - Comment 2: There needs to be more recognition and facilitation of the capacity building and training of data steward.
  - Comment 3: Lots of skills for data stewardship is not domain-specific, but also domain-specific skills are needed. This clearly needs to be a combination of the two.

7. Presentation: National Data Sharing Coalition across sectors: Economic growth through creation of a digital ecosystem, a generic agreement system and FAIR data. Maarten Bakker (Innopay, EZK-Data Sharing Coalition).

Maarten Bakker presented the National Data Sharing Coalition. This is a coalition mostly between private companies, with strong support of the Ministry of Economic Affairs and Climate (EZK). He mentioned that data is important and that the data economy is growing. This is reflected by the increase in data professionals. He pointed out that data needs to be shared because most organisations don't have enough data themselves. This raises legal and technical challenges. There are different ways of sharing data.

Data can be shared from one party to another. Here trust is of the essence and contracts have to be in place. Secondly, data can be shared from many parties to many parties via a specific platform. In this setting the data can cause monopoly, generated by transactions between party A to B. He noted that we should avoid data monopolies. Thirdly, data can be shared via agreements between multiple stakeholders. In his vision we should move toward these sorts of agreements.

He further mentioned that most data sharing initiatives focus on certain sectors or domains (using standard taxonomies and agreements). However, there should be more effort towards common taxonomies, bridging different sectors. This could be done via a cross-sectoral data sharing coalition (DSC) which can create a trust relationship. Today, there are few use cases that can be classified as a data sharing coalition and these follow a first version of a 'trust framework'. For each new use case a new trust framework has to be created. He mentioned that they are still looking for new cross-sectoral use cases. He asked if the life sciences field could potentially come up with a private/public data sharing use case.

<div style="border:1px solid black; padding:1em;">

Questions/remarks/ answers raised during the presentation:

- Comment: "The National Data Sharing model fits perfectly with the M4M / FIP concept for driving convergence on domain-relevant community standards on FAIR implementations (FAIR Principle R1.3). It would be nice to FAIRify data sharing agreements. These could then become part of the FIP for those sectors." (Erik Schultes/GO FAIR-DTL)
- "Does the data stay where it is (distributed)? You didn't mention that in the presentation." Answer: For a lot of use cases this makes a lot of sense, sometimes however large data sets are created and are then shared from a central place. So both scenarios are possible, depending on which type of data.
- Comment: "This could be interesting for the biobank data collections. Could this be a use case option?"
- Comment: "This might be a [relevant recent paper in this context: The TRUST Principles for digital repositories](#)."

</div>

8. PAC Exchange: DTL Alliance as data sharing coalition in the Dutch national data landscape?

How can the life sciences field be best represented in the National Data Sharing Coalition (public and private)? Should DTL become a partner in the National Data Sharing Coalition?

Ruben asks if people could raise their hands if they feel the need to be included as participants or designers. Several company representatives remarked that this could be of value for their organisations. Ruben will organise a follow-up meeting with Maarten Bakker and set up a process involving DTL partners to discuss a potential follow up in working towards a data sharing coalition for the life sciences sectors.