DTL Partner Advisory Committee meeting 09-10-2019

Report of the 11th DTL Partner Advisory Committee meeting
October 10th 2019, 14:00-16:30h

Meeting location: Intel Hotels Utrecht Centre, Smakkelaarshoek 24 3511 EC Utrecht
Invited: Formal representatives of DTL Partners and representation of DTL Board
Meeting documents: https://www.dtls.nl/about/organisation/pac/documents-pac-meetings/

Report (including the presented slides).

1. Welcome / arrival

2. PAC Opening

Report of the 10th meeting of the DTL Partner Advisory Committee (Annex)
There are no comments or amendments to the minutes of the previous meeting.

3. Presentation Health-RI

Leone Flikweert (Kwartiermaker Health-RI) presented the process Health-RI is making entitled 'Creating a national data services infrastructure for integrative life science research and innovation: Health-RI as example for the Health sector'. She introduced the reasoning why a data infrastructure for Health-research is necessary, touching upon topics such as reproducibility of data and re-usability of data. She mentioned that the main challenges to implement a good data infrastructure are of cultural, organizational and social nature rather than of technical nature. However, there are still technical hurdles such as making better quality data and tools accessible for research, in a sustainable way. Health-RI is needed to profit from big data (and advanced analytics such as AI). With lots of preparatory work done already (incl. in DTL and GO FAIR), now the timing is right for a collaborative effort such as Health-RI. Health projects are still often organized in silos, and therefore it is necessary that we organize ourselves (and form a collective voice).

What is DTL's role in Health-RI?
Leone mentioned that DTL has an important community role, as DTL has specific domain expertise. DTL is actively participating in Health-RI's programmes, for example in coordination of data stewardship support and training.

Health-RI's bundling of health-related data services

Services & tools which are easily searchable and findable via Health-RI's portal will be continuously extended (contact Health-RI if you want your organisation's services to be added). A support desk for such services and related health data questions is planned. This overview of services is valuable for researchers and data managers. However, Health-RI represents a complex community, which poses unique challenges on how to structure such an overview. Health-RI's main stakeholders are NFU, Health Holland, Lygature, DTL, ELIXIR-NL and Principal Investigators in academia. Also funding agencies like ZonMw and the National Health Care institute (ZiN) are actively involved. Another challenge which needs to be addressed is that the rules for acceptance of tools & services are not yet fully defined. This needs to be done to safeguard a high quality.

Here are some main discussion point that came up after the presentation:

- **Future FTE's:**
  Leone mentioned that there will be around 15 fte working for Health-RI in 2020- half of them centrally, the other half in parner institutes.
- **How to create/safeguard good quality data:**
  There was an active discussion on how Health-RI can contribute to the creation and safeguarding of good quality data. Leone mentioned that guidelines and principles could help. It was mentioned that DTL's Enabling Technology Hotels creates FAIR data at the source and can thus also positively contribute to good quality data creation. But it was also noted that FAIR data standards do not determine the quality of the data and that also provenance is important. FAIR standards do supply a lot of that provenance. Once data is FAIR it is also easier to expose flaws in data, which thus in return improves the quality of data before it is published.
- **Addressing the knowledge gap on data tools/standards:**
  In the discussion it was mentioned that there still is a considerable gap in knowledge of researchers on good data standards. Not a lot of researchers know what FAIR is. To bridge this gap, Health-RI will actively communicate, collaborate with other stakeholders and try to provide tools that will assist researchers.
- **Aligning with international initiatives:**
  Through existing relations, Health-RI is already aligned with some international initiatives such as BBMRI, ELIXIR and GO FAIR. But this connection can and will be extended to other relevant international initiatives.
- **Connection to policy makers in the NL:**

Leone mentioned that Health-RI communicates with ministries of Health, Education and Science and Economic Affairs. Health-RI's expertise is valued and asked for by policy makers.

4. Updates DTL

DTL director Ruben Kok gave a general update. Recent updates included:

- **ZonMw's Project 'Professionalising Data Stewardship' has been finalised**
  Three data stewardship roles have been defined working in the areas of: 1) Policy, 2) Research, and 3) Infrastructure. All results can be found here: https://zenodo.org/record/3474789#.XZ2SB0YzbIV and in DTL's zenodo community https://zenodo.org/communities/nl-ds-pd-ls/
- **Digital Competence Centers in academic institutes will be co-funded by NWO**
  NWO reserved 4.5M and a request for proposals is expected soon. The institutional DCCs will play a crucial role for educating and employing data stewards. There will also be a call for inter-institute digital competence centers: each specific for a science domain. DTL has already put some thoughts on paper on how these inter-institutional DCCs could work. The NPOS (National Plan Open Science) project group, a collaboration of KNAW, NWO, SURF and DTL is working on an analysis of the Dutch data landscape.
- **Updates from EOSC**
  There have been 5 working groups established by the EOSC Secretariat. Rob Hooft (DTL) and SURF are working in the "FAIR" working group. A National reference group set up by NWO will provide their input on national perspectives to EOSC.
- **Designing an internet of FAIR data and services (GO FAIR)**- *commented by Barend Mons*
  "Center of the hourglass" is what the "Fair Digital Framework" is called. This is GO FAIR's equivalent to the TCL/IP protocol for the internet. GO FAIR has reached a consensus on how this will be built (Digital Object Architecture and Linked Data Platform standards are forming parts of this).
- **ELIXIR CONVERGE has started**
  ELIXIR CONVERGE is a H2020 project which has been awarded to ELIXIR to create a European network of Data Stewardship expert centres. ELIXIR-NL will be actively participating in this project. It will fit perfectly to our national data management network and offer visibility to tools such as the Data Stewardship Wizard.
- **Heavy investments into development of artificial intelligence:**
  The Dutch Government is doubling their investments in AI to 128M per year. Currently

there are two AI coalitions: Dutch Digital Delta & VNO-NCW. They are in close contact and now merging their agenda's. Availability of well-structured data is paramount for AI to be effective and this matches the DTL and Health-RI agenda's. Questions that we as a network should address: how can data be made Fully AI-Ready?

- **Together with Helis Academy, DTL has organized data stewardship courses**-*commented by Celia.* DTL has had some requests from organisations asking for similar in-company training. DTL will think about how to build such training. In the meantime all the training materials are openly available.
- **The magazine Data Intelligence had a special issue on globally emerging FAIR practices**. DTL will provide a copy to all partners.
- **The 5th ETH Call (ZonMw) has just closed.** About 60 applications have been received by ZonMw and the review and finding decisions will take place in the coming period.

5. Breakout sessions discussing 'DTL's current and future strategy'

DTL has and has had the goal to make "integrated life sciences" a reality. With integrated life sciences DTL means the power of combined data from different disciplines resulting in new insights that had previously not been achievable, such as working towards personalized health, diets, precision breeding in agriculture, etc. In order to combine data from disparate disciplines within life sciences lots of work has still to be done for tools, technologies, standards, infrastructure, etc. Whereas for example certain tools are specific to the life sciences only, lots of these are and will be valuable for other disciplines as well.

In several breakout sessions PAC members have discussed different topics crucial for the refining of DTL's current and future strategy on implementing 'integrated life sciences', The first discussion evolved around the question of what integrated life sciences means for each partner.

Note: Input below is gathered from participants, sometimes in quotes. Text boxes provide key take-home messages and conclusions drawn by the DTL team.

5.1 What is Integrated Life Sciences for you?

1.) **Being able to re-use & combine data** (for example heterogeneous datasets from the clinic) by setting data standards & ensuring interoperability and removing silo walls.
2.) Preparing for the actual research work through a solid **data infrastructure** (meaning finding appropriate data sets, use of terminologies, identifiers ect)
3.) **Complex data analysis** (through software and algorithms)
4.) A **professional expertise community** (meaning findability of expertise, facilitating interdisciplinary collaborations in combination with integrated trainings provided by a learning community)
5.) **Knowledge Organisation Systems** (meaning information bundling and dissemination)

Quotes from participants:
"Integrated Life Sciences means that you can finally start asking interesting questions and do the actual research. This is how it could look like: you program your experiments and the robots will do the experiments. There will be intelligent environments for doing research"

"We are interested in following practical user discussions and in learning from each other. Learn what is new."

"We expect to be networking, to be informed, what is going on in the world, and to inform others."

"DTL should be the linking pin, the broker. The partners should bring the expertise."

"We should be actively involving the patient. Include patients into your teams. That way you'll get a broader support through inclusion and this will lead to for example, more biopsies in biobanks"

"Collaborative data sciences could be an alternative term for integrated life sciences (possibly integrated data)"

5.2 What should DTL focus on in the future?

Next PAC members have discussed what role they see for DTL now and in the (near) future. The discussion has been summarized to highlight the main takeaways points.

DTL should form a collective voice (for academia & industry).

This includes:

- making sure that solutions are made sustainable
- standardize
- lobby for funding
- keep pushing the importance of FAIR data (advocacy)
- push for "DTL"-like organisations for different science domains and work on interoperability between these national data organisation

However there were also critical voices about underrepresenting industry's needs and the dilemma of funding conflicts in academia- industry collaborations.

Quotes from participants:
"DTL should involve the commercial partners more in the discussions."

"Involving industry makes it possible to develop sustainable solutions. As an industry we are providing the tools, but will not always be the owners of the data."
"When academia and industry collaborate there is often a conflict in funding. Public funding is not eligible for exchanging with companies, this issue is not resolved yet and this limits cooperation. How some public investments could be used by private companies need to be addressed. This is a legal, ethical and standardisation problem. There needs to be more alignment toward the same infrastructure (academia and industry)."

"Funding for standardization should become available. We should address this to our funding agencies. We could for example prepare white papers"

Key takeaways:

DTL should form a collective voice for academia and industry.

DTL, involving it's public and private partners, now and in the future will focus on standardization, interoperability, lobbying for funding, advocating FAIR data principles and work towards sustainability of solutions, aligning the DTL public-private network with (inter)national organisations within and outside of the life sciences.

DTL should raise awareness around data sharing and provide training (e.g. Data Stewardship) to increase data literacy.

Several conversations evolved around outreach to scientists, helping with raising awareness and encouraging to share one's data. Another important topic of discussion was the need to train scientists in data handling.

Quotes from participants:

"There is no reward system for data sharing. How do we give incentive for data sharing? What is the value of data?"

"We need a cultural change for scientists to share their data. How do we do that? The cultural environment for sharing data is different for e.g. bioinformaticians than for medical researchers. One possibility could be to train PhD students formally."

"Motivation to share data is top down now. Should we define the role of universities better and include the funders (financial & administrative incentives). How can we motivate researchers to share their data? One option could be researchers who collaborate with data scientists. This could give them new possibilities to publish, new insights and automated data pipelines, to name a few."

"We have great training needs: academic institutes need to incorporate data science aspects in their curricula. There is a need for continuous education too. Generic, cross-disciplinary, as well as discipline specific."

Key takeaways:

DTL and its network encourage raising awareness for data sharing and professional data 'handling'. DTL does this by strengthening the role of Data Stewards and Data Scientists (as ambassadors and to reach scientists in partner organisations. For this group DTL invests in developing and giving training and professionalising their roles. DTL sees the important long term effects of train-the-trainers. As Data Stewards professionalise, they can reach and train more scientists in their network.

In this way DTL envisions a cultural change for data sharing and proper handling amongst researchers in the near future.

DTL should provide overviews of tools, standards, expertise and current developments.

Several participants noted that it is hard to get an overview over the field of data 'handling'. What is very much needed is a clear overview over tools, standards and best practices. Another issue that was discussed is how to find the right experts to tackle common problems and how to stay up-to-date with recent developments in the field.

There is thus a need to create overviews of:

- software tools
- standards
- best practices and templates
- expertise (people) and current knowledge
- current developments (in form of news ect) and updates on the field

Quotes from participants:

"There is an overwhelming amount of things that are needed, any help is welcome. We need best practices (templates to fill in for your own organizing) and a guide on how to deal with all other aspects (how to start sharing different types of data)"

"It is still a challenge to be visible and get recognition for your work on 'best practices'. There is still a lot of fragmentation out there."

Key takeaways:

Overviews are urgently needed. They need to be created and curated, extensive but findable and always up-to-date. This is a big task.

DTL's approach to this is to promote the involvement in the European research infrastructure ELIXIR, which has platforms for data resources, tools, standards and training and a broad collection of key services. DTL also contributes actively to the Health-RI portal, which has just been launched to start offering overviews of tools & services, standards and best practices.

DTL harnesses it's broad network of experts to advance (technical) knowledge and discusses issues within it's network through training, meetings (focus groups, interest groups, ad-hoc meetings), Interest Groups (e.g. on Data Stewardship), news and updates through (targeted) emails, social media (LinkedIn, twitter) and DTL's website.

DTL should be more involved with AI and advanced analytics.

The developments and the recognition of the potential of AI and advanced analytics has been an important topic in the discussions. There are still lots of hurdles when it comes to data being AI-ready. In general the agreement was reached that the analysis and it's insights are and will only be as good as the data it is based on. Here, DTL and partners need to invest more time and energy to tackle this challenge.

These are the main challenges that have been raised:

- establish an overview of the right tools for machine learning
- enable connections and collaborations around this topic (connecting people)
- define AI-ready data
- work towards responsible data science - FACT* data as the equivalent to FAIR data
- work on data interpretation (acting upon data insights, translating insights)

*Responsible data science depends on findable, accessible, interoperable and reusable (FAIR) data and the algorithms & tools it uses should be fair, accurate, confidential and transparent (FACT)

Quotes from participants:
"Responsible data science should be the goal for strategies and policies."

"Advanced analytics should not be mindless science. We should build on existing knowledge (e.g. in biology) in AI through semantic interoperability. "

"We need to keep up-to-date while working in a responsible manner. Countries that do not take e.g. privacy very seriously shouldn't have an advantage"

Key takeaways:

DTL should be more involved with AI and advanced analytics.

DTL recognizes the potential and hurdles that need to be tackled when it comes to advanced analytics. DTL will pay more attention to data science and combine its agenda on data stewardship and data infrastructure into a comprehensive agenda towards *'Digital Life Sciences'*. With upcoming strategies on AI and data science, DTL's partner network of bioinformaticians and computational life scientists, already assembled in research school BioSB, can play a key role in advancing this field for the life sciences.

DTL should help strengthen international connections.

There are numerous initiatives, projects and organizations world-wide who have overlapping goals (within and outside of life sciences). In the discussions it became clear that DTL should form a bridge between the partners in its network and suitable international initiatives. This helps to avoid duplication of efforts and learn from each other.

Quotes from participants:
"We see an important role - as a network- to make international connections"

"How does DTL fit into other European initiatives? We should be combining national and international level initiatives to not ignore similar initiatives and to learn from each other"

Key takeaways:

DTL should help strengthen international connections.

DTL is already very well connected to international initiatives, projects and organisations such as ELIXIR, GO FAIR, EOSC and through the projects that DTL's network is working on within these organisations such as FAIRsFAIR, EOSC LIFE and ELIXIR CONVERGE, to name a few.

Also, DTL partners are embedded in a broad range of international initiatives. We invite partners to communicate about these connections through the DTL partner network.

DTL is actively searching for initiatives to the field of 'integrated and digital life sciences' we are not yet connected to, and is reaching out to them.

5.3 What should DTL's future role be as an organization?

To summarize the above mentioned discussions and discussion points two views emerged for DTL's possible role in the future: DTL as umbrella organization and DTL as incubator of new initiatives.

DTL should be an 'umbrella' organization.

Being an umbrella would imply to provide a thin layer around it's partners and act as a network organisation. If you turn the umbrella upside down and choose the symbol of a salad bowl it would now imply facilitating coherency of a healthy mix, rather than 'shielding it's partners from the outside world'. As an organisation DTL sees the 'salad bowl' principle more fit to it's tasks.

Within this task it should:

- operate as an expertise and learning community (promoting data literacy)
- connecting people who are working on the same challenges

Quote from participants:
"DTL should be a learning community, where you can exchange knowledge, information, e.g algorithms designed for certain infrastructure. We would like to have access to expertise and engage in lots of discussions during projects, as well as forming more interest groups."

DTL should have an incubator role: be up to date with new developments, start a new initiative and then let go.

Another take on DTL's role in the past, present and in the future is that DTL works as an 'incubator' where it initiates new projects/initiatives, builds them until they gain momentum and then lets them go. Examples: of spin offs to DTL: GO FAIR, Health-RI, ...

Quotes from participants:

"DTL should be aware every time of the next challenge, stop what we have been doing until that moment. Avoid growing the organisation: hand over finished work to an institute or commercial partner. E.g. now that AI comes in, FAIR is out!"

"DTL can set things up where needed, once they are running we can get out and start doing something new."

" DTL could advocate certain solutions, showing a good example and that way inspire other

people"

"DTL-activity needs to limit itself to things individual institutes cannot do."

Key takeaways:

DTL has *both* an umbrella role *and* an incubator role. Both roles need active involvement of DTL's partners. In particular, DTL does not see its role as an implementer: this is the role of partners or spin off initiatives. DTL's network role helps to bring parties together and its incubator role is geared to help create initiatives that can take care of implementation, possibly involving individual partners. Here, DTL may continue to guide its partners in finding implementation partners, create overview of these and organise training where this is still lacking.

6. Closure and drinks

The chair thanks all partner representatives and Leone Flikweert as speaker, and closes the meeting.