



RESEARCH ARTICLE

REVISED **The Dutch Techcentre for Life Sciences: Enabling data-intensive life science research in the Netherlands [version 2; referees: 2 approved, 1 approved with reservations]**

Lars Eijssen¹, Chris Evelo¹, Ruben Kok², Barend Mons²⁻⁴, Rob Hooft^{2,3}, and the other founding members of DTL Data (see Acknowledgements)

¹Department of Bioinformatics - BiGCaT, Maastricht University, 6229 ER Maastricht, Netherlands

²Dutch Techcentre for Life Sciences (Foundation office), Catharijnesingel 54, 3511 GC Utrecht, Netherlands

³Netherlands eScience Center, Science Park 140, 1098 XG Amsterdam, Netherlands

⁴Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, Netherlands

v2 **First published:** 02 Feb 2015, 4:33 (doi: [10.12688/f1000research.6009.1](https://doi.org/10.12688/f1000research.6009.1))

Latest published: 06 Jan 2016, 4:33 (doi: [10.12688/f1000research.6009.2](https://doi.org/10.12688/f1000research.6009.2))

Abstract

We describe the Data programme of the Dutch Techcentre for Life Sciences (DTL, www.dtls.nl). DTL is a new national organisation in scientific research that facilitates life scientists with technologies and technological expertise in an era where new projects often are data-intensive, multi-disciplinary, and multi-site. It is run as a lean not-for-profit organisation with research organisations (both academic and industrial) as paying members. The small staff of the organisation undertakes a variety of tasks that are necessary to perform or support modern academic research, but that are not easily undertaken in a purely academic setting. DTL Data takes care of such tasks related to data stewardship, facilitating exchange of knowledge and expertise, and brokering access to e-infrastructure. DTL also represents the Netherlands in ELIXIR, the European infrastructure for life science data. The organisation is still being fine-tuned and this will continue over time, as it is crucial for this kind of organisation to adapt to a constantly changing environment. However, already being underway for several years, our experiences can benefit researchers in other fields or other countries setting up similar initiatives.

Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
REVISED version 2 published 06 Jan 2016	 report	 report	
	↑	↑	
version 1 published 02 Feb 2015	 report	 report	 report

- 1 Steffi Suhr**, European Bioinformatics Institute UK
- 2 Karin Verspoor**, The University of Melbourne Australia
- 3 Bryn Williams-Jones**, ConnectedDiscovery UK

Discuss this article

Comments (0)

Corresponding author: Lars Eijssen (l.eijssen@maastrichtuniversity.nl)

How to cite this article: Eijssen L, Evelo C, Kok R *et al.* **The Dutch Techcentre for Life Sciences: Enabling data-intensive life science research in the Netherlands [version 2; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2016, 4:33 (doi: [10.12688/f1000research.6009.2](https://doi.org/10.12688/f1000research.6009.2))

Copyright: © 2016 Eijssen L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work has been funded by the author's home institutes.

Competing interests: No competing interests were disclosed.

First published: 02 Feb 2015, 4:33 (doi: [10.12688/f1000research.6009.1](https://doi.org/10.12688/f1000research.6009.1))

REVISED Amendments from Version 1

Based on the suggestions made by the reviewers, we have adapted several paragraphs of the manuscript. In the abstract and introduction we make more clear that the paper mainly discusses DTL Data, and introduces DTL to provide context. We put the timeline in a separate box to support the structure of the manuscript. We added a paragraph and an extra Figure on the organisational structure of DTL. Furthermore, we improved the text on the background and needs underlying the setup of DTL (Data) and compared its setup to that of other projects. We added some information to clarify the tasks of DTL Data and its role and position in the national context, with respect to its support of tools, and in connection to ELIXIR. Finally, we shortened the detailed part on the meetings. Also, we rephrased some unclear sentences and corrected some smaller textual aspects.

We thank the reviewers for their valuable feedback and suggestions to improve our manuscript.

See referee reports

Introduction

In this introduction we will explain the origin of the Dutch Techcentre for Life Sciences (DTL) and its three programmes Data, Technologies and Learning. Furthermore, we discuss how the activities of the DTL Data programme fit in the parallel development of data stewardship and knowledge structuring initiatives in science overall.

Why was DTL started?

The initiative for DTL was based on the growing complexity of life sciences projects requiring multidisciplinary collaboration, coinciding with an increase in variety and volume of data. Starting 2003 several technology specific institutes developing technical services and techniques for life sciences have been set up in the Netherlands and operated until 2013. These include the Netherlands Bioinformatics Centre (NBIC), Netherlands Metabolomics Centre (NMC), Netherlands Proteomics Centre (NPC), and the Netherlands Consortium for Systems Biology (NCSB). These institutes were put in place to foster technology research, drive the exchange of methodology among labs and translate these into technical services that other scientists could use. Around 2012 it became clear that in the future the development and use of these technologies would no longer receive similar direct funding, and that research projects that apply a technology would need to budget for that. Coinciding with the increasing need for support of multidisciplinary collaboration and research projects, the technology institutes decided to develop a framework in which they would continue to exchange expertise across technology disciplines, build up a collective and well-accessible research infrastructure (RI) and deliver the services required. This has led to the formation of DTL.

The major advantages of this form of organisation are:

- Technology programmes working together in a single organisation enable the application of what we call *integrated life science research* requiring the use of multiple technologies in a single research project, and the integration of generated and already available data.

- Members of DTL can collaboratively draw attention to the fact that the fundamental developments in the technology fields require more attention of both the collaborating research organisations as well as the national funding agencies. Together we can look for solutions to tackle these challenges.
- Establishing a collective technology platform of the major research organisations in the Netherlands provides further chances to establish international partnerships for individual member organisations or as a collective.

An important distinction between DTL and other institutes with similar functions in other countries is that DTL was not set up as an institute by a (national) funding organisation (like e.g., the National Centers for Biomedical Computing in the US¹, and NECTAR in Australia, <http://hdl.handle.net/1885/8993>), but as a collaboration institute funded primarily by partner organisations. Where such bottom-up efforts to set up a supporting organisation are seen, they are often localised to a single research institute², and rarely started as a public-private partnership.

Timeline

During the preparatory phase of ELIXIR (elixir-europe.org), in 2012, several high profile bioinformatics and systems biology representatives started an initiative called DISC, The *Data Integration and Stewardship Centre*. They met several times to discuss the implementation of ELIXIR in the Netherlands. In parallel to this, the initiative to establish the Dutch Techcentre for Life Sciences was launched on the 31st of October 2012. The DTL organisation was started as a platform of leading universities, research institutes, university medical centres, science funders, government funding sectors ('topsectoren' in the Netherlands) and private companies from the health, nutrition, agrigenomics and industrial microbiology and information engineering sectors. We soon discovered that there was a significant overlap in the goals of the two initiatives, and it was decided to merge DISC into DTL as its Data programme. Starting from the 1st of January 2014, organisations have been signing up for formal membership of DTL.

DTL programme areas

As mentioned above, DTL has organised its actions in three areas, Data, Technologies and Learning, which run as individual but cross-connected programmes within the organisation.

DTL Learning, to start with the third area, manages an inventory of all training needs and offerings in life science technologies. It forms the bridge to the national Research School on Bioinformatics and Systems Biology (BioSB, biosb.nl) and other related *research schools*, and maintains contacts with all academic institutes that offer bioinformatics bachelor and master programs or postdoctoral training. DTL Learning also bundles expertise available in the DTL network, and organises both *ad hoc* and repeated training and courses on diverse subjects related to developments in the Data and Technologies programmes.

DTL Technologies bundles more than 100 research labs that offer support to life scientists with different technologies (so called

technology hotels). These technology hotels include a wide coverage of a variety of experimental (e.g. next generation sequencing, proteomics, metabolomics, bioimaging) technologies as well as bioinformatics and systems biology expertise. DTL Technologies facilitates the contact between the technology hotels and external researchers as potential customers *e.g.* through the organisation of funding calls that encourage new collaborative projects. In the DTL Technologies programme we will also work on harmonising and optimising access to hotels to make it easier for life scientists to use the latest technological opportunities and access multiple facilities in parallel.

DTL Data brings together experts on every aspect of data stewardship, tools and databases, and e-infrastructure. DTL Data builds relations for the people involved in the other DTL programmes and partner organisations and connects to international initiatives such as ELIXIR, the pan-European life science research data infrastructure. The setup of DTL Data has gone hand in hand with more generic developments related to data and knowledge handling in the life sciences that we will address below, before describing the DTL Data programme in more detail.

Whereas all scientific work in each of the three programmes is carried out within the participating institutes, DTL is run as a small not-for-profit organisation that has a central operational team governed by a board of domain representatives. The team takes care of organisational tasks to support and manage the functioning of the three programmes. The coherence between the three programmes is guaranteed by the participation of the three programme leaders. Operations are monitored by a scientific advisory committee consisting of senior staff from academia and industry and the board

is advised by executive representatives from the partner institutes (Figure 1).

Parallel developments: data stewardship and knowledge structuring

Driven by the rise and wide application of modern data-intensive technological approaches, present-day research projects in the life sciences collect much data that intrinsically has more value than the project itself will extract. This valuable and costly to obtain data offers many opportunities for re-use, which makes serious investment in keeping the data available and properly annotated (data stewardship) a more effective strategy than reacquiring the data. This also opens up the possibilities for research to reach new conclusions based on existing data. Furthermore, good data stewardship is required to make the work reproducible. Initially in the US and later in Europe, funding agencies have started demanding data stewardship to be an integral part of all scientific research projects. Besides data stewardship, integrative research requires proper structuring of knowledge based on aggregated and possibly curated findings of previous research.

DTL facilitates data stewardship and knowledge structuring in all associated projects through participation in the development and deployment of an initiative to make data Findable, Accessible, Interoperable, and Reusable (FAIR, datafairport.org). Making data and knowledge sources findable and accessible by both humans and computer systems requires a standardised description of metadata and study capturing as well as long-term storage and proper licensing. Interoperability and reusability require the representation of data and knowledge in such a way that they can be easily combined and used for further analytical processing³.

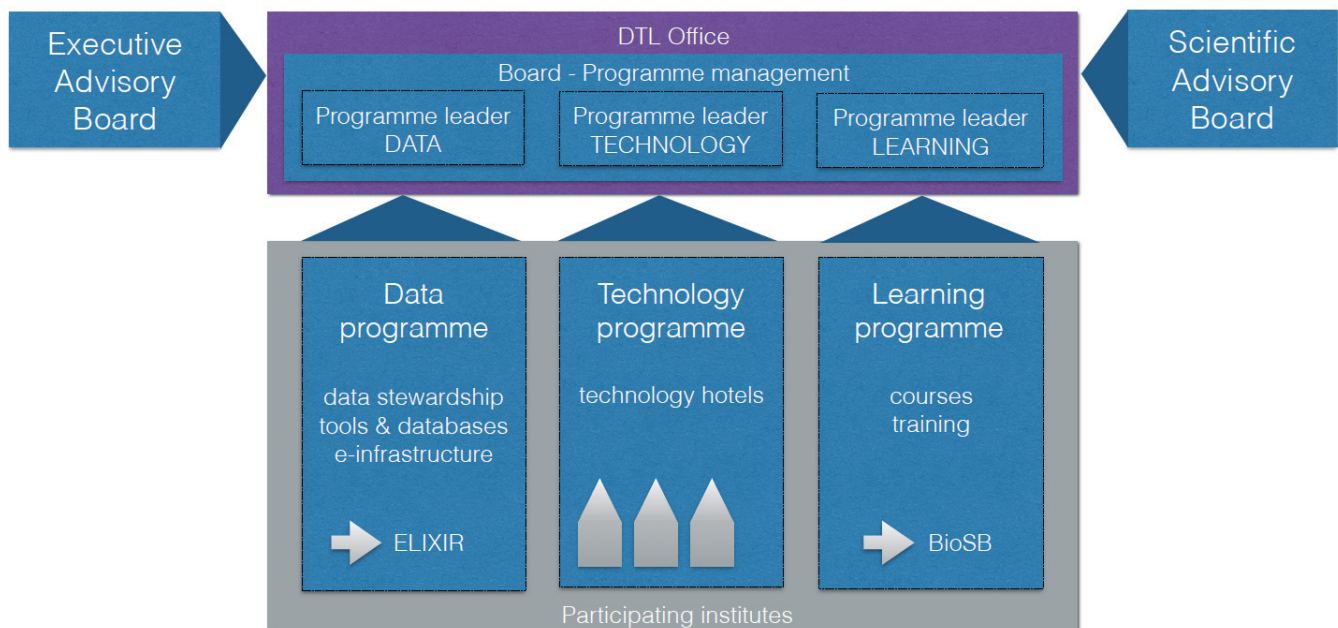


Figure 1. Organisation of DTL and its programmes. The board of DTL is advised by executive and scientific committees. For each of the programmes, Data, Technology, and Learning, it contains a programme leader who coordinates the work done by the DTL partners in the programme and takes care of the coherence with the other two programmes.

To support practical implementation of good data stewardship, DTL and its Data programme are on a mission to bring together all experts that can help life scientists with different aspects of their data management, and to show life scientists that it is not efficient to do everything in house using local solutions.

The remainder of this paper describes the organisational structure and approaches of the DTL Data programme in more detail.

Content of the DTL Data programme

DTL-associated scientists and engineers are responsible for data integration and stewardship in various life science initiatives in different life science sectors. They bring in expertise, reusable tools and databases that have been developed in the Netherlands or elsewhere, and have access to a shared e-infrastructure.

Bioinformatics and medical informatics expertise

DTL brings together experts with a very diverse professional expertise in life science data management. This expertise is classified along four independent dimensions:

- The life science *sector*: current activities are in health, agri/food, nutrition, and industrial biotechnology.
- *Location*: even though the Netherlands is a relatively small country, a local expert is sometimes preferred for an advice or in a collaboration.
- *Phase* in the data lifecycle: we distinguish expertise in planning an experiment, collecting data, data processing, data analysis, data and knowledge integration, and modelling. There is also underlying expertise in biostatistics, systems biology, instrumentation, data security, computing infrastructure, and computer science approaches.
- Technical *discipline* and type of data: e.g. genomics, proteomics, metabolomics, bioimaging, biobanking, knowledge representation.

To make expertise fulfilling requirements on all four dimensions available to life scientists all over the country, we are working on setting up a network of local expert centres at different sites. Such expert centres can function as help desks: places where information can be obtained about the expertise available locally as well as elsewhere. Representatives of the expert centres are in frequent contact with each other to learn about new developments and learn of each other's experience (both in techniques and in organisation). Over time, DTL will also extend its own help desk that can guide people to the right expert centres.

A very important mission of DTL Data is to prevent projects from running into problems because of *unconscious incompetence*⁴; we want to facilitate early interaction between life scientists with a specific plan and experts in all the technical fields that they need to engage, to avoid underestimating technological tasks or risks.

Tools & databases

Many researchers and other experts have (co)developed reusable tools and databases. They have ample experience to implement these in different projects. Such tools can often be implemented

in a new project using an existing deployment with dedicated user support. In other cases, specialised installations of the software can be made, tailored to the project. DTL has a strong preference for reuse of existing tools, which have proven their value in earlier national or international projects. Advantages of such tools are that they have overcome their teething problems, that their continued development benefits multiple projects, and that the reuse increases interoperability with other tools and existing data.

DTL, at this moment, is not performing any selection or endorsement of tools and databases. DTL Tools are supported by DTL partners that have developed the tools or by power users in the organisation. DTL, however, is actively involved in the inventory of Dutch tools for the ELIXIR registry (<https://bio.tools/>). In future, DTL may suggest selected unique, scalable, and internationally supported tools to be accepted for inclusion in the ELIXIR programme.

e-Infrastructure

In the past, many life science labs have each been taking care of their own needs for computing. More and more, however, the data to be processed becomes too large to handle. Furthermore, server system maintenance is not a core competence of a life scientist, and keeping a local cluster running should not be the task of a PhD candidate. Computing and data storage are becoming an infrastructure: equipment that nobody can do without, and which is inefficient to duplicate for every project. Many groups are therefore no longer willing to maintain the needed infrastructures themselves, and set up institutional services together employing specialised people for maintaining the computing equipment. Additional benefits of such centralisation efforts are flattening-off peak demands and allowing individual projects to be run at relatively short notice. Also, it reduces the need for synchronising purchase of new equipment with the start of new projects, which without central facilities results in waste for short projects and the use of outdated computing resources for longer projects. DTL brings experience from centralisation efforts together, and ensures alignment with the national centres for computing as well as international e-infrastructure projects like the European Grid Initiative (EGI, www.egi.eu) and EUDAT (www.eudat.eu). DTL links to and between the people that work on harmonising the computer centres so that migration of computing work and federation of resources become easier. When a new data intensive life science project is started with new demands for computing or storage, the best solution for the location of such computing is found in collaboration.

The e-infrastructure that can be shared is not limited to the computer racks (Infrastructure as a Service, IaaS). We also investigate possibilities for sharing higher level platforms (Platforms as a Service, PaaS), for example the workflow supporting software Galaxy⁵, which has been supported by the Netherlands bioinformatics centre in the past, and potentially other shared infrastructures for systems biology. We are also working together on a shared data publishing infrastructure based on experience from the Open PHACTS project⁶.

Organisation of the DTL Data programme Organisational structure and facilitation

The DTL Data programme is coordinated by a programme manager from the DTL Team. All projects are executed by DTL partners

outside of the team. The primary organisation of DTL Data is per sector of life science research (Figure 2). We organise several kinds of meetings for different target groups, which we have identified as fulfilling an urgent need: project leader meetings, programmer meetings and so-called *focus meetings*. We also identify people with similar interest and facilitate interest groups and working groups with their own meetings. Each of these types of events will be described in more detail.

Project leader meetings

Parallel running scientific projects often share common needs, which could be addressed by common efforts. Also, projects can often benefit from previous experience by others that are not directly involved. To detect such synergies, DTL organises project leader meetings for each life science sector in which project representatives meet regularly to discuss their progress and intentions. For the healthcare sector, this is already in place; the other sectors (agrigenomics, nutrigenomics and industrial biotechnology) are now setting up similar meetings. The principal project leaders of the four sectors will be meeting together on a monthly basis to discuss progress and to identify synergies between the sectors.

Programmers meetings

Many of the programmers involved in the bioinformatics projects in the different sectors of DTL Data are so-called *embedded programmers*, often the only bioinformatician in a biology or medical setting. Others work together in groups. In order to keep

each other informed and to encourage interactions between these programmers, DTL Data calls them together every two months for lectures and workshops on topics ranging from programming techniques to biological applications.

Focus meetings

During our work we regularly recognise similar problems or solutions being raised in more than one context. These common interests can be signalled by the programme managers or brought up by DTL scientists. For such topics we organise *focus meetings*. A focus meeting brings together a group of people that preferably have never met in that composition, to discuss a subject that is either crossing borders between technologies or between sectors. Focus meetings are not only organised by DTL Data, but also by the DTL Technologies and DTL Learning programmes. A focus meeting often contains a few short lectures, followed by a well-prepared discussion that engages the whole audience. After the meeting, a white paper is written by the organisers of the meeting that is published on the DTL website.

Interest and working groups

If a group of people, *e.g.* after a focus meeting, feels the need to exchange experience more often, they can form a so-called *interest group* within DTL. DTL facilitates these interest groups with meeting rooms, and tries to find a young researcher as a champion of the group to keep it going. This is modelled after “Project and Area Liaisons” (PALs) from earlier EU and UK projects⁷.

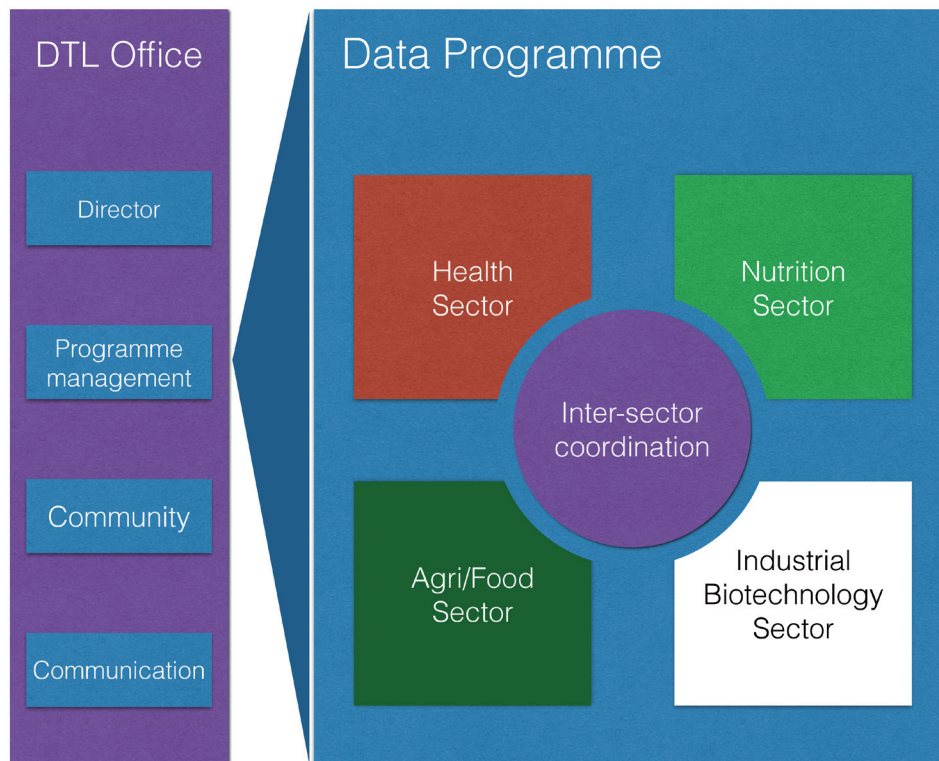


Figure 2. Organisation of DTL Data. DTL has a small office that organises and coordinates programme management, community and communication for the different programmes. The actual DTL Data projects are run by the partners outside of the DTL Office. The coordination structure is primarily divided into four sectors.

PALs are rewarded for introducing new ways of working: they are provided with extra support for their work and direct influence on the development of the new working methods.

An interest group that has identified an issue they want to work on together can form a *working group*. A working group needs to be supported by a part-time project leader to take the practical work out of the hands of the principal investigators. Each working group must deliver a practical result (deliverable) after a limited time. DTL is looking for ways to support the working groups by providing resources for the project leaders.

Both interest groups and working groups can be supported with a good software development environment, mailing lists, a website and a wiki to exchange information.

Relations with other DTL programmes

The Data programme interacts with many organisations, both internal to DTL (other programmes and partners) as well as external. We first describe the internal interactions.

Help desk, training and education

In the day to day operations of the Data programme, we frequently come across needs for training: both training for data scientists to broaden their knowledge with newly developed technologies, as well as training for life scientists to make them aware of and teach them how to use solutions that are being developed in DTL Data projects. This is expected to become even more important once the development of local data desks in different institutions will be realised. The setup of these data desks will bring together experienced data scientists from different institutes, and they will find out that others have complementary expertise that they sometimes need to replicate. Also, life scientists with less experience will have a low barrier to approach their local data desk for advice, bringing in more demand for basic data awareness training. All of these training needs will be developed with the DTL training Programme, which is very well connected to people and organisations that can support this effort.

Data-related technology hotels

Many of the people involved in DTL Data offer their services to life scientists as a Data hotel in the DTL Technologies Programme. DTL Data works with DTL technologies to define the needs of and requirements for these data-specific hotels. An overview of current DTL hotels is available at www.dtls.nl/expertise-services/hotels.

Relations with external programmes

External connection of DTL Data include for instance projects from the Innovative Medicine Initiative (IMI, www.imi.europa.eu) and infrastructures under the European Strategy Forum on Research Infrastructures (ESFRI) scheme, including ELIXIR.

ELIXIR

Synchronous with the development of the DTL organisation, bioinformatics institutes and laboratories all over Europe have set up the European research infrastructure for life science data and bioinformatics, ELIXIR. ELIXIR is organised as a hub hosted at the EBI in Hinxton, UK, and nodes in each of the member countries.

In the Netherlands, DTL hosts the ELIXIR node (ELIXIR-NL). Each of the nodes contributes specific expertise to the ELIXIR network. The Dutch node contributes expertise related to interoperability, learning, and computer and network infrastructure. Through the network, DTL-associated scientists can benefit from all European contributions.

DTL and ELIXIR have developed the concept of so-called *Bring Your Own Data* (BYOD) meetings as a platform to bring together data owners and data experts. Also biological domain experts are invited where relevant. The main goal of these meetings is to get data owners acquainted with the possibilities to connect and functionally interlink their data with other datasets and knowledge resources by applying FAIR principles. Researchers can suggest a BYOD party and DTL will assist with the logistics and invite data experts.

Other ESFRI programmes and national projects

Europe has many other Research Infrastructures in the life sciences, each with their own special focus. Also in the Netherlands several larger project organisations are active in life science research. All of these have their own research data and associated challenges. In the Netherlands we make sure that the people working with that data are co-developing and steering the DTL Data Programme. This ensures that the methods and tools they use are compatible with the ELIXIR choices and avoids unnecessary duplication of development efforts.

Conclusion

Life science research becomes more and more data intensive and cross-disciplinary at unprecedented scales. Individual research groups do not have the resources and the interest to keep in contact with all expert providers and keep informed of the progress of other related projects at such scales. In the Netherlands we have developed a networked approach to accommodate for the challenges posed by modern data-intensive life science research. The establishment of DTL as a collective platform that brings together experts in various technological disciplines across life science domains, facilitated by a small core team, allows projects to run efficiently. Already in the preparatory period and in the first year of operations we have identified synergies between parallel running research projects and found common interests across researchers with a focus in surprisingly different disciplines. The growing community of experts involved in DTL Data makes sure that necessary data-related expertise can be located for any researcher in the life sciences starting on any new project. At the publication date of this article DTL had over twenty confirmed member organisations. The current partner list can be found at www.dtls.nl/about/partnership/.

Contact

To find out how DTL Data can support your challenges or for more inquiries about the setup of the organisation, contact Rob Hooft (programme leader) at rob.hooft@dtls.nl

For further information on the other programmes of DTL contact Ruben Kok (director DTL) at ruben.kok@dtls.nl

Website: www.dtls.nl

Author contributions

All authors have been involved in the setup of DTL Data and have contributed to the text of the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

This work has been funded by the author's home institutes.

Acknowledgements

Next to the authors, the following people have played instrumental roles in setting up DTL Data, and all share a DTL affiliation:

- Jan-Willem Boiten, CTMM-TraIT, Eindhoven
- Luiz Olavo Bonino, Dutch Techcentre for Life Sciences (Foundation office), Utrecht
- Jildau Bouwman, TNO, Zeist and Netherlands Metabolomics Centre, Leiden

- Richard Finkers, WUR, Wageningen and Plant Research International, Wageningen
- Femke Francissen, Dutch Techcentre for Life Sciences (foundation office) and BioSB research school
- Celia van Gelder, Dutch Techcentre for Life Sciences, Utrecht and Radboudumc, Nijmegen
- Martien Groenen, WUR, Wageningen
- Jaap Heringa, VU University Amsterdam and BioSB research school (Deputy Head of Node ELIXIR-NL)
- Irene Nooren, SURFsara, Amsterdam
- Merlijn van Rijswijk, Dutch Techcentre for Life Sciences (foundation office), Utrecht and Netherlands Metabolomics Centre, Leiden
- Marco Roos, LUMC, Leiden
- Morris Swertz, UMCG, Groningen

References

1. Berg J: **National Centers for Biomedical Computing: from the BISTI report to the future.** *J Am Med Inform Assoc.* 2012; **19**(2): 151–152.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Kallioniemi O, Wessels L, Valencia A: **On the organization of bioinformatics core services in biology-based research institutes.** *Bioinformatics.* 2011; **27**(10): 1345.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Mons B, van Haagen H, Chichester C, *et al.*: **The value of data.** *Nat Genet.* 2011; **43**(4): 281–3.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Kruger J, Dunning D: **Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments.** *J Pers Soc Psychol.* 1999; **77**(6): 1121–34.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Goecks J, Nekrutenko A, Taylor J, *et al.*: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol.* 2010; **11**(8): R86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Williams AJ, Harland L, Groth P, *et al.*: **Open PHACTS: semantic interoperability for drug discovery.** *Drug Discov Today.* 2012; **17**(21–22): 1188–98.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Goble C, Wolstencroft K, Owen S, *et al.*: **SysMO-DB: A pragmatic approach to sharing information amongst Systems Biology projects in Europe.** In: *Proceedings of the UK e-Science All Hands Meeting.* 2009.
[Reference Source](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 09 February 2016

doi:[10.5256/f1000research.8266.r11903](https://doi.org/10.5256/f1000research.8266.r11903)



Steffi Suhr

BioMedBridges, European Bioinformatics Institute, Hinxton, UK

The authors have done a substantial amount of work to restructure, clarify and refocus the manuscript. I am happy to approve this new version.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 02 February 2016

doi:[10.5256/f1000research.8266.r11902](https://doi.org/10.5256/f1000research.8266.r11902)



Karin Verspoor

Department of Computing and Information Systems, The University of Melbourne, Melbourne, VIC, Australia

I thank the authors for their excellent work in overhauling this manuscript. I find it much improved in terms of clarity and structure. The paper does a nice job of introducing the Dutch Techcentre for Life Sciences, and describing an alternative model to providing ongoing support for modern data-oriented life sciences research.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 24 March 2015

doi:[10.5256/f1000research.6430.r7550](https://doi.org/10.5256/f1000research.6430.r7550)

**Bryn Williams-Jones**

ConnectedDiscovery, London, UK

This article outlines the background, formation and to an extent the operation of the Dutch Techcentre for Life Sciences (DTL) with a focus on the data programme, DTL-Data. Given this type of organisation is of relevance to data researchers in other nations, researchers in other fields, and set against a background of increasing data complexity and institutional networking in the wider community, this could provide a model for other initiatives.

In common with the other reviewers, there are some issues with the presentation which would help readers new to DTL to understand the background, structure and operation of this entity. To encourage similar organisations in other nations, the historical background is relevant, but could be split to a separate section - as a text box/diagram or similar to aid the flow of the description of DTL operations as they are now.

Whilst DTL-Data is discussed in some detail, the other sections are briefly mentioned and its challenging to gain a picture of the holistic DTL and the competencies necessary to establish a similar organisation in other nations. Again, as noted by the other reviewers, an outlook comparing other national efforts, their relation to DTL, what may have particularly aided development of DTL in the Netherlands from both a structural/funding and cultural/operational perspective would be useful. Similarly examples of FAIR data principles, information on some of the tools and databases mentioned but not named could help researchers in other nations contextualise the assertions made. This may be of more value to those less familiar with the operations of DTL than detailed description of the meetings which would be better summarised in less detail.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 29 Dec 2015

Lars Eijssen, Maastricht University, Netherlands

Thank you for your feedback. Herewith our responses to the points raised.

Response to "In common with...as they are now":

According to the suggestion, we have added a separate 'timeline' box to describe the historical context of DTL. Furthermore we have made more clear right from the opening of the paper that it is mainly about the Data programme, and described the remaining DTL context more in connection to this focus.

Response to "Whilst DTL-Data...in other nations":

Even when trying to keep the text on the other parts of DTL brief, and making more clear in the text that this paper specifically addresses the DTL Data programme, we have added an extra short paragraph and a diagrammatic figure (Figure 1 and the paragraph directly preceding it) on the organisation of DTL as a whole.

Response to "Again, as noted by...summarised in less detail":

We rephrased the 'Why was DTL started' subsection to more clearly introduce its benefits and reasons for setup and to extend the description of the projects that were running in the Netherlands before DTL, to better sketch the national operational and funding perspectives. We have substantially reduced the level of detail in our description of the meetings. For (stewardship and other) tools and databases, we deliberately had not included examples in order not to select some above others, since DTL partners are the entities that are offering and supporting tools, rather than DTL centrally, which does not enforce specific ones. We now included a paragraph to the manuscript (also in response to a similar point made by the first reviewer) clarifying this aspect and mentioning the relationship between DTL and ELIXIR with respect to tool selection, referring to the latter's tool registry ('Tools & databases' subsection).

Competing Interests: No competing interests were disclosed.

Referee Report 16 March 2015

doi:10.5256/f1000research.6430.r7547



Karin Verspoor

Department of Computing and Information Systems, The University of Melbourne, Melbourne, VIC, Australia

This paper describes the structure and role of the Dutch Techcentre for Life Sciences (DTL), with a specific emphasis on the DTL Data programme. The DTL is an important initiative in the Netherlands to support effective and efficient data-intensive life science research throughout the Dutch research community by facilitating the connection of researchers with computational tools, expertise, and infrastructure.

As the first reviewer has identified, there are some structural issues in the presentation of this paper that could improve its focus and clarity. There are also some more minor issues, for instance a lack of background on certain referenced programs when first mentioned (ELIXIR is introduced in the second paragraph of the Introduction, but not defined until several paragraphs later; not all readers may be familiar with the program).

Elaborating on the first reviewer's point about e-infrastructures, it would be helpful for the authors to relate the model described in the paper, as developed for the DTL, to other possible models that exist, for instance the National Centers for Biomedical Computing in the US (see e.g., <http://jamia.oxfordjournals.org/content/jaminfo/19/2/151>) or bioinformatics core facilities in place at a more local level (see e.g., the discussion at <http://bioinformatics.oxfordjournals.org/content/27/10/1345>). Given the emphasis on data management and computing infrastructure of the DTL Data programme, cloud-based generic e-research infrastructure supported at a national level (e.g. for social sciences research or the Australian National eResearch Collaboration Tools and Resources (NECTAR), e.g., the Genomics Virtual Laboratory and other Australian research infrastructure programs) are also relevant. I believe that understanding how the DTL Data model is different or unique in the global context is important to support the authors' goals of providing insight to new efforts. Furthermore, it would be helpful to more clearly distinguish the role of the DTL from the national centres for computing to which the DTL is aligned (p. 4).

There are a few minor issues with wording choices that the authors may wish to revisit, e.g. the phrase

"the path to professionalisation" -- in what sense is the organisation being "professionalised"? (this recurs in the phrase "professional data stewardship") -- and the phrase "unconscious incompetence" which sounds perhaps more severe than the authors intend.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 29 Dec 2015

Lars Eijssen, Maastricht University, Netherlands

Thank you for your feedback. Herewith our response.

Response to "As the first reviewer...familiar with the program).":

We improved the introduction of ELIXIR in the text. In the abstract we added context by: "ELIXIR, the European infrastructure for life science data". Furthermore both in the abstract and the timeline we removed the mentioning of the term 'ESFRI', to only introduce this (but now written in full) in the 'Relation with external programmes' section by: "...and infrastructures under the European Strategy Forum on Research Infrastructures (ESFRI) scheme, including ELIXIR".

Response to "Elaborating on ... DTL is aligned (p. 4).":

An important distinction between DTL and other institutes with similar functions in other countries is that DTL was not set up as an institute by a (national) funding organisation (like e.g., the National Centers for Biomedical Computing in the US, and NECTAR in Australia), but as a collaboration institute funded primarily by partner organisations. Where such bottom-up efforts to set up a supporting organisation is seen, they are often localised to a single research institute and rarely started as a public-private partnership. We have also included this text in the manuscript, including suggested references, in the 'Why was DTL started?' section.

Related to the last point, DTL has more a role as an orchestrator between research projects, generic computing initiatives, and national computing centres. As suggested by the first reviewer, we added some international initiatives to this phrase, widening the scope. Furthermore, we modified the next phrase ("Together these people work...") to "DTL links to and between the people that work..." to make DTL's role as a connector more clear.

Response to "There are a few minor...the authors intend.":

We removed both occurrences of professionalisation/professionalised, as they were superfluous. The phrase 'unconscious incompetence' was a deliberate choice, as a term coined to indicate the issue not being aware of not doing something according to standards. We added a reference to a paper by Kruger *et al.* on this topic.

Competing Interests: No competing interests were disclosed.

Referee Report 25 February 2015

doi:[10.5256/f1000research.6430.r7548](https://doi.org/10.5256/f1000research.6430.r7548)

**Steffi Suhr**

BioMedBridges, European Bioinformatics Institute, Hinxton, UK

The authors describe the drivers and motivation behind setting up the Dutch Techcentre for Life Sciences and describe the organisational setup so that the experiences gained can benefit researchers in other fields or countries setting up similar initiatives. As the number of initiatives seems to be steadily growing, this is helpful and could indeed provide some useful background against which other efforts could model themselves.

However, trying to come at this from the perspective of someone who might want to implement such an infrastructure, there are some bigger and smaller points that would need to be addressed to be able to give this paper an "approved".

1. The focus is slightly unclear: the abstract and introduction describe the DTL in its entirety, including its history, while more than three quarters are focused on the services related to the last area of activity listed, "DTL Data". The manuscript would benefit hugely from a reorganisation that makes the main focus clearer right at the beginning. It might also be worth considering to put the description of the background and origin (timeline, why was DTL started) at the end of the manuscript. Alternatively, the authors might want to consider leaving the history out of the manuscript completely: at present, the first point that seems to be mentioned is that funding became more scarce, rather than focusing on the benefits.
2. Abstract: as mentioned under (1), the focus on data-related issues should be reflected. It would also be helpful to (very briefly) give examples of "tasks ... that are not easily undertaken in a purely academic setting", as this seems to be one of the core drivers behind DTL. Examples for why and how the environment is changing, and how DTL can address this, would be good in the introduction (rather than leaving this hanging in the abstract).
3. The part about "governance and organisational programmes" is confusing: governance structures are not really described in any detail (who is on the board, and why?). In general, it is unclear that this information adds value to the manuscript - as organisational structures will often heavily depend on local framework conditions, the paragraph on the board and SAB could be left out and the section simply renamed "DTL programme areas" or similar.
4. Section "parallel developments..": the opening paragraph on open data is quite weak and possibly incorrect (e.g. the push towards open data does not originate from the application of data-intensive technologies - arguably, data intensive technologies provide a challenge for open data). For the purpose of this manuscript, less would be more.
5. Going into the Data programme, section on Tools & databases: which tools are supported by DTL? How are they selected and/or prioritised? What is the DTL contribution to these tools?
6. e-infrastructure: It might be interesting to explain if there any interactions with European-wide e-infrastructures.
7. Meetings described: There is an odd level of detail. What is missing is who is involved and why, in particular in the case of the project leader meetings. If these are a continuation of TraiT meetings it is clear that there would be a spirit of collaboration, but what about the "other sectors" mentioned -

how were participants identified, how will the aims be stated, how will members be motivated?
Focus meetings and working groups: who prioritises the topics, and based on what?

8. Relations with external programmes: ELIXIR - "...gives us the possibility to reach out to experts and tools all over Europe" - how about the other way around, can needs also be identified across DTL that would better be met at the European level?

Minor points (mostly language):

- Page 2: "...and let life scientists each replicate the EXPERTISE at their own institutes, the technology institutes decided to develop a FRAMEWORK in which they..."

- Page 3:

"..., and Reusable (datafairport.org). ALLOWING data and knowledge sources.."

"All expertise can be classified along those four dimensions" - delete sentence (redundant). Next sentence, change to: "To make expertise fulfilling all four dimensions available,..."

"Representatives of the expert centres are in frequent contact.." (delete "involved").

Change to: "More and more, however, the data to be processed becomes too large to handle."

- Conclusion: "...and found common interests across researchers with a focus in surprisingly different disciplines. ... makes sure NECESSARY data-related expertise.."
- ELIXIR should be listed by its name (without "ESFRI"), with the full name at first mention

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 29 Dec 2015

Lars Eijssen, Maastricht University, Netherlands

Thank you for the extensive and detailed feedback. Herewith a point-by-point response.

1. We rephrased the abstract to directly introduce the Data programme of the DTL, instead of DTL as a whole. We still kept some background on DTL in the introduction, since this gives the context in which DTL Data was set up and lives, but also there we adapted the first lines to mention the programmes, including Data, right away. Furthermore, we put the timeline in a separate box to improve the structure of the text. Also, we rephrased the text to indicate the benefits that clearly underlie both changes in funding policies as well as the Data programme.
2. Data-related issues get more focus in the adapted abstract, as described in the response to the first point. We also added a phrase to exemplify the tasks referred to ("DTL Data takes care of such tasks related to data stewardship, facilitating exchange of knowledge and expertise, and brokering access to e-infrastructure."). We now also already briefly mention the generic environmental changes right at the opening of the introduction: "The initiative for

DTL was based on the growing complexity of life sciences projects requiring multidisciplinary collaboration, coinciding with an increase in variety and volume of data.” For DTL Data specifically, this is further detailed in the ‘Parallel developments: data stewardship and knowledge structuring’ subsection right after the introduction of the three programmes.

3. We have adapted the paragraph in accordance with these comments. We renamed the section and removed the opening line on the board. At the end, we included a short paragraph and a figure on the governance structure, but more focused at its relationship to the programmes including Data.
4. We agree this was not clearly formulated. It is an indirect ‘push’: because of the more generic value such data has, making the data open gets even more important. We rephrased the opening sentence to: “Driven by the rise and wide application of modern data-intensive technological approaches, present-day research projects in the life sciences collect much data that intrinsically has more value than the project itself will extract”. Also, we reformulated the start of the paragraph to indicate that this is the reason for data stewardship (and for funders to ask for this) rather than funder’s requirements being its original cause.
5. DTL, at this moment, is not performing any selection or endorsement of tools and databases. DTL Tools are supported by DTL partners that have developed the tools or by power users in the organisation. DTL, however, is actively involved in the inventory of Dutch tools for the ELIXIR tools registry (<https://bio.tools/>). Selection does take place only when the initiative is taken to start supporting the tools or databases in an international context. In such cases the tools are verified to be one of a kind, best of breed and have sufficiently scalable support infrastructure. DTL can then suggest to include the tools in the ELIXIR programme. To be accepted into the ELIXIR programme, tools have to be part of the node plan, which must be accepted by the ELIXIR SAB. No tools have been added to the ELIXIR NL portfolio yet. A shortened version of this text has been included in the ‘Tools & databases’ subsection.
6. To name some important initiatives, we added the clause: “...as well as international e-infrastructure projects like the European Grid Initiative (EGI, www.egi.eu) and EUDAT (www.eudat.eu).”
7. We removed a substantial part of the detailed description, including the background related to the project leader meetings. Also we added a short line on who initiates the topics for focus meetings (“These common interests can be signalled by the programme managers or brought up by DTL scientists”), but not in too much detail to keep this part short.
8. Yes, obviously it would be a two-way interaction, as is the basic idea of ELIXIR. To make this more clear we added: “Each of the nodes contributes specific expertise to the ELIXIR network. The Dutch node contributes expertise related to interoperability, learning, and computer and network infrastructure. Through the network, DTL-associated scientists can benefit from all European contributions.”

Minor points: all adapted according to suggestion.

Competing Interests: No competing interests were disclosed.

