European Infrastructure for Life Sciences Data
ELIXIR

Scientific area:	Life Sciences (Bioinformatics)
Host country:	United Kingdom
Infrastructure type:	Distributed
Dutch node or similar:	ELIXIR-NL
Legal Entity:	Consortium Agreement
The Netherlands member since:	2014
Phase:	Operational

Duration of agreement
Ongoing

Terms of withdrawal
An ELIXIR Member State may withdraw from ELIXIR at the end of a Financial Year by providing one year's written notice sent to the Chair of the ELIXIR Board. The ELIXIR Board shall formally record the withdrawal. An ELIXIR Member that withdraws from ELIXIR is not entitled to claim any compensation or goodwill value whatsoever and shall continue to contribute to the ELIXIR Budget until its withdrawal is effective. Any outstanding contributions will be paid and obligations fulfilled before withdrawal of membership is confirmed.

Access to facilities
n/a

Access to data
ELIXIR has various platforms and communities, and provides various services. Communities and platforms are gated via institutional login. Services are open and online accessible. These services include: Compute, Data resources, Interoperability and standards, Software tools, Training, and are organised via tags.
Access to training is based on application procedure.

User definition
The definition of a user will vary across the various types of services, but it remains that ELIXIR's services are free and open to all types of users (academia, industry), wherever they are based on the planet. Users only require an internet connection to access services and typically no application will be made. Some exceptions exist in case of depletable resources such as compute capacity.

Description
ELIXIR is a virtual and distributed research infrastructure for life science data, organised along a hub and nodes model where a main hub coordinates the work done in the nodes. A node is a collection of research institutes within a member country. ELIXIR consists of 22 member countries; EMBL-EBI constitutes a special node that is not associated with a country. Nodes run the resources/services that are part of ELIXIR. ELIXIR's nodes run 300+ bioinformatics services, or 'resources' which span databases, tools, compute, interoperability and training.

<u>**Financial details**</u>
Mean Dutch membership over 5 year period (€K):   292 (*)
Mean Dutch share of contribution over 5 year period (%):   4,9

Year	NL membership (k€)	NL Contribution (% of total)	Total membership (k€)	Total expenditure (k€)	Turn-over (k€)
2016	205	5,5	3710	2922	4660
2017	241	4,8	5035	4461	6016
2018	329	4,8	6913	7128	8063
2019	338	4,8	7089	6916	7648
2020	345	4,8	7233	5172	7969

Employee statistics

	Female	Male	Other	Total
Total*	18	9	n/a	27
Of which Dutch**	n/a	n/a	n/a	n/a

[2019 in Full Time Equivalents (FTE)]

*Table *The information provided above relates to the ELIXIR Hub, which acts as a coordinating secretariat for the research infrastructure. The ELIXIR Hub coordinates different types of bioinformatics services across the Nodes. Whilst the total number of scientists involved in ELIXIR's activities exceeds 700, data is not currently kept on the gender balance across ELIXIR as a whole, though it is provided in this submission for the ELIXIR Hub and the ELIXIR Netherlands Node.*

***The ELIXIR-NL Node coordinating core team consisted in 2019 of (part-time) involvement of employees of VU-University Amsterdam, UMC-Groningen, Maastricht University, Stichting DTL Projects, Stichting DTL and Lygature.*

Use of the infrastructure
Number of users
n/a

Comments RI (2020)
The individual Nodes within ELIXIR monitor the usage of the services that they run. ELIXIR is working towards better coordination of user-number monitoring across all services. To give an idea of scale, database users of the set of 19 so-called ELIXIR Core Data Resources were estimated using 'monthly unique IP addresses' for the period 2014 to 2018: annually, they represented 2.1 million in 2014, increasing to 3.2 million in 2018. In addition, many of these resources are built in collaborations with resources around the globe, e.g. USA and Japan, so these numbers represent only a fraction of the overall usage. In terms of user training, ELIXIR Nodes ran 1,103 training events during the period 2016-2019, attended by a total of 26,303 trainees. The ELIXIR Authentication and Authorisation Infrastructure (AAI) provides a single sign-on service to currently 189 ELIXIR and other ESFRI services, typically registering 4,500 logins monthly. Another example is the GALAXY bioinformatics analysis platform, which is supported by ELIXIR, and which typically has more than a thousand active users per month. As described in other questions, calculating the number of users can not be done at the European level, as ELIXIR coordinates hundreds of different types of resources from databases to software tools. However, when we assess overtime the usage of individual resources, including Core Data Resources, ELIXIR's AAI and the provision of training, the trend of usage grows higher with each year, including usage of these resources from individuals in the Netherlands. ELIXIR's strength lies in the combined portfolio of core data services brought together and initially funded by its nodes, based upon national and European (EMBL-EBI) institutes involved. These services are activated for the entire European Science and Innovation community across all life science fields through three basic ELIXIR Platforms (1] Data Collections (mostly data at molecular and (sub-)cellular level (genes, proteins/enzymes, metabolites), 2] Tools (bioinformatics software applications and validated workflows to support bioinformatics analyses, and 3] Computing Facilities. Adding further value to this portfolio of selected individual data services is done through two cross-connecting ELIXIR Platforms: 4] FAIR-based Interoperability of data services and 5] Training and Capacity Building in bioinformatics and data stewardship. Having these distributed European data resources findable, accessible, interoperable and reusable through a central infrastructure is of crucial value to all life scientists in the Netherlands in academia and industry. This means that numbers of users cannot (easily) be counted.

User statistics of individual ELIXIR services are not (yet) collected at the ELIXIR level, although user statistics are collected by many individual service providing institutes, such as access statistics and downloads of EMBL-EBI data, the Swedish Human Protein Atlas (HPA), and the ELIXIR Core Data Resources BRENDA and SILVA.

Collectively, Dutch life scientists use these services on a daily basis and in very high numbers. EMBL-EBI user data (hundreds of thousands of individual users) are separately provided via EMBL-EBI. Statistics for use of three non-EBI provided core data resources are as follows:

HPA statistics show access by Dutch Scientists as follows:
- 2018: 8448 total individual users (of which 2887 male and 5561 female users)
- 2019: 8129 total individual users (of which 3017 male and 5112 female users)
BRENDA statistics show access by Dutch scientists as follows:
- 2018: 4155 total individual users
- 2019: 5266 total individual users
SILVA statistics show access by Dutch scientists as follows:
- 2018: 1462 total individual users
- 2019: 1423 total individual users

Two Dutch services Molgenis and BridgeDB have been selected as so-called 'Recommended Interoperability Resources' as part of the ELIXIR Interoperability Platform. Unfortunately, user statistics of these resources are not available in terms of individual users or access numbers, as - due to their nature - they are mostly embedded in other bioinformatics services. These services are used in increasing number of international projects and by mainstream bioinformatics tools.

Part of the ELIXIR services portfolio is provided as 'depletable' services, where access is limited due to limited available capacity, e.g. compute services or training and workshops. In the period 2016-2019 ELIXIR-NL (co-)organised 52 training courses (with a total of 166 days of training) that attracted 1061 participants (on average 65% of which are Dutch students/participants). In addition to training, ELIXIR-NL actively organises technical workshops and events that have a strong dissemination component. For example, in 2019, ELIXIR-NL (co-)organised 29 ELIXIR-related events: 4 conferences and 25 other workshops and events.

Users of ELIXIR services typically do not require to register to use them, hence the individual Nodes and thus the infrastructure have limited information on these. Exceptions exist, for instance in the case of training, for which career stage is recorded: for ELIXIR-branded training events during the period 2016-2019, 9.7% were Senior researchers, 73.3% were Students, 3.2% were Technical personnel, and 9.6% others. We do not gather specific characteristics of Dutch scientists using ELIXIR services. When we look at the participants of the ELIXIR-NL courses we see that between 60-70 % of the participants are PhD students (on average 40%) and postdocs (on average 20%). The majority of the participants is from Academia (85-90%) but there are also participants from Industry and non-profit organisations.

Application information
n/a

Comments RI (2020)
Users of ELIXIR services are not typically required to make an application to use them - one notable exception relates to depletable resources such as compute capacity when applications are submitted to local computing centres, but for databases, tools, interoperibility resources and online training (face to face training can require registration), users simply access the resource that they want when they want, without making an application. Users of non-depletable ELIXIR services are not typically required to apply to use them; for example, databases are freely accessible except when they contain possibly personal or sensitive information. Notable exceptions relate to depletable resources such as compute capacity and training (courses) where it makes sense to arrange for access requests.

Sample and data request information
Number of sample requests to the RI
n/a
Comments RI (2020)
ELIXIR is a virtual data infrastructure, hence no physical samples are kept; ELIXIR does not provide access to biomaterials.

Number of requests for data to the RI
n/a

Comments by the RI
ELIXIR, through its Nodes, provides 110+ data-related services. Datasets are freely accessible via the internet, without the need to formally request them. To give an idea of the scale of their usefulness to the global research community, usage of data in 19 ELIXIR Core Data Resources was estimated using numbers of "mentions" (name, accession number) and citations in scientific publications. Mentions increased from 41,000 to 51,000 (annually) between 2013 and 2017, and citations of 'key resource articles' increased from just under 3,000 to approx. 7,000 (annually) for the same period. The corresponding EMBL submission provides more detailed information on usage of many of the large archives run by EMBL-EBI and their usage from researchers in Sweden, Finland and Netherlands. To avoid double-counting, we do include the same duplicate statistics in this response, but refer you to the EMBL submission. The high usage of those resources is facilitated by engagement in ELIXIR, and ELIXIR's Commissoned Services (see Supplementary Information document) aim to connect scietists in ELIXIR Nodes with those running the EBI archives. Major archives are also run by other ELIXIR Nodes, and these also show extensive usage from scientists in Sweden, Finland and Netherlands. For example, the Human Protein Atlas (ELIXIR Sweden) and the Brenda and Silva databases (ELIXIR Germany) show high usage from the Netherlands. ELIXIR assembles individual data resources across Europe and has set up a schema of selection of so-called Core Data Resources. As published (Bioinformatics 36, 15 April 2020, p2636–2642), the ELIXIR CDR programme establishes a fundamental infrastructure for the life sciences. Access to these datasets is open, and therefore these datasets are not specifically requested. ELIXIR-NL has a strong focus on Interoperability, Training and Compute, and also a strong representation in ELIXIR User Communities. ELIXIR-NL not provide access to national datasets, yet. The Netherlands does have globally unique life science data collections, but in most cases these have been assembled under the framework of BBMRI-NL (clinical data collections and population cohorts) or other infrastructures (e.g. DiSSCo for biodiversity collections), and hence they strictly fall outside ELIXIR-NL. The parties responsible for these resources all assemble in DTL, and recently also in Health-RI (for parties in the biomedical field). At this level, ELIXIR-NL aligns Dutch ESFRI nodes active in the Life Sciences and strives to have these important data resources recognised as ELIXIR Core Data Resources in the future. In the future, we expect that access to datasets will be increasingly done through distributed analysis rather than aggregation, the "FAIR Data Train" model of analysis where local FAIR datasets are visited by analytics algorithms rather than being centralised. ELIXIR-NL has designed two applications of this approach:
1] Personal Health Train
2] Farm Data Train, implemented with ELIXIR-NL partner Wageningen University and Research for access to farm and agricultural data.
One notable exception to the above relates to sensitive human data - dataset requests are traditionally fed to a committee which regulates access to ensure ethical and legal compliance, with ongoing work to create electronic passports for identity and access management as part of ELIXIR's Authentication and Authorization Infrastructure (AAI). The centralised system for access sensitive data is run jointly by EMBL-EBI (Cambridge, UK) and CRG (Barcelona, ES).

Contributions provided by organisations or companies in the participating countries

Year	Contributions by organisations or companies in the participating countries (k€)		Dutch percentage out of contributions provided by organisations or companies	
	NL	Other countries	Academic	Non academic
2016	950	n/a	92	8
2017	1400	n/a	94	6
2018	1000	n/a	92	8
2019	1270	n/a	94	6

Comments by the RI
The ELIXIR Hub manages the membership contributions from particiapting Members. The financial accounts are open for all to see in each corresponding ELIXIR Annual Report (2019). However, due to the distributed nature

of ELIXIR , there is no concept of national-level (in-kind) contributions to the research infrastructure in the same way that a physical RI will receive in-kind contributions from members (for an overview of how ELIXIR is funded, please consult the website). What is possible to describe is ELIXIR's success with European Union funding, building on the early success of the ELIXIR-EXCELERATE project (2015-2019) which saw ELIXIR being the only life science research infrastructure to be prioritised for 19 million euro funding. The Supplementary Information sheet also provides details of the financial return to Sweden, Finland and Netherlands as a result of participating in ELIXIR's EU-funded grants, which is substantial.

The list of Dutch contributions to the ELIXIR services portfolio is described in the ELIXIR-NL Service Delivery Plan (SDP), and can be found through the ELIXIR website. An important source of international funding for our contribution to the ELIXIR technical infrastructure is through participation in international (EU-funded) projects. Around 1.7M€ of European (H2020 InfraDev calls) funding has been acquired for Dutch institutes through participation in European projects associated with, or coordinated by ELIXIR (e.g. ELIXIR-EXCELERATE, EJP-RD, EOSC-Life, and Helis Academy). Finally, to align and contribute services among nodes, ELIXIR-NL participates in so-called ELIXIR-funded Implementation Studies. Since 2015, ELIXIR-NL has participated in a coordinated manner in 39 out of 88 ELIXIR Implementation Studies. The investment in these ELIXIR-funded projects is not included in the figures above.

Total sum spent on other deliveries such as equipment, services and consumables
n/a

Comments by the RI
Such investments are not needed as part of ELIXIR.

Income from user fees
n/a

Comments by the RI
The services coordinated by ELIXIR are typically free to use, hence in the small number of cases where users do pay for ELIXIR services, these are for depletable resources such as compute capacity or data management consultancy, which would be handled by the Node providing those services rather than centrally by the Hub.

Additional questions to the RI
What is the Dutch contribution to the RI?
- ELIXIR-NL is actively involved in 4 out of 5 ELIXIR Platforms and co-direct the Platforms Interoperability and Training
- ELIXIR-NL is involved in 9 out of 12 selected ELIXIR User Communities (Federated Human Data, Human Copy Number Variation, Rare Diseases*, Proteomics, 3D-BioInfo, Intrinsically Disordered Proteins, Metabolomics*, Microbial Biotechnology*, Plant Sciences (* marks a co-lead role).
- ELIXIR-NL participated in 39 out of 88 ELIXIR Implementation Studies, while co-leading six of these.
- ELIXIR-NL participated in several major ELIXIR-related European projects, such as EXCELERATE and CONVERGE.
- ELIXIR-NL has leading roles in crucial international organisations, initiatives and projects with a direct link to ELIXIR, such as GOBLET, RDA, GOFAIR, EOSC, the 1+Million Genomes initiative and the European Joint Programme on Rare Diseases.
- ELIXIR-NL has played a pivotal role in developing and advocating the now globally adopted FAIR Data approach (launched early 2014, Leiden) underpinning the ELIXIR Interoperability Platform and resulted in the publication of the FAIR Principles in 2016 (Nature Scientific Data 2016 (3) 160018).
- The FAIR data approach has meanwhile resulted in a global movement to implement FAIR across disciplines. It has become a core element of Open Science programmes globally and is being adopted by numerous science funders worldwide (incl. NWO and ZonMw) as a key element of data management planning and FAIR-based data stewardship.

- ELIXIR-NL contributes actively to FAIR implementation and to ELIXIR-related training and capacity building on this topic (https://zenodo.org/communities/nl-ds-pd-ls/]) and has co-developed the Data Stewardship Wizard (https://ds-wizard.org).

Currently, are there any RI´s that provide similar kinds of research infrastructure and services as yours in the world?
As the ELIXIR Nodes collectively run more than 300 services (i.e. resources), some level of overlap with other service providers worldwide is likely. Some of this is by design: for instance and in the case of the ELIXIR Core Data Resources, these have for many years been created and developed within international partnerships. The role of ELIXIR is to improve coordination across its 22 Members to build, maintain and further develop infrastructure for molecular biology across national boundaries to ensure non-duplication of services and efficient use of the often limited funding available for bioinformatics services. Service can be freely used by users across the globe (with some exceptions). The main direct benefit of this Open Science approach is research efficiency, which translates to a more effective use of national-level financial support to research infrastructures in the participating countries. As a result, ELIXIR has been recognised as a research infrastructure of global significance by the Group of Senior Officials on Research Infrastructures of the G7. ELIXIR has also taken a leadership role in Europe through the coordination of cluster projects (e.g. CORBEL, EOSC-life), which gather numerous life science research infrastructures, and ELIXIR's role as the life science component of the European Open Science Cloud.

What are the overlaps and what are the main differences? To which extent do you cooperate or compete?
In the spirit of Open Science and its coordination, ELIXIR cooperates more than it competes with research infrastructures in the same field. ELIXIR has a leadership role in Europe through the coordination of cluster projects (e.g. CORBEL, EOSC-life which gather several life science research infrastructures. Looking beyond Europe, ELIXIR has launched a three-year collaboration strategy with the Australian BioCommons research infrastructure to exploit synergies between the two entities (2020). Another example of close cooperation is with the US National Institutes of Health, where ELIXIR will formalise a similar agreement with the NIH Data Commons. NIH funds some key resources in the ELIXIR network (e.g. UniProt, an international protein database, is run collaboratively by the Swiss Institute for Bioinformatics (SIB), EMBL-EBI and the PIR in Georgetown, USA). The key difference between ELIXIR and these two country-specific initiatives is the scale at which ELIXIR operates, (across several countries) and the number of institutes it coordinates.
ELIXIR is a founding partner of the Global BioData Coalition, which aims to better coordinate the funding, and hence long-term survival of international databases, core to scientific research across the globe. This initiative brings together global funding agencies including NIH, Wellcome, the European Commission, AMEC (Japan) to name a few.

What are the RI's major educational and outreach activities?
- User training is one of five domains of technical implementation for ELIXIR, under the so-called Training Platform, which was established to develop a training community that spans all ELIXIR Member countries. ELIXIR country Nodes run 24 training-related services. Particular highlights include the ELIXIR's training registry TeSS where users can find training events, workflows and materials, and ELIXIR's Training Metrics Database to access training quality and impact data for the thousands of ELIXIR training events. The ELIXIR Training Platform also supports the development of new training materials, selected according to yearly gap analysis on bioinformatics training needs across the ELIXIR Member countries.
- ELIXIR runs numerous events encompassing workshops, webinars, project meetings and conferences. In terms of outreach to the wider public, some of the ELIXIR Member countries run public engagement activities including annual conferences listed in the ELIXIR events page.