

Data Science Nodes

*the new national research infrastructure services
for data-driven life science - a local perspective*

Thomas Svensson, Senior Researcher & Support Manager

Dept. of Life Sciences, Chalmers

NBIS Support – Cell & Molecular Biology

What is data-driven life science?



Data-driven life science is a field of research that focuses on **using data, computational methods and artificial intelligence** to study biological systems and processes. This approach includes **assembling, sharing, integration and advanced analysis of large amounts of data** from diverse sources, including experiments, observations, and simulations, in order to gain a **better understanding of how living organisms function.**

What is data-driven life science? (ChatGPT)



Data-driven life science research refers to the **use of large datasets** and **advanced computational techniques** to gain insights into biological systems and processes. The goal of this type of research is to **identify patterns and correlations in the data that can help researchers better understand the underlying biology.**

The field of life sciences generates vast amounts of data through experiments, clinical studies, and other sources. This data can include genetic sequences, physiological measurements, imaging data, and more. By **analyzing this data using techniques such as machine learning**, researchers can identify new relationships and patterns that would be **difficult to detect using traditional experimental approaches.**

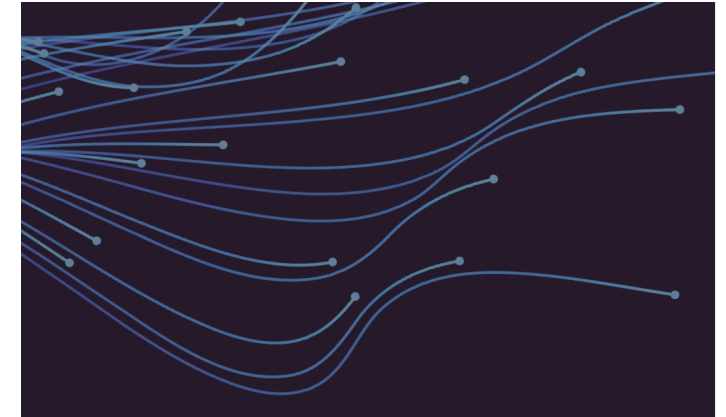
DDL S program overview



Research









Education and
Training



Data management
and data services

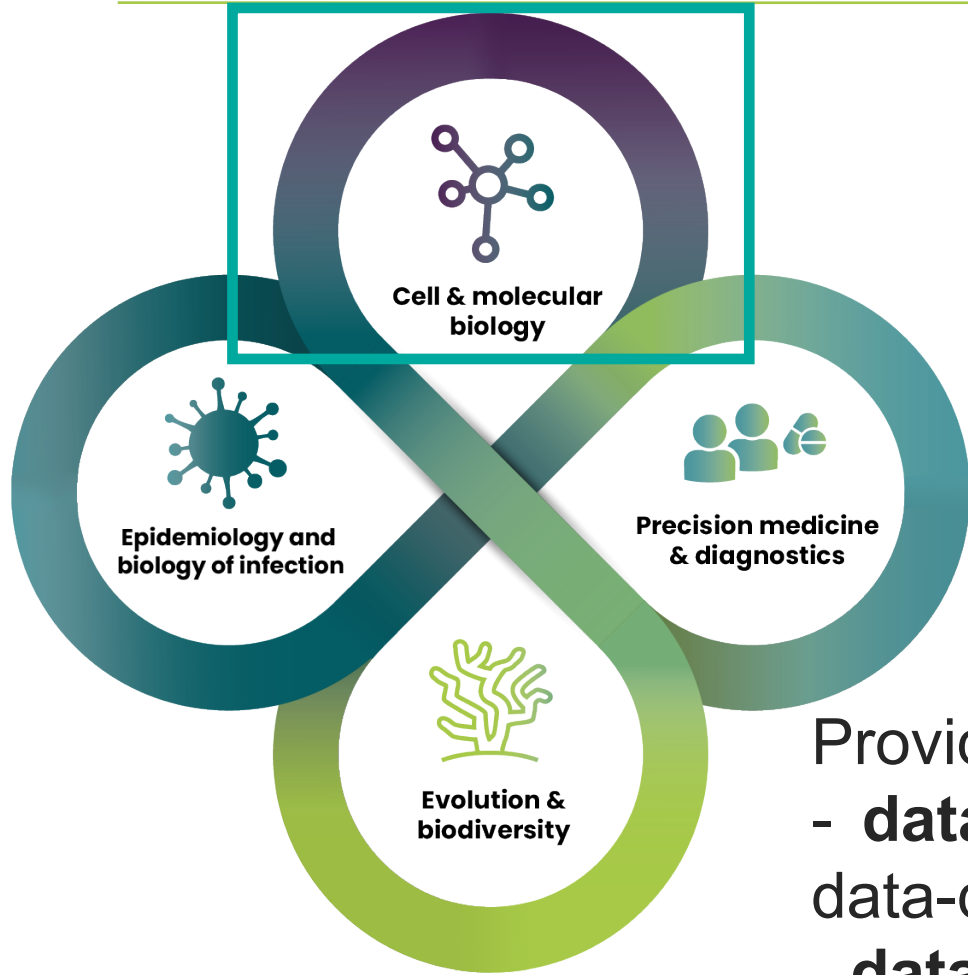
Overall 12-year plan for the DDLS program



Recruitment	 DDLS Fellows (incl. PhDs, post docs)	39 DDLS Fellows 78 PhDs and 78 postdocs
Education	*  PhDs and Industry PhDs	140 PhDs in academia and 45 industry PhDs
Training & networking	*  Post docs and Industry post docs	90 postdocs and 45 industry postdocs
Data, AI collab society impact	 WASP WASP-HS	210 MSEK WASP 35 MSEK WASP- HS
Bioinfo & comp capabilities	 WABI (incl. Cryo-EM)	235 MSEK
Data nodes, FAIR, AI, compute	 Data support, data bases	670 MSEK

* From 2024 - onwards

DDL S program - Data Science Nodes

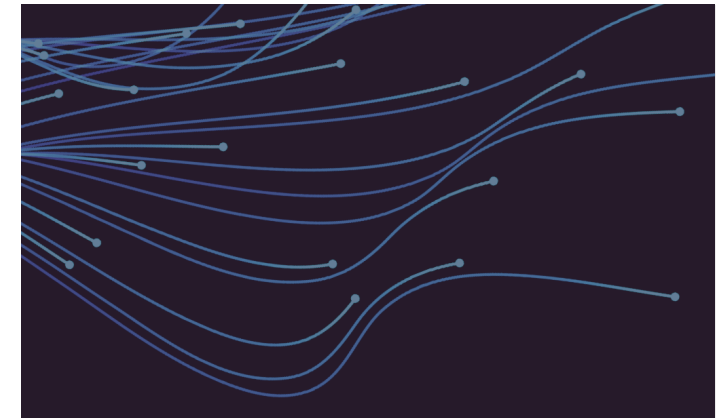


5 FTE per RA

- 2 NBIS/WABI
- 3 Data Center

Provide:

- **data support** to enable data-driven projects
- **data analysis support** in data-driven projects
- **added value** in data-driven projects



Data management
and data services

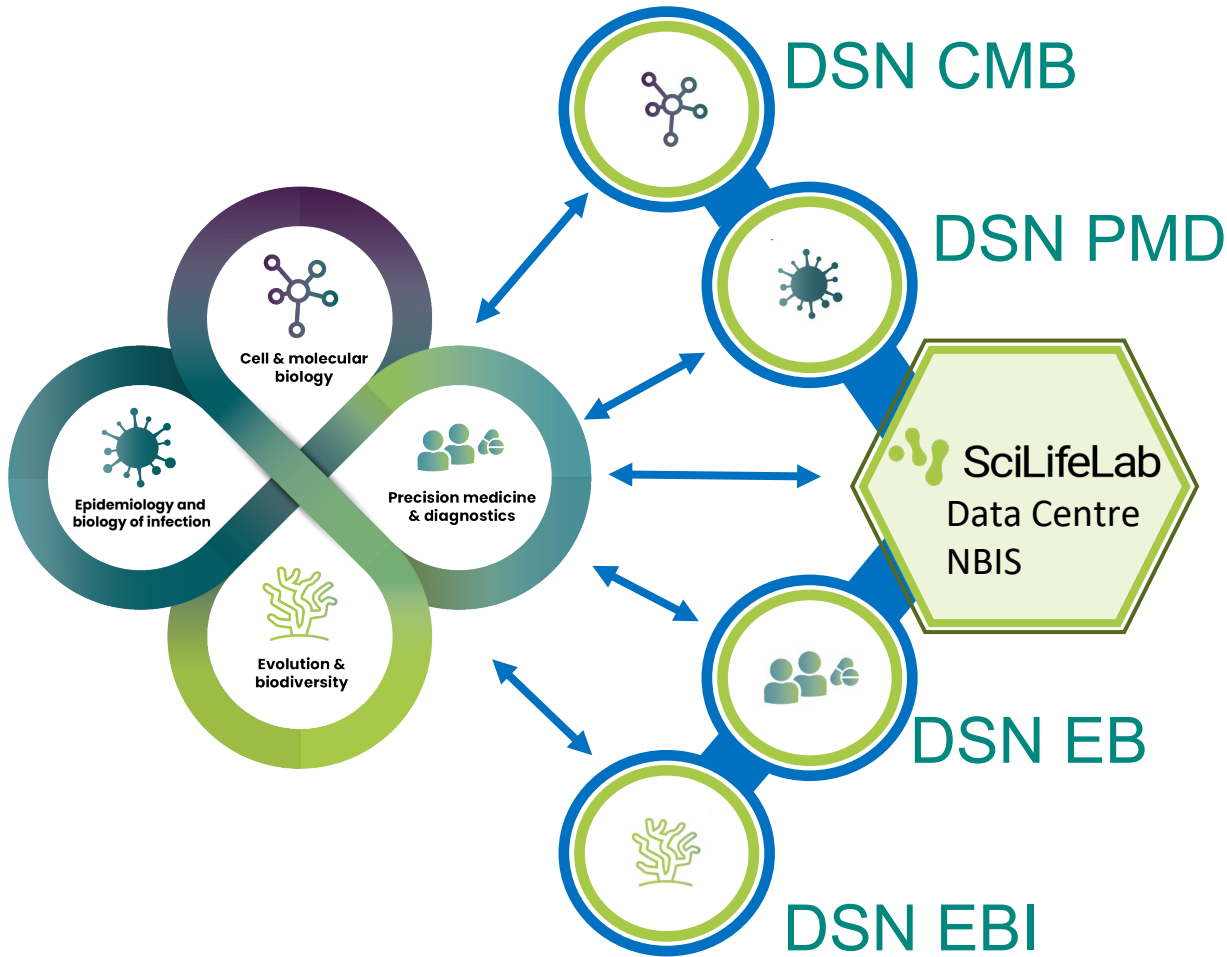
How to organize research infrastructure support for DDLs? (ChatGPT)



Organizing research infrastructure support for data-driven life science research can be a complex process, but there are several key components that are essential for success:

- ✓ - **Data management and storage:** One of the most critical components of data-driven research is the ability to store and manage large amounts of data. This requires robust data management systems that can handle complex datasets, ensure data privacy and security, and provide efficient data access and sharing.
- ✓ - **High-performance computing:** Data-driven research requires significant computational power to analyze large datasets. Therefore, access to high-performance computing resources such as clusters, clouds, or GPUs is essential to enable researchers to process large datasets quickly and efficiently.
- ✓ - **Software and tools:** Developing and using software and tools that enable efficient data analysis and visualization is crucial for data-driven research. Infrastructure support must provide access to such software and tools that allow researchers to work with complex data types and extract meaningful insights from the data.
- ✓ - **Expertise and training:** Data-driven research requires specialized expertise and training. The research infrastructure support must provide access to experts in data science, machine learning, statistics, and computational biology to enable researchers to develop and apply advanced analytical methods.
- ✓ - **Collaboration and communication:** Data-driven research often involves multiple stakeholders, including researchers, clinicians, data managers, and patients. Therefore, the research infrastructure support must provide mechanisms to facilitate collaboration and communication among these stakeholders, such as shared data platforms and online communication tools.
- ✓ **Overall**, organizing research infrastructure support for data-driven life science research requires a **multifaceted approach** that addresses the data management, computational, software, expertise, and collaboration needs of the researchers. By providing these essential components, the research infrastructure support can help to facilitate successful data-driven research that can lead to new discoveries and improved patient outcomes.

DDL S Data Science Nodes - National Services!



As a PI in data-driven research, interact with the relevant RA Expert groups.

There will be transparent processes set up in order to access support

You will get access to new and existing services via SciLifeLab. If you work in Gothenburg and don't find your way - talk to SciLifeLab Gothenburg Site (or ask ChatGPT...).

Research area specific expert groups



Data-Driven Precision medicine and diagnostics

Gunnar Cedersund, Linköping University
Sven Nelander, Uppsala University
Lars Klareskog, Karolinska Institutet
Johan Trygg, Umeå University
Patrik Georgii-Hemming, Karolinska Institutet
Päivi Östling, KI (adj. SciLifeLab Precision Medicine Capability lead)
Francis Lee (adj. WASP-HS representative in DDLS)
David Gisselsson Nord (adjunct as GMS representant)
[Janne Lehtiö, chair \(DDLS SG member\)](#)

Data-Driven Epidemiology and biology of infections

Joakim Dillner, Karolinska Institutet and Karolinska Univ. Hospital
Patrik Medstrand, Lund University
Birgitta Henriques-Normark, Karolinska Institutet
Erik Kristiansson, Chalmers University of Technology
Staffan Svärd, Uppsala University
[Oliver Billker, chair \(DDLS SG member\)](#)

Data-Driven Evolution and biodiversity

Christine Bacon, University of Gothenburg
Tanja Slotte, Stockholm University
Erik Svensson, Lund University
Matthew Webster, Uppsala University
Courtney W Stairs, Lund University
[Fredrik Ronquist, chair \(DDLS SG member\)](#)

Data-Driven Cell and molecular biology

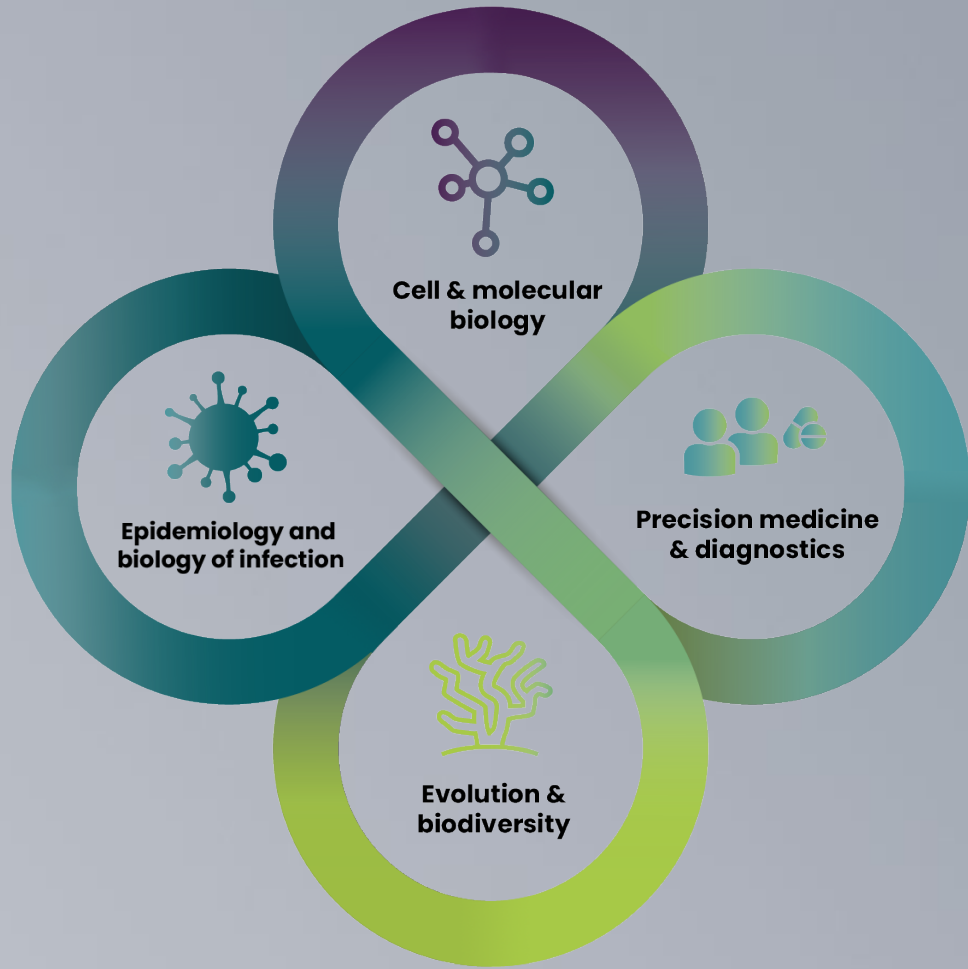
Arne Elofsson, Stockholm University
Ola Engkvist, AstraZeneca
Margit Mahlapuu, University of Gothenburg
Eduardo Villablanca, Karolinska Institutet
Marc Friedländer, Stockholm University
Ida-Maria Sintorn, Uppsala University
Sebastian Westenhoff, Uppsala University
[Erik Lindahl, chair \(DDLS SG member\)](#)



Vision & Mission

Vision: The future of life science is data-driven

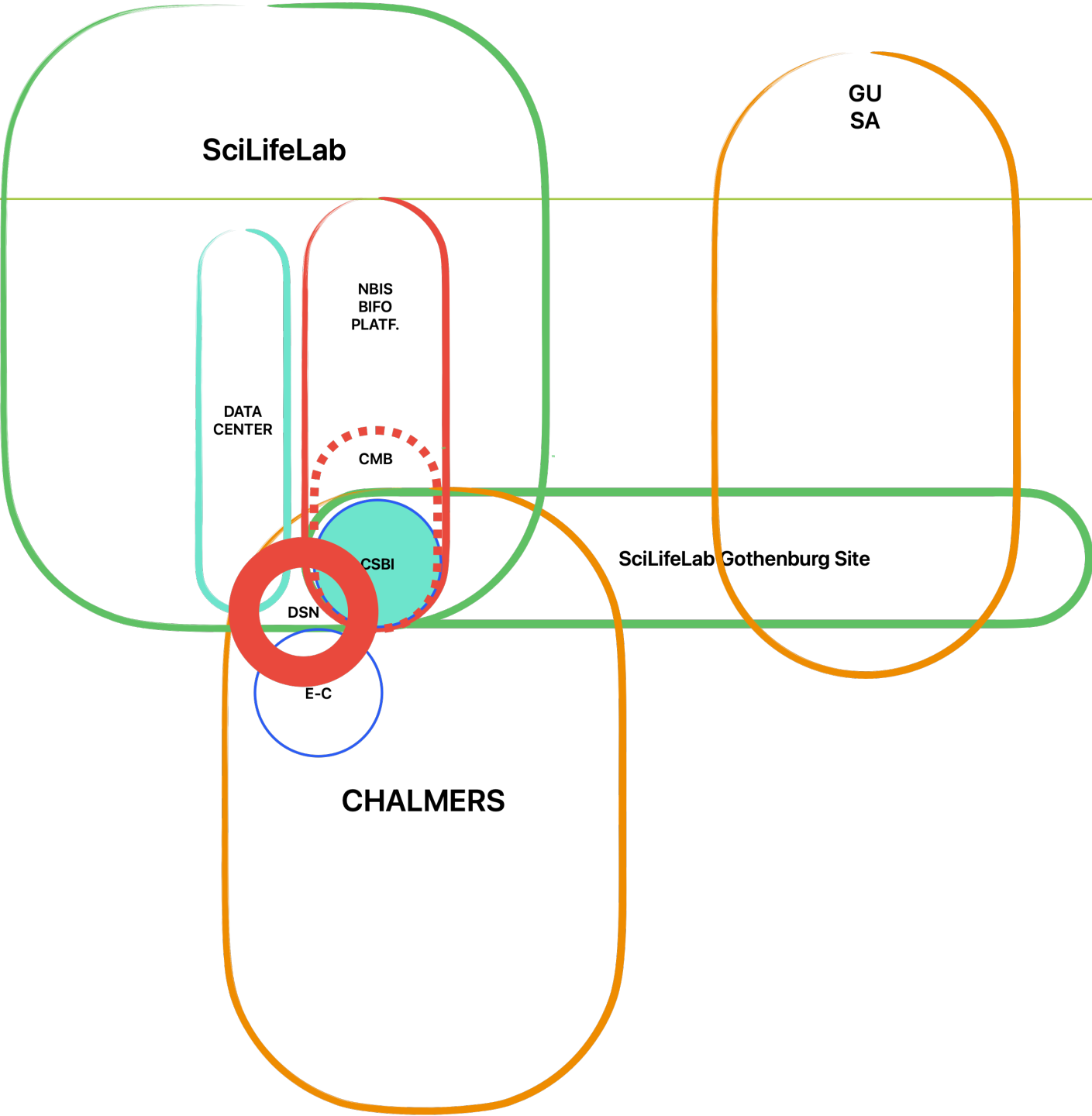
Mission: As a national research and training program across 11 partners, SciLifeLab & Wallenberg National Program for Data-Driven Life Science, DDLS, focuses on how data science and computational approaches enable and energize life science as well as applications in health, environment and industrial research.



Strategic research areas

Change how life science is practiced

Gothenburg context





CHALMERS