

Utmaningar och möjligheter med forskningsdata i arbetet med den svenska Covid-19-portalen

*2021-04-27*

Johan Rung

SciLifeLab Data Centre

[johan.rung@scilifelab.uu.se](mailto:johan.rung@scilifelab.uu.se)

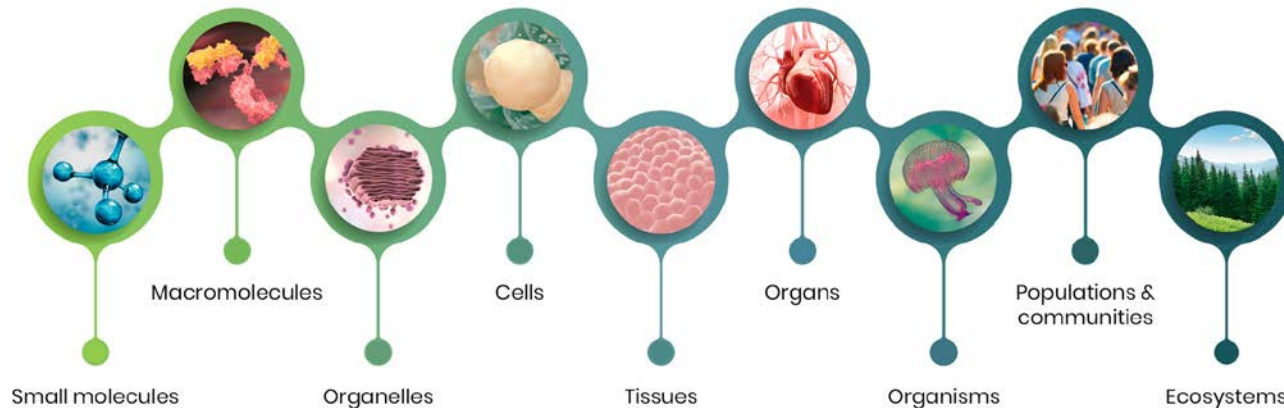
# Datadelning inom life science



- Första databasen med DNA-sekvenser 1980
- European Bioinformatics Institute 1995
- Spegling av sekvensdatabaser EU – USA – Japan
- EBI har idag 300 PB över hela life science-fältet
- Datadelning är en global drivkraft för life science
- FAIR som begrepp kommer från life science
- Reproducerbarhet och transparens



EMBL-EBI



# COVID-19-dataportalen



The screenshot shows the homepage of the COVID-19 Data Portal Sweden. The header includes the logo, navigation links (About, European Portal, Support Services, Submit data), and language options (en, sv). A main navigation bar lists categories like Genomics, Proteins, Imaging, Biochemistry, Health, and More. The main banner reads "Accelerating research through data sharing". Below this, there are sections for "Available data" with links to various data types (Genomics, Proteins, Imaging, Biochemistry, Health, Public health, Serology, Drug discovery, Other data, and Recent SARS-CoV-2 testing results). A "Data highlights" section features three featured articles with charts: "Proteomic profiling in COVID-19 patients as biomarkers to monitor disease severity", "Swedish COVID-19 symptom data contribute to accelerating research about pandemic", and "SARS-CoV-2 wastewater data from cities across Sweden now available". A pink box encourages sharing new COVID-19 data. The footer includes "External support services" with a Bioinformatics link.

**19,000+** besökare  
**95,000+** sidvisningar

**1200+** publikationer  
**170+** projekt med svensk finansiering  
**20** "data highlights" – nyheter med datafokus

Access till  
**530,894** virussekvenser  
**14,151** sekvenser från värdorganismer  
**10,000+** dataset för proteiner, bilddata, etc.

(2021-04-27)

Del av Europeiskt nätverk, koordinerat av EMBL-EBI  
<https://covid19dataportal.org>

Sverige första nationella noden (3:e juni 2020)

<https://covid19dataportal.se>

# Nationell COVID-19 dataportal



Plattformar



Forskare



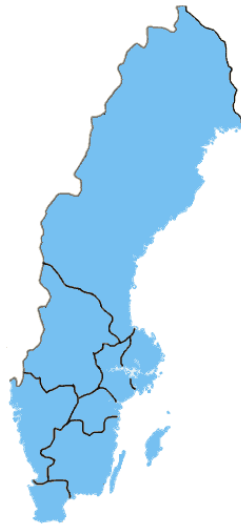
Hälsovård



Data-driven forskning



Industri och innovation



- Datatyper
- Tjänster
- Datavisualisering
- Nyheter med datafokus
- Svenska COVID-19-projekt
- Svenska publikationer
- Svenska studier med öppna data



# COVID-19-dataportalen



**Data types**

All data types (115)

**Genomics & transcriptomics data (16)**

Protein data (14)

Imaging data (4)

Biochemistry data (32)

Health data (60)

Drug discovery data (5)

Serology data (7)

Public health data (23)

Other data (13)

Showing 1 to 10 of 16 entries.

Project	Last updated
<b>Time series analysis and mechanistic modelling of heterogeneity and sero-reversion in antibody responses to mild SARS-CoV-2 infection.</b> Manisty C, Treibel TA, Jensen M, Semper A, Joy G, [...], Moon JC <i>EBioMedicine</i> <b>65</b> 103259 10.1016/j.ebiom.2021.103259 Abstract ▾	2021-03-01
<b>Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity</b> Fadista J, Kraven LM, Karjalainen J, Andrews SJ, Geller F, [...], None <i>EBioMedicine</i> <b>65</b> 103277 10.1101/2020.12.15.20248279 Abstract ▾	2020-12-16
<b>The MERS-CoV receptor gene is among COVID-19 risk factors inherited from Neandertals</b> Zeberg H, Paabo S <i>bioRxiv</i> 10.1101/2020.12.11.422139	2020-12-12

**COVID-19 Data Portal SWEDEN** Sample Collection Database [back to the COVID-19 Data Portal Sweden](#)

16 biobanks with 28 collections matching the search criteria

« < 1 2 > »

**Västerbotten biobank** Sample collection types: Healthcare sample collection  
Juridical person: Region Västerbotten

**Västra Götaland biobank** Sample collection types: Healthcare sample collection, Research sample collection  
Juridical person: Västra Götalandsregionen

Collection name	Type	Materials	Number of samples
<a href="#">Ascov-19</a>	Healthcare sample collection	Buffy Coat, cDNA / mRNA, DNA, microRNA, PBMC, Peripheral blood cells, Plasma, RNA, Whole Blood, Serum	1000 - 10.000
<a href="#">The Sahlgrenska COVID-19 biobank</a>	Research sample collection	Buffy Coat, Cerebrospinal fluid, Feces, microRNA, PBMC, Peripheral blood cells, Plasma, RNA, Serum, Urine, Whole Blood	1000 - 10.000

**Materials**

- Buffy Coat
- cDNA / mRNA
- Cell lines
- Cerebrospinal fluid
- DNA
- Feces
- microRNA
- Nasal swab
- Other
- Pathogen
- PBMC
- Peripheral blood cells
- Plasma
- RNA
- Saliva
- Serum
- Throat swab
- Tissue (frozen)

Svenska studier med öppna data

# COVID-19-dataportalen



**COVID-19 Data Portal SWEDEN**

Genomics Proteins Imaging Biochemistry Health More | Biobanks Research projects

## Publications on COVID-19 and SARS-CoV-19

Home / Publications on COVID-19 and SARS-CoV-19

This section presents a list of published scientific journal articles and preprints on COVID-19 and SARS-CoV-19 where at least one author has an affiliation with a Swedish research institute. Note that this is a manually curated database and as such may not be exhaustive. If you would like your publication to be added here or information about your publication to be corrected, please get in touch with us ([datacentre@scilifelab.se](mailto:datacentre@scilifelab.se)).

Last updated: 2021-01-28

Showing 1 to 7 of 7 entries (filtered from 908 records in total). Search:

Publication	Published
<b>Seropositivity in blood donors and pregnant women during 9-months of SARS-CoV-2 transmission in Stockholm, Sweden</b> Dopico XC, Muschiol S, Christian M, Hanke L, Sheward DJ, Grinberg NF, Bogdanovic G, Mcinerney GM, [...], Karlsson Hedestam GB <i>medRxiv</i> . DOI: 10.1101/2020.12.24.20248821 <b>preprint</b> <b>data available</b>	2020-12-26
<b>SARS-CoV-2 serological tests can generate false positive results for samples from patients with chronic inflammatory diseases</b> Kharlamova N, Dunn N, Bedri SK, Jerling S, Almgren M, Faustini F, Gunnarsson I, Rönnelid J, [...], Fogdell-Hahn A <i>medRxiv</i> . DOI: 10.1101/2020.11.13.20231076 <b>preprint</b> <b>data available</b>	2020-11-13
<b>SARS-CoV-2 exposure, symptoms and seroprevalence in healthcare workers in Sweden.</b> Rudberg A, Havervall S, Månberg A, Jernbom Falk A, Aguilera K, Ng H, Gabriëlsson L, Salomonsson A, [...], Thålin C <i>Nat Commun</i> <b>11</b> (1) 5064. DOI: 10.1038/s41467-020-18848-0 <b>data available</b>	2020-10-08
<b>Development, clinical translation, and utility of a COVID-19 antibody test with qualitative and quantitative readouts.</b> Bortz RH, Florez C, Lauderemilch E, Wirchnianski AS, Lasso G, Malonis RJ, Georgiev GI, Vergnolle O, [...], Chandran K <i>medRxiv</i> . DOI: 10.1101/2020.09.10.20192187 <b>preprint</b> <b>data available</b>	2020-09-11
<b>Blood type A associates with critical COVID-19 and death in a Swedish cohort.</b> Hultström M, Persson B, Eriksson O, Lipscey M, Frithiof R, Nilsson B <i>Crit Care</i> <b>24</b> (1) 496. DOI: 10.1186/s13054-020-03223-8 <b>data available</b>	2020-08-12

**Field/Topic**

- Biochemistry
- Drug discovery
- Genomics & transcriptomics
- Health
- Imaging
- Proteins
- Public health
- Serology
- Other

**Data availability**

- Data available

**Publication type**

- Journal article
- Review
- Preprint
- Other

**Year**

- 2020
- 2021

**Funder**

- H2020
- KAW/SciLifeLab
- NordForsk
- Swedish Research Council
- Vinnova

## Covid-19 Publications

This is a database of publications on SARS-CoV-2 and COVID-19 where at least one author is affiliated with a Swedish research institute. The database is curated manually and as such may not be exhaustive.

**Publications**

- 2021 (55)
- 2020 (854)
- All (909)

**Labels**

**Journals**

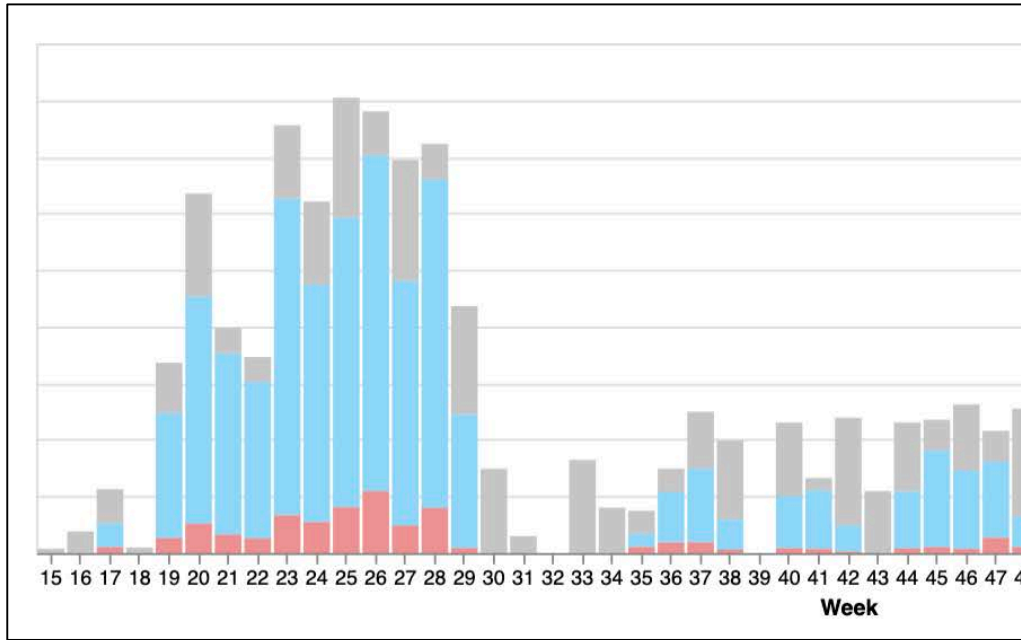
### The 10 most recent publications

- Comprehensive Landscape of Heparin Therapy for COVID-19**  
Shi C, Tingting W, Li JP, ..., Deng B, Zhang Y  
*Carbohydrate Polymers* **254** (-) 117232 [2021-02-15; online 2021-02-15]  
**Type: Review**
- Should neurosurgeons continue to work in the absence of personal protective equipment during the COVID-19 era?**  
Eijkholt M, Hulsbergen A, Muskens I, ..., Tisell M, Broekman M  
*Acta Neurochir (Wien)* - (-) - [2021-01-20; online 2021-01-20]  
pIi: 10.1007/s00701-021-04703-8  
**Type: Review**
- Nursing students' experiences of a pedagogical transition from campus learning to distance learning using digital tools.**  
Langegård U, Kiani K, Nielsen SJ, Svensson P  
*BMC Nurs* **20** (1) 23 [2021-01-19; online 2021-01-19]  
pIi: 10.1186/s12912-021-00542-1  
**Data type: Public Health** **Research Area: High-throughput and high-content serology** **Type: Journal article**

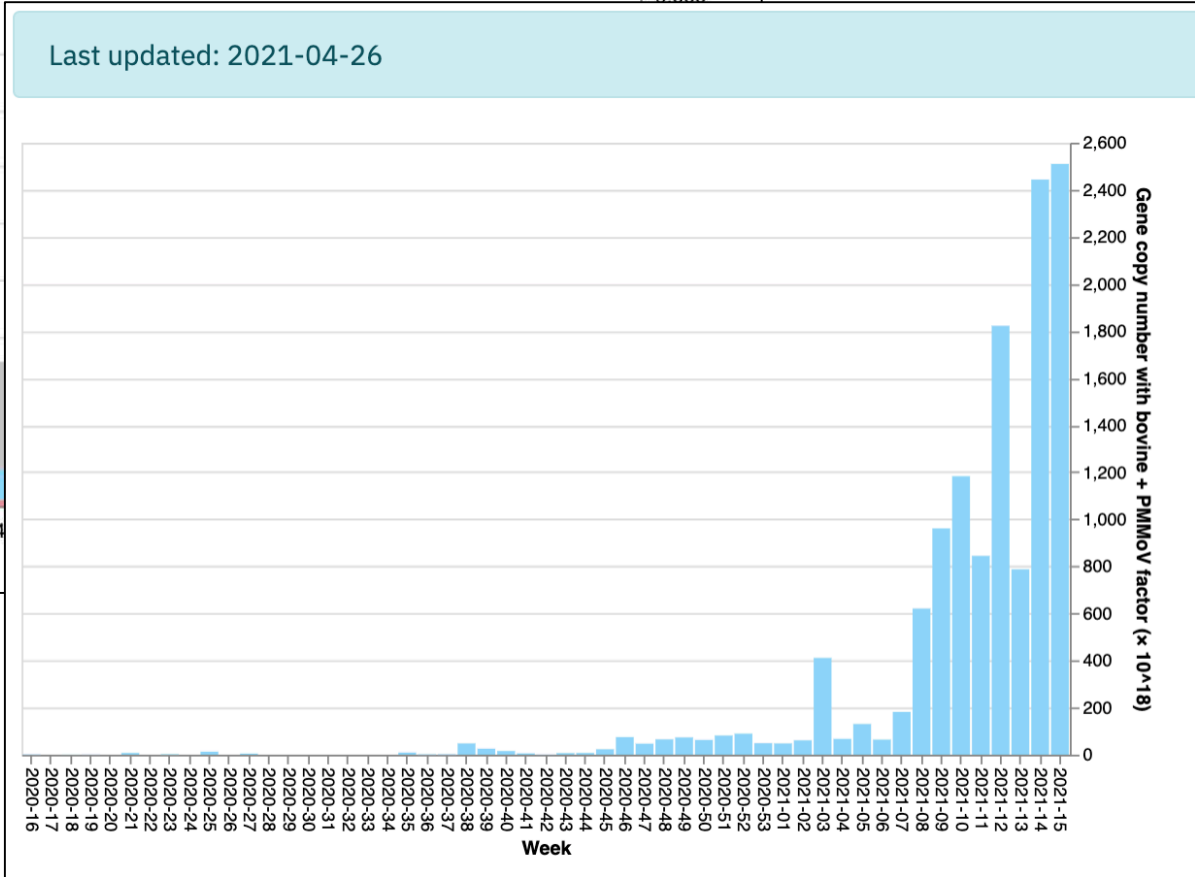
Publikationsdatabas och rapporteringsverktyg  
Open source: <https://github.com/pekrau/Publications>

Publikationer med svenskt deltagande  
(2021-01-29: N=1,254)

# Testdata



Antikroppstestning



Virusalt i avloppsvatten

# Nationella portaler bygger på öppen källkod från Sverige



The screenshot shows the GitHub interface for the repository `ScilifelabDataCentre/covid-portal`. The repository has 7 stars and 16 forks. The main content area displays a list of files and folders with their commit history:

File/Folder	Commit Description	Time Ago
<code>.github</code>	Update markdownlint.yml	last month
<code>DOCUMENTATION</code>	Update adding_new_data_highlights.md	last month
<code>archetypes</code>	Move mockup into a new Hugo site	12 months ago
<code>container</code>	Update Hugo to 0.80.0	3 months ago
<code>content</code>	Correct typo about Älvkarleby	20 hours ago
<code>data</code>	Pubs db update	4 days ago
<code>layouts</code>	Forgotten bit of text into Swe	11 days ago
<code>static</code>	Merge pull request #424 from ScilifelabDataCentre/csss-map-page-s...	13 days ago
<code>.gitignore</code>	Ignore backup files	12 months ago
<code>LICENSE</code>	add license	12 months ago

The right sidebar contains the 'About' section, which describes the repository as 'The Swedish COVID-19 Data Portal' and provides the website URL [www.covid19dataportal.se/](http://www.covid19dataportal.se/). It also lists tags: `bootstrap`, `markdown`, `hugo`, and `covid-19`. Below the 'About' section is the 'Contributors' section, showing 15 contributors with their profile pictures and a link to '+ 4 contributors'.

<https://github.com/ScilifelabDataCentre/covid-portal>



# Varför var Sverige först?

---



- Snabb svensk respons på Europakommissionens initiativ
- Snabbt, tydligt uppdrag för nationell koordinering
- Existerande nationell infrastruktur med naturlig roll och tillgängliga resurser och kompetens
- Välorganiserade forskningsprogram med uttalade villkor för öppna data och datahantering
- Finansiering

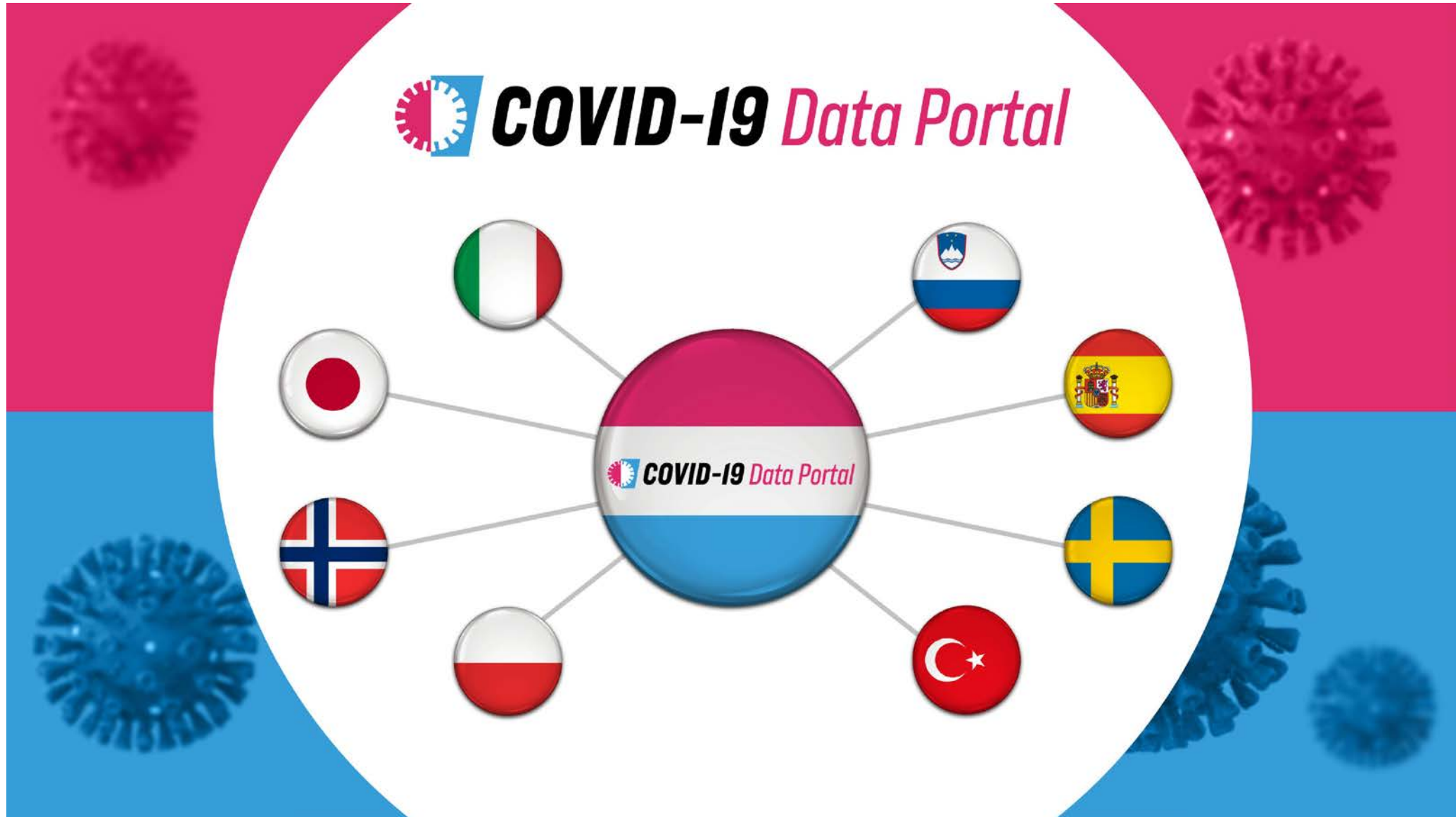
# Utmaningar och möjligheter?

Praktiska frågor

# Meta-data och standarder



# Internationell samverkan





# Delning av känsliga data



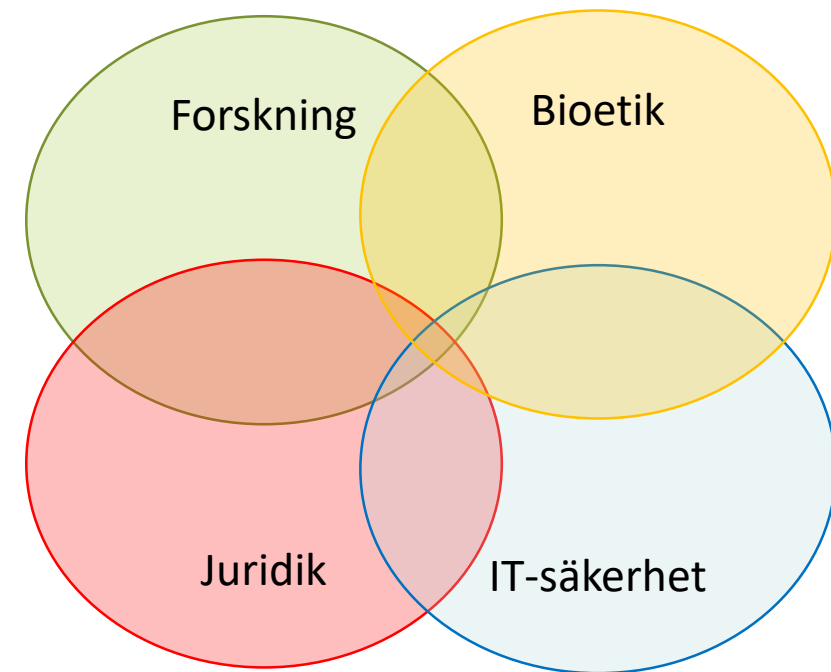
- Infrastruktur, t.ex. federerade system för lagring och beräkning
- Stödfunktioner och system för accesskontroll
- Nationell samordning
- Bioetik, t.ex. samtyckeshantering
- Longitudinella studier

## Här slängs proverna – som kan ge svaret på gåtan med corona

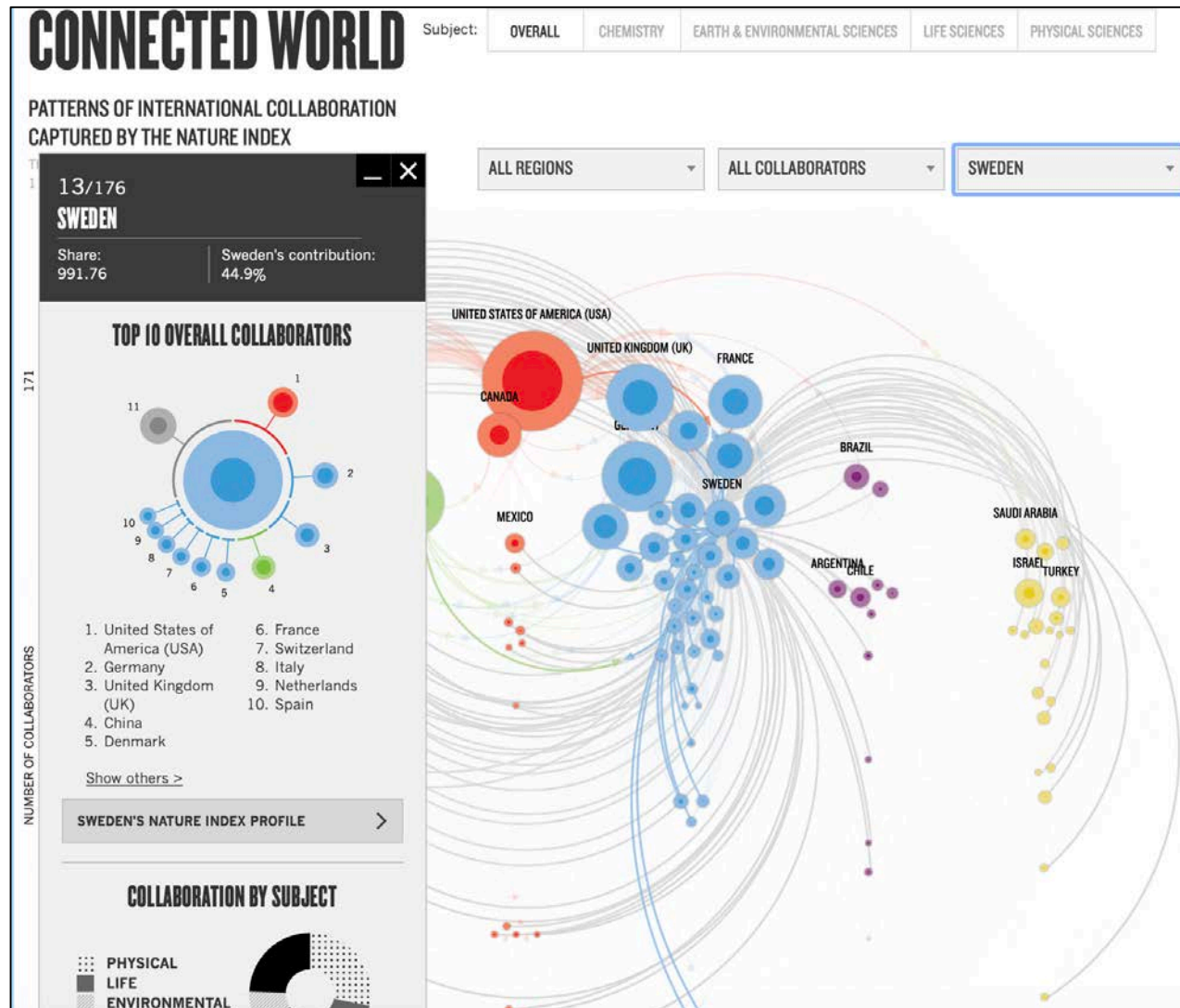
LENA HALLENGREN Publicerad 25 apr 2021 kl 18.49



Uppsala biobank Expressen 2021-04-25



# Resurser och tjänster som stödjer samverkan



Forskningen bedrivs i global samverkan

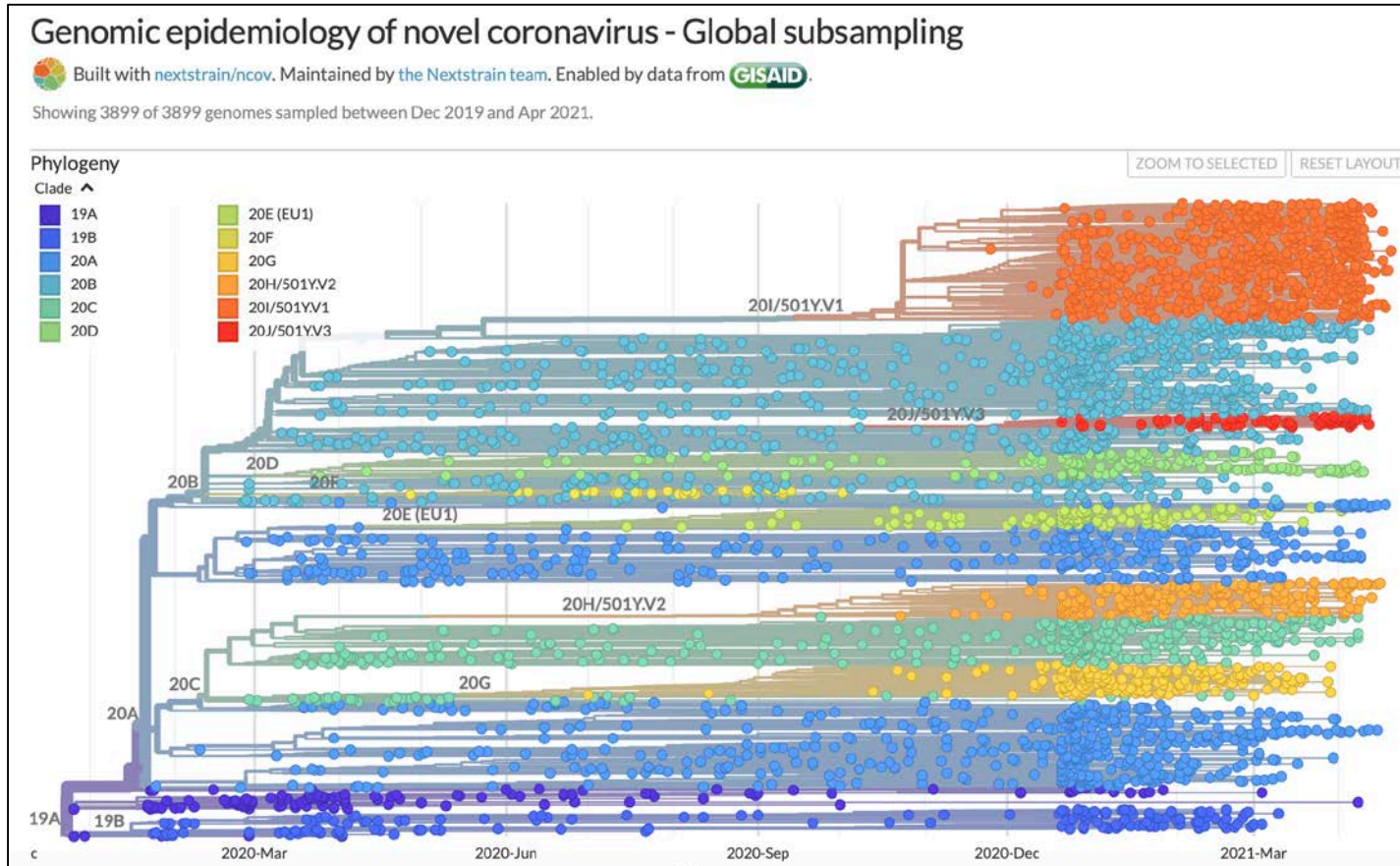
Stark nationell samverkan inom landet

...men behov finns för nationell samordning, och system för forskare vid sina lärosäten som möjliggör interaktioner med andra

Långa processer för att dela data även mellan lärosätena

Lagring

# Datadelning kräver resurser och motivation

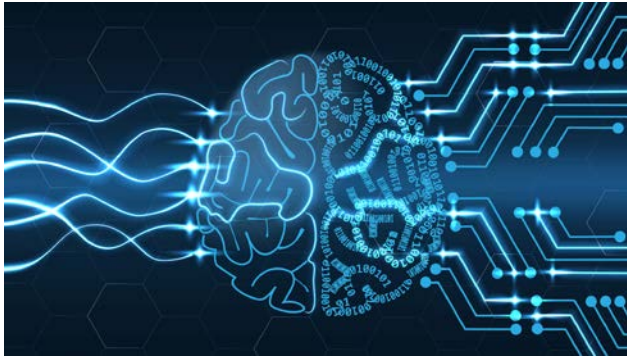


<https://nextstrain.org/ncov/global>

Öppen datadelning händer inte av sig självt.

- Tydliga uppdrag och datapolices
- Meritering
- Finansiering
- Ämnesspecifika, forskarnära stödfunktioner
- Möjlighet till direkt projektsupport
- Utbildning

# Framtidens life science är datadriven



Träning, delning, annotering och tillämpning av AI-modeller

Delning av kod, metoder och containers, och meta-data för dessa



Precisionsmedicin och hälsodata

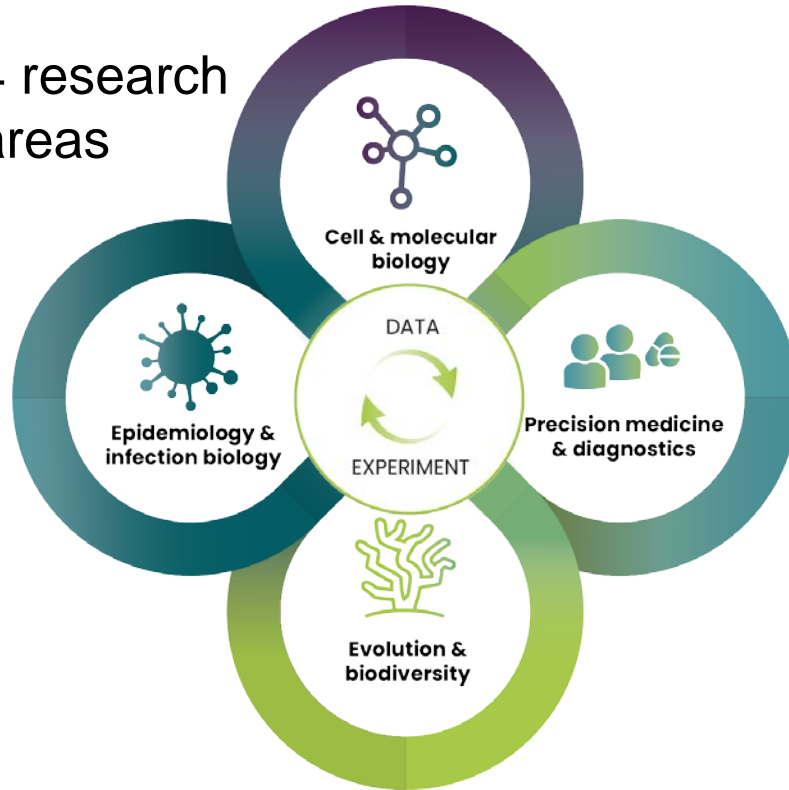
...och allt i en mycket större skala!



# Data-driven life science program: 12 years, 3.1 BSEK, 11 partners



4 research areas



- 1) National Framework for FAIR Data-Driven Life Science
- 2) Attract Scientific Excellence
- 3) Train the next generation of data driven life scientists
- 4) National research programs
- 5) Bridge the life science (DDLS) and data science (WASP)
- 6) Create partnerships and impact on society: industry, health care and other national and international links
- 7) DDLS community and policy actions

Knut och Alice  
Wallenbergs  
Stiftelse

 SciLifeLab

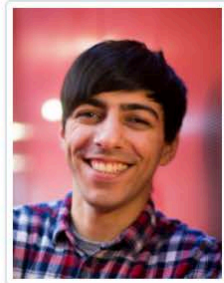
WALLENBERG CENTRES FOR MOLECULAR MEDICINE

WACQT | Wallenberg Centre  
for Quantum Technology

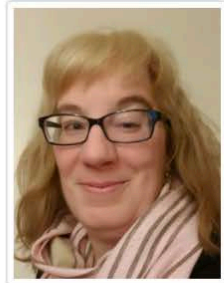
WASP  
WALLENBERG AI  
SOFTWARE SYSTEMS  
AND SOFTWARE PROGRAM



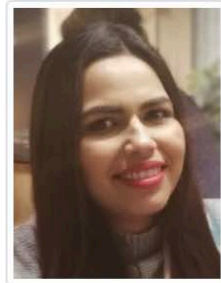
# COVID-19 dataportal



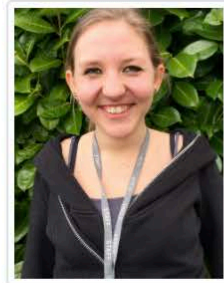
**Arnold Kochari**  
Webmaster for the  
COVID-19 Data  
Portal, SciLifeLab



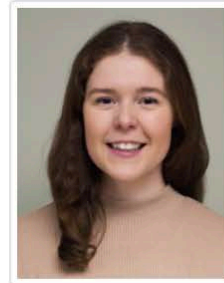
**Katarina Öjefors  
Stark**  
Coordinator for  
COVID-19 projects,  
SciLifeLab



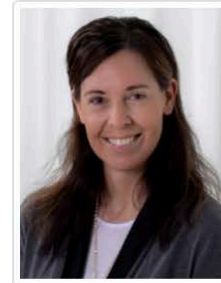
**Parul Tewatia**  
Data steward,  
SciLifeLab



**Liane Hughes**  
Data Engineer,  
SciLifeLab



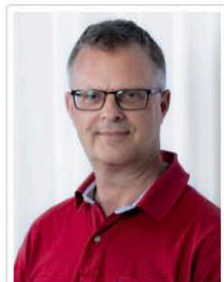
**Anna Asklöf**  
Data Steward,  
SciLifeLab



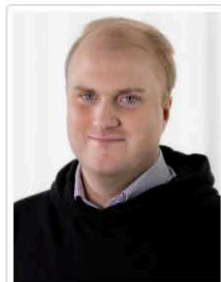
**Hanna Kultima**  
Data manager,  
coordinator,  
SciLifeLab



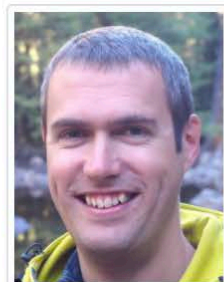
**Johan Rung**  
Head of the Data  
Centre, SciLifeLab



**Per Kraulis**  
System development  
team leader, data  
engineer, SciLifeLab



**Linus Östberg**  
Systems developer,  
system  
administrator,  
SciLifeLab



**Phil Ewels**  
Bioinformatics  
development lead,  
NGI

**Finansiering:**  
SciLifeLab  
Vetenskapsrådet

**Frågor?**  
[johan.rung@scilifelab.uu.se](mailto:johan.rung@scilifelab.uu.se)  
[datacentre@scilifelab.se](mailto:datacentre@scilifelab.se)