

Data Centre services and tools workshop

Fellows retreat

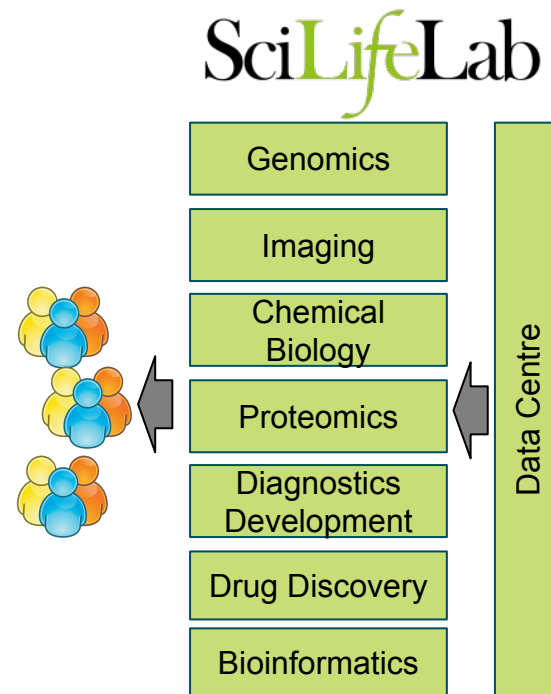
2020-10-20

Johan Rung, Hanna Kultima and Per Kraulis

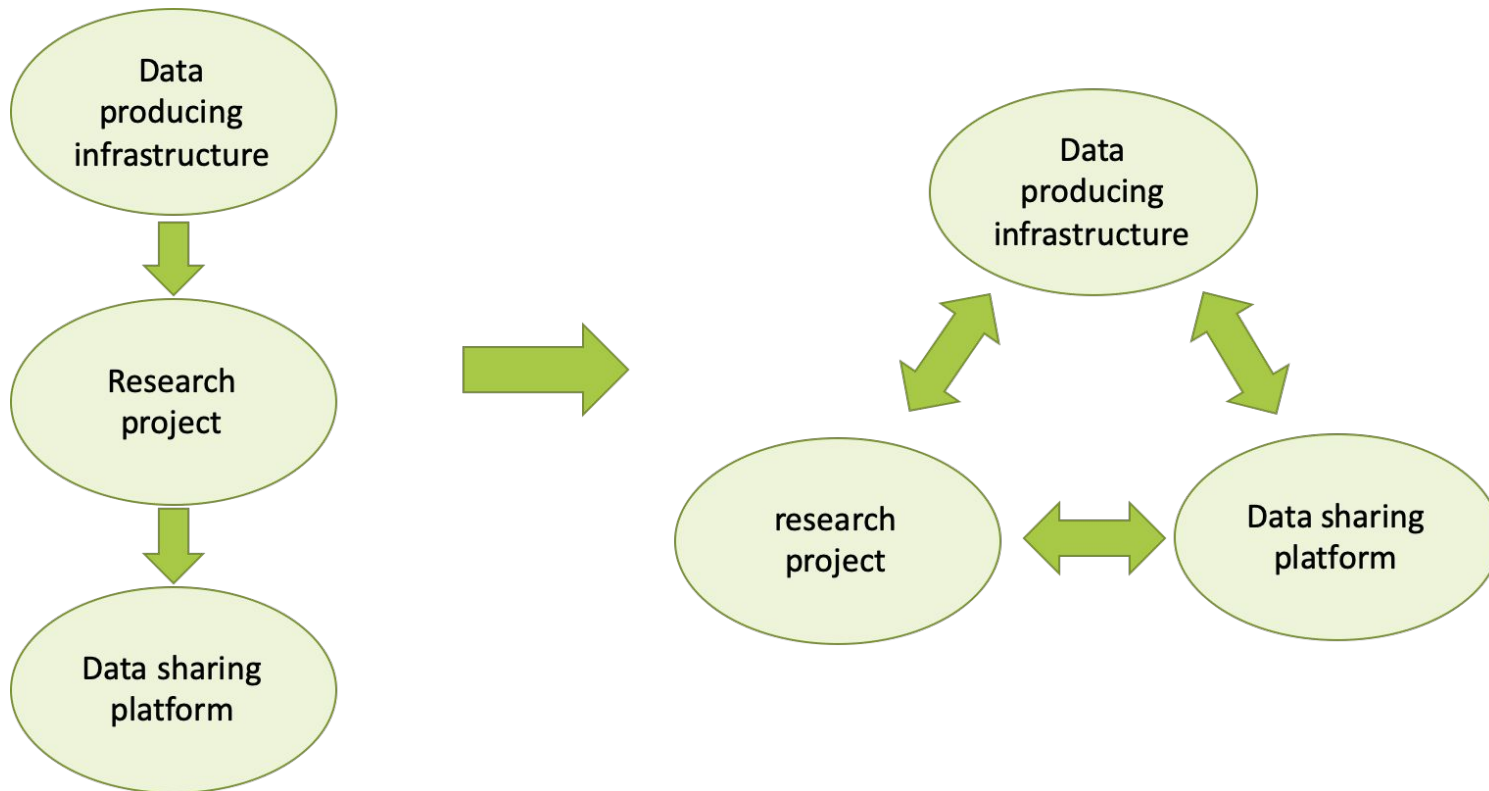
SciLifeLab Data Centre



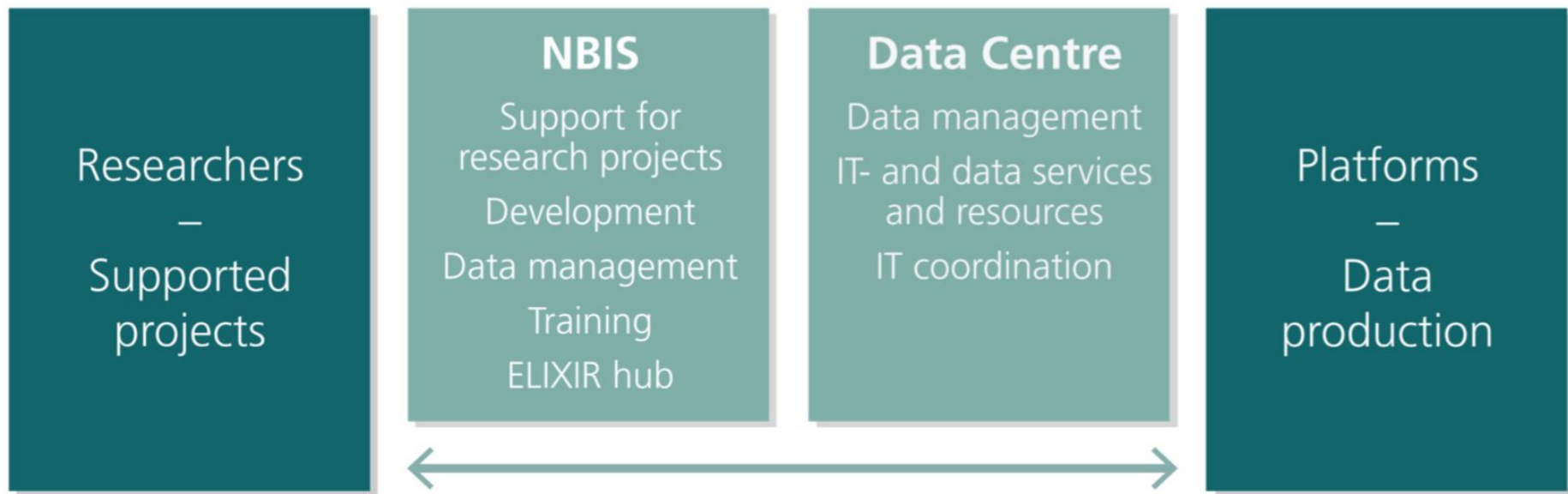
- Central support function in SciLifeLab, not a facility.
- Created in 2016 to address needs to make SciLifeLab produced data FAIR, as one of Sweden's first research data offices.
- Close connection between IT services to data producers and services for FAIR and research data management.
- Very positive feedback from IAB evaluation 2019
- Primarily supports data producers:
datacentre@scilifelab.se
- For bioinformatics and analysis support to research projects, contact NBIS: support@nbis.se



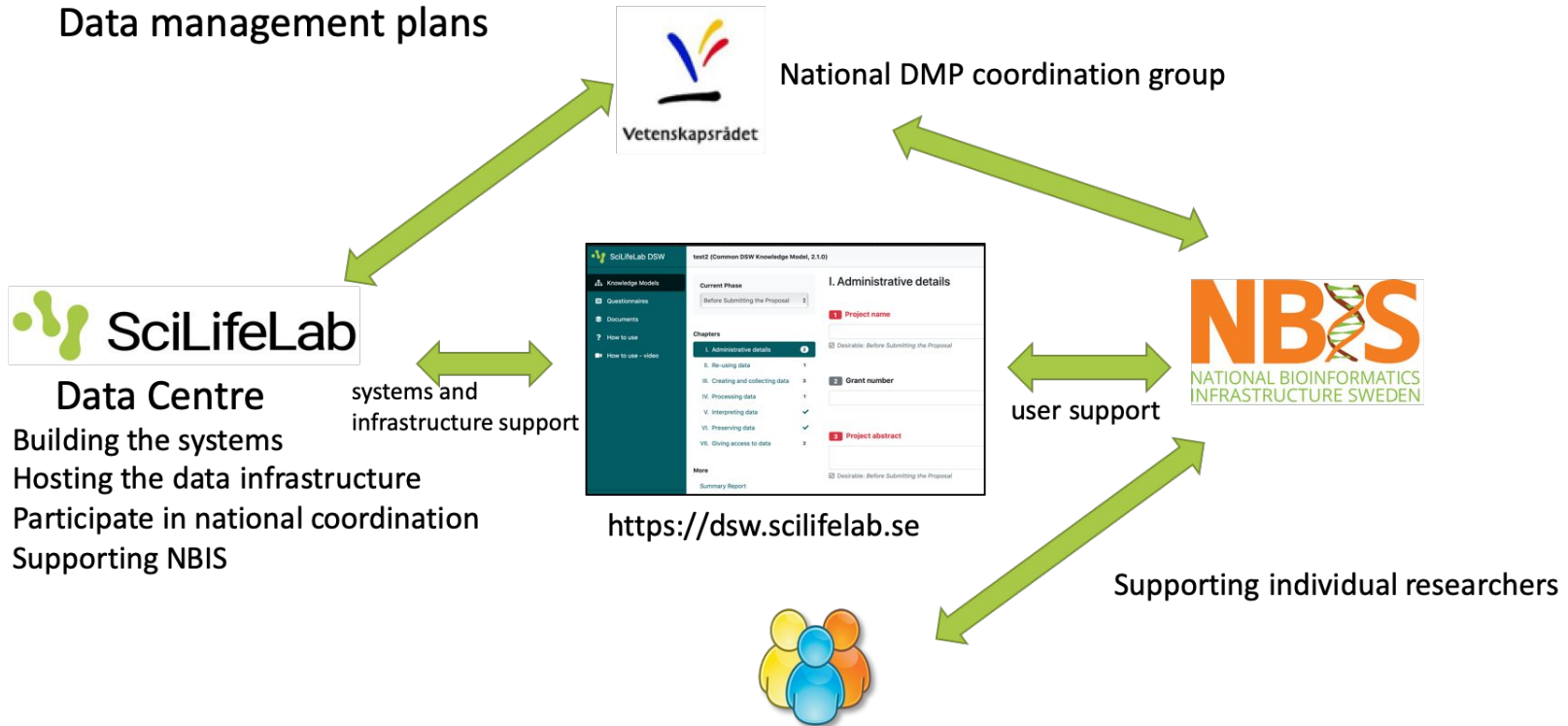
Bringing data production closer to data sharing



Data Centre / NBIS differences



Typical way of working DC-NBIS



Swedish Covid-19 Data Portal



The screenshot shows the homepage of the COVID-19 Data Portal Sweden. At the top, there is a navigation bar with the portal's name and logos, and links for 'About', 'European Hub', and 'Submit data'. Below this is a horizontal menu with categories: 'Genomics & transcriptomics', 'Protein data', 'Imaging data', 'Compound and target data', and 'Health data'. A large blue banner with the text 'Accelerating research through data sharing' is prominently displayed. Below the banner, a paragraph describes the portal's purpose: 'The Swedish COVID-19 Data Portal provides information, guidelines, tools and services to support researchers to utilise Swedish and European infrastructures for data sharing, in particular the European COVID-19 Data Portal.' A link 'För information på svenska, klicka här' is provided. At the bottom, there are two sections: 'Data Types' with a dropdown menu showing 'Genomics & transcriptomics', 'Protein data', and 'Imaging data'; and 'Support Services' with links to 'Bioinformatics', 'General data repository', and 'Resources'.

Structured around data types and support services in Sweden

Method circumventing RNA extraction in RT-PCR COVID-19 testing published along with data

Published: 2020-10-01

COVID-19 infection is most commonly tested using reverse transcription polymerase chain reaction (RT-PCR), but RNA purification constitutes a logistically demanding and costly prerequisite for conventional RT-PCR based diagnostics. To accommodate expansion of COVID-19 testing, researchers at the Karolinska Institute (PI: Björn Reinius; first authors; Ioanna Smyrlaki, Martin Ekman) developed a simple, cheap, and RNA extraction-free RT-PCR-based test, retaining high accuracy in identifying positive and negative cases. The method has broad implications to simplify COVID-19 testing within as well as outside health-care-facility settings and for periodic testing of asymptomatic people, benefitting patient care and infection control.



Source: Figure 1 from Smyrlaki et al., Nature Communications, 2020

From start, the researchers continuously shared their results on the medRxiv.org pre-print server (first version April 17th 2020), gaining considerable attention. Data from the paper is freely available on [Nature Communications](https://www.nature.com/articles/s41467-020-18611-5) and computational code is openly provided at github.com/reiniuslab/COVID19.

The project is part of the SciLifeLab National COVID-19 Research Program founded by Knut and Alice Wallenberg Foundation and SciLifeLab. Additionally supported by the Swedish Research Council and Ragnar Söderbergs Stiftelse.

Article

DOI: [10.1038/s41467-020-18611-5](https://doi.org/10.1038/s41467-020-18611-5)

Smyrlaki I, Ekman M, Lentini A, de Sousa N R, Papanicolaou N, Vondracek M, Aarum J, Safari H, Muradrasoli S, Rothfuchs A G, Albert J, Högberg B, Reinius B. Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-PCR. *Nat Commun* **11**, 4812 (2020)

Data

Link to source data available under the DOI above.

Data sharing focused news

Swedish Covid-19 Data Portal

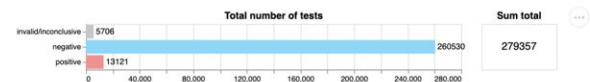


National Pandemic Centre SARS-CoV-2 (COVID-19) test statistics

Home / Data Types / Health data / National Pandemic Centre SARS-CoV-2 (COVID-19) test statistics

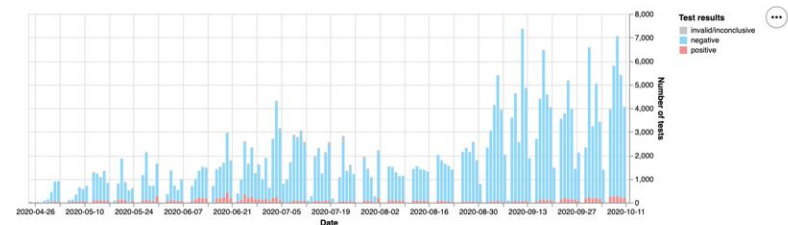
Total NPC test numbers

The total number of SARS-CoV-2 (COVID-19) tests run at NPC since the start, split up into positive, negative and invalid/inconclusive results.



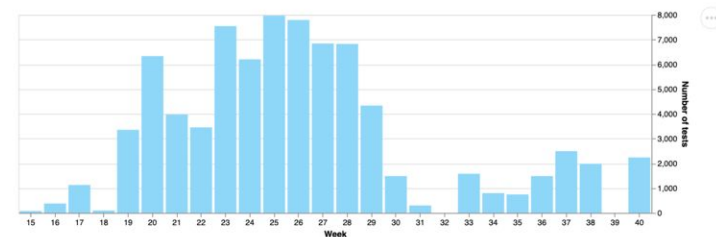
Daily NPC test numbers

The number of SARS-CoV-2 (COVID-19) tests run daily, split up into positive, negative and invalid/inconclusive results.



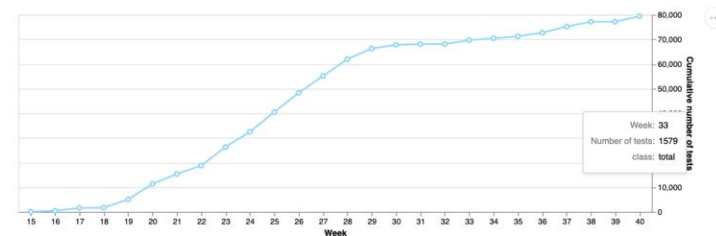
Weekly serology test numbers

The number of tests run at SciLifeLab Autoimmunity and Serology profiling facility weekly.



Cumulative serology test numbers

The sum of all tests run at SciLifeLab Autoimmunity and Serology profiling facility since the start.



Updated test statistics from virus sequencing and serology testing labs

Swedish Covid-19 Data Portal



The screenshot shows the 'Submit data' page of the COVID-19 Data Portal Sweden. The header includes the portal logo, navigation links (About, European Portal, Support Services, Submit data), and language options (en, sv). A secondary navigation bar lists data types: Genomics & transcriptomics, Protein data, Imaging data, Compound and target data, and Health data. The main heading is 'Submit data'. Below it, a breadcrumb trail reads 'Home / Support Services / Submit data'. The text explains that users should publish their COVID-19 research data in a public repository with descriptive metadata. It provides contact information for SciLifeLab (datacentre@scilifelab.se) and NBIS (support@nbis.se) for consultations. It also mentions the European Bioinformatics Institute (EBI) as a repository. A section for 'Support for data management planning' describes how SciLifeLab and NBIS can assist with data management plans and standards. A link to 'Request Data Management consultation from NBIS' is provided at the bottom.

Data submission guidelines and support

The screenshot shows the SciLifeLab website. The header features the SciLifeLab logo and a navigation bar. The main heading is 'Discover research from Science for Life Laboratory'. Below this, there are statistics: 4,289 views, 307 downloads, and a link to 'more stats...'. A navigation bar includes 'ALL', 'CATEGORIES', 'GROUPS', and 'SEARCH'. A 'sort by: Posted date' dropdown is visible. The main content area displays four dataset cards: 'StratFreshDB v1.0' (posted 12.10.2020 by Moritz Buck), 'Talk at Alicante's metagenomic forum' (posted 21.09.2020 by Moritz Buck), 'Genome Informatics 2020 --- Poster' (posted 18.09.2020 by Moritz Buck), and 'Improved computational analysis of ribosom...' (posted 17.09.2020 by Vicent Pelechano).

Institutional data repository (Figshare)

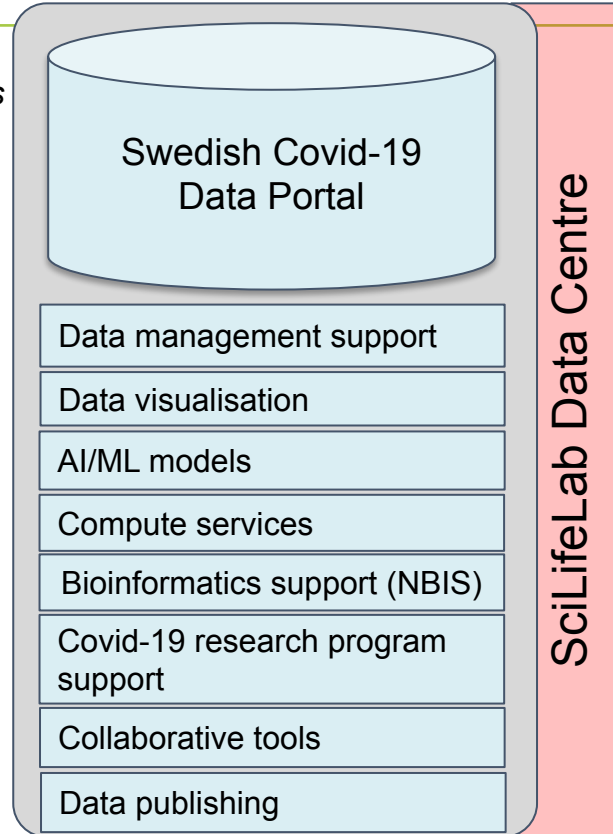
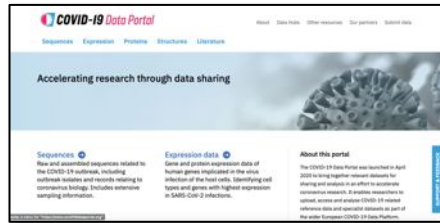
<https://covid19dataportal.se>

SciLifeLab Covid-19 Data Portal



Virus analyses
Antibody analyses
Epidemiological studies

External relations
Industry
Health care



SciLifeLab Covid-19 research program

Technology platforms

Genomics

Imaging

Chemical
Biology

Proteomics

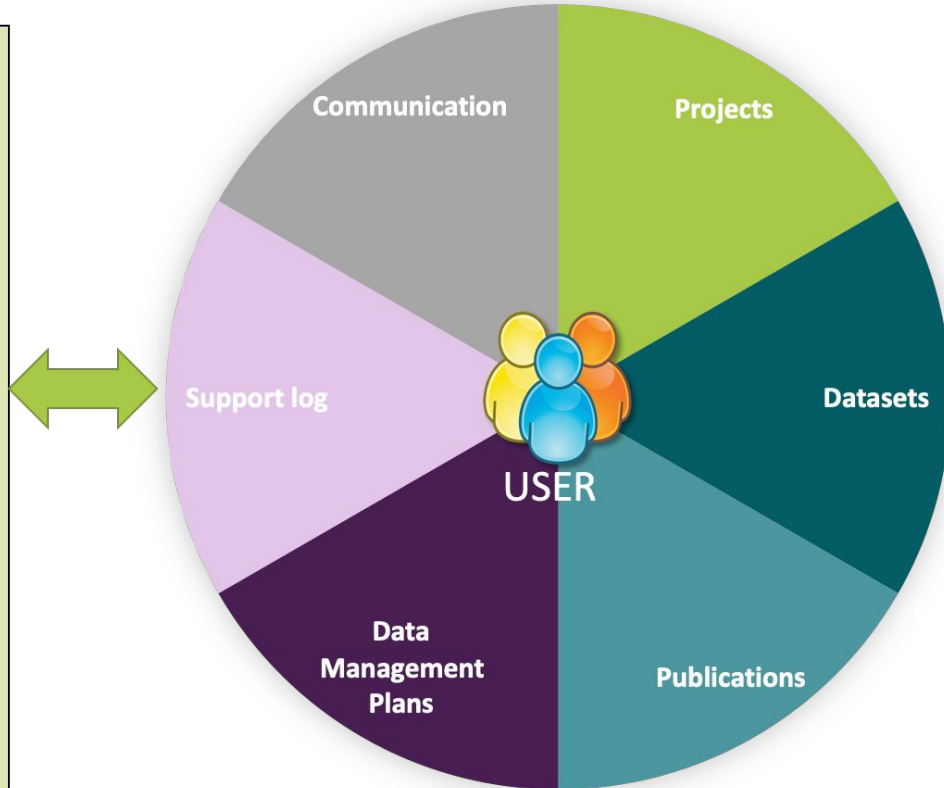
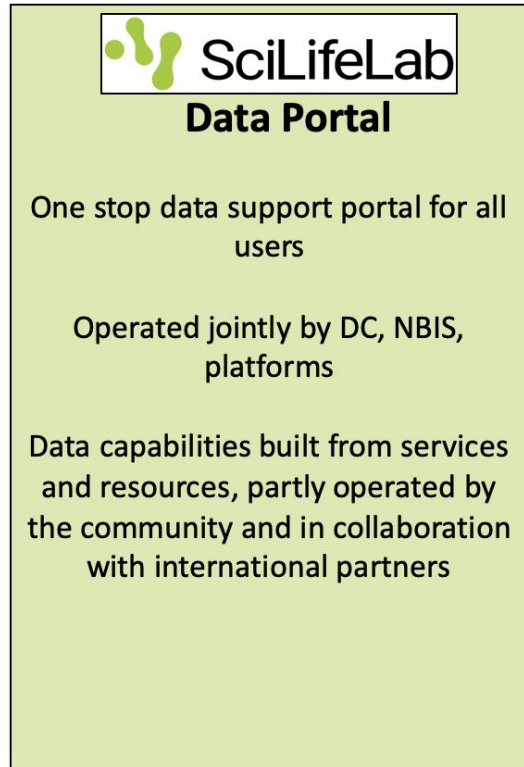
Diagnostics
Development

Drug Discovery

Bioinformatics



Future plan: SciLifeLab Data Portal



Future plan: SciLifeLab Data Portal



Platform data services



Systems, tools and support to platforms

Collaborative tools



Tools and systems for team science

International archives



Submission pipelines and technical collaboration

Bioinformatic services



Support, training, infrastructure

SciLifeLab Data Resources



Repositories for data sharing and publishing

ELSI and sensitive data services



Tools and resources for sensitive or clinical data

Supported Access to Computational services



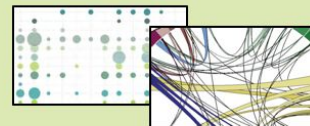
...and close coordination with university IT

Hosted Data Resources



Community created resources and services

Data modeling and visualization



Tools for AI and data mining and exploration

Research data lifecycle



- The research data life cycle illustrates the different stages the research data goes through, from planning to data-use.
- Research Data Management spans over all stages of the research cycle

Overview



- Planning
 - Data Management Plans
- Project tools
 - Slack
 - Confluence
 - Nextcloud
 - Storage
- Publications
 - Publication Database
 - Figshare
 - EBI services
 - Hosting @ DC
 - DbShare
 - DataGraphics

Data management plans



- A data management plan describes how the research project will generate/collect data, how it is to be handled, organized, stored and published. A DMP also describes how data will be retained and archived after the project has ended.
- All who are awarded a grant from the Swedish Research Council must have a data management plan if the research generates research data. The plan should address the following aspects of the research project:
 - Description of data
 - Documentation and data quality
 - Storage and backup
 - Legal and ethical aspects
 - Accessibility and long-term storage
 - Responsibility and resources
- The universities provide the digital tool DMPonline where researchers can create, update and share data management plans. Employees at the university can create an account in DMPonline.

DSW – Data Stewardship Wizard *demo to come*

The screenshot displays the SciLifeLab DSW (Data Stewardship Wizard) interface. On the left is a dark teal sidebar with navigation links: Users, Knowledge Model Editor, Knowledge Models, Projects, Documents, Templates, How to use, How to use - video, Settings, Help, and Hanna Kultima. The main content area is titled 'ResearchProjectA' and includes a 'Share' button. Below the title are tabs for Questionnaire, TODOs, Metrics, Preview, Documents, and Settings. The 'Current Phase' is set to 'Before Submitting the Proposal'. A 'Chapters' list on the left shows 'I. Overview' selected, with other chapters (II-VIII) marked with checkmarks. The 'I. Overview' section contains three numbered steps: 1. Title of the project, 2. Aim of the project, and 3. Revision date. Each step has a text input field and a checkbox labeled 'Desirable: Before Submitting the DMP'. The interface is clean and modern, with a light blue header and a white main content area.

Create data management plans for your project <https://dsw.scilifelab.se>

Slack – chat style team communication



Browse Channels

Create Channel

Search channels

Show: All channels Sort: Channel name

#irma-nestor

Chat for people working on nestor and irma

Created by Phil Ewels on September 1st, 2014

#meetups

Discussion and planning for events in the Uppsala/Stockholm Research Engin...

Created by Samuel Lampa on January 22nd, 2015

#metagenomics

General purpose channel for metagenomics discussion on methods, database...

Created by Fredrik Boulund on January 20th, 2017

#microbiome

Everything microbiome related

Created by Mauricio Barrientos on May 7th, 2018

#mip

discuss MIP related issues

Created by Kenny Billiau on January 16th, 2019

#mushroomhunting

NGI Monthly Social Even Coordination

SciLifeLab

Jump to...

More unread

#scilifelab-tracking-gr...

#swefreq-development

#uppmix

Channels

#covid_symptoms_grant

#covid-19-informatics

#covid19-confluence-con...

#covid19-diagnostics-ma...

#covid19-discourse-confi...

#covid19-informatics-coor...

#covid19-mailredaktion

#covid19-serology

#datacentre_internal

#datadriven

#figshare

#gamma-5

#general

#mg-corona

#music

#navet

#random

#uppnex

Direct Messages

Slackbot

#general

8,750 | 1 | Big announcements that everyone needs to see.

Tuesday, March 24th

Sebastian DiLorenzo 10:06 AM

FYI: Due to the current Covid-19 situation the NBIS Bioinformatic drop-in sessions will be held online for all sites. Visit www.nbis.se/events for all the information you need to join and talk to our bioinformatics experts. We have three sessions starting within the next hour!

10:00 Umeå

10:30 Stockholm

11:00 Linköping

Artem Zhmurav 10:14 AM

joined #general along with 41 others.

Friday, March 27th

Phil Ewels 9:42 AM

Please use #covid-19

Carl Nettelblad 9:54 AM

joined #general along with 5 others.

Phil Ewels 8:35 PM

Please post COVID-19 related posts in the #covid-19 channel, or one of it's siblings. This channel is called #general and is for announcements only (it's the only slack channel that people cannot remove themselves from and has 740 people listening as I type this). Thanks!

Saturday, March 28th

Mohammad 6:44 PM

joined #general along with 5 others.

Message #general

Confluence – collaborative workspaces



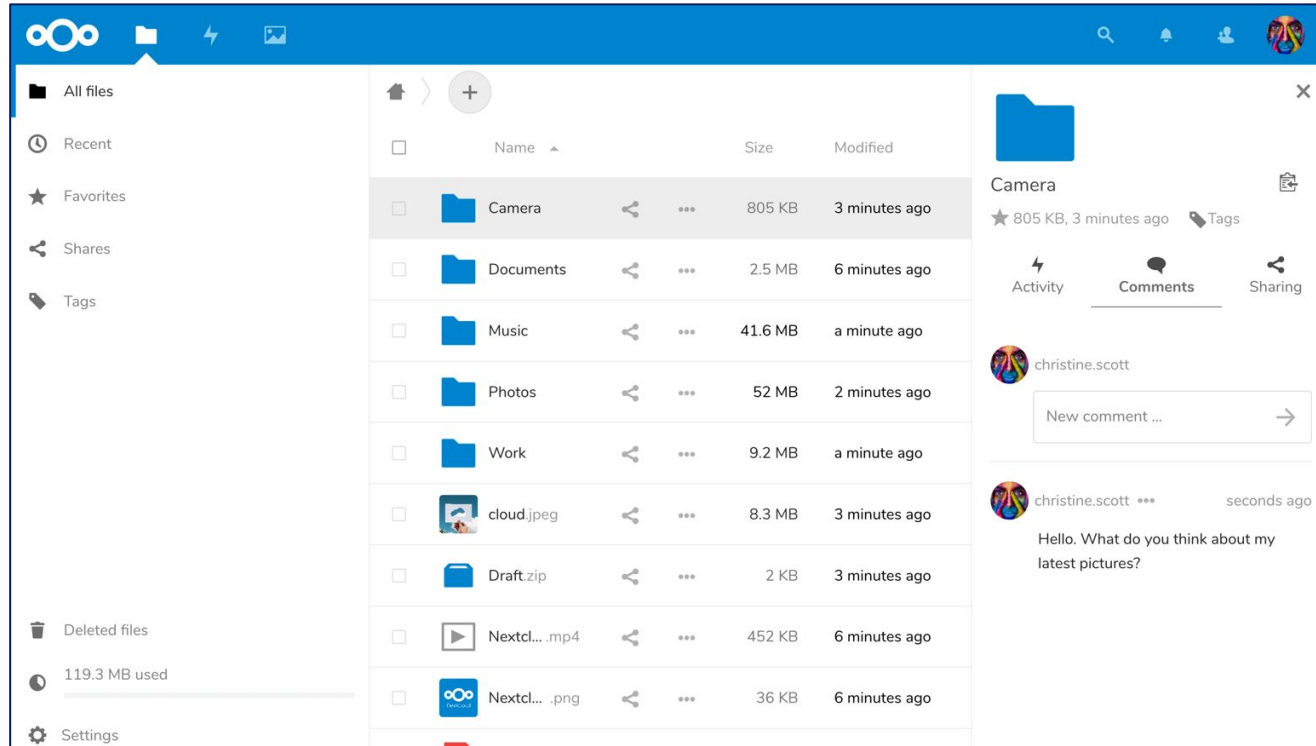
The screenshot shows the Confluence interface for a workspace named 'coronastatus'. The left sidebar contains navigation options: Overview (selected), Space Settings, and SPACE SHORTCUTS (Meeting notes). Below these are PAGES, including 'Meeting notes' and two links: 'Corona status website...' and 'References for survey...'. The main content area is titled 'Meeting notes' and includes a 'Create meeting note' button. Below this is a section 'Incomplete tasks from meetings' with a table of tasks. The table has columns for Description, Due date, Assignee, and Task appears. The tasks listed are: 'Proposal writing (due tomorrow)' (due 2020-03-25), 'Swedish translation, @Malin Klang is looking into this.' (due 2020-03-25, with a GitHub link), and 'Investigate survey questions and suitability of the website and dashboard - @Irene Stevens' (due 2020-03-25, assigned to Irene Stevens). Below the tasks is a section 'All meeting notes' with a table of meeting notes. The table has columns for Title, Creator, and Modified. The notes listed are: '2020-03-26 Meeting notes' (created by Phil Ewels, modified 2020-03-27) and '2020-03-25 Meeting notes' (created by Phil Ewels, modified 2020-03-25).

Description	Due date	Assignee	Task appears
<input type="checkbox"/> Proposal writing (due tomorrow)	2020-03-25		Meeting notes
<input type="checkbox"/> Swedish translation, @Malin Klang is looking into this. https://github.com/SciLifeLab/coronastatus/issues/3	2020-03-25		Meeting notes
<input type="checkbox"/> Investigate survey questions and suitability of the website and dashboard - @Irene Stevens	2020-03-25	Irene Stevens	Meeting notes

Title	Creator	Modified
2020-03-26 Meeting notes	Phil Ewels	2020-03-27
2020-03-25 Meeting notes	Phil Ewels	2020-03-25

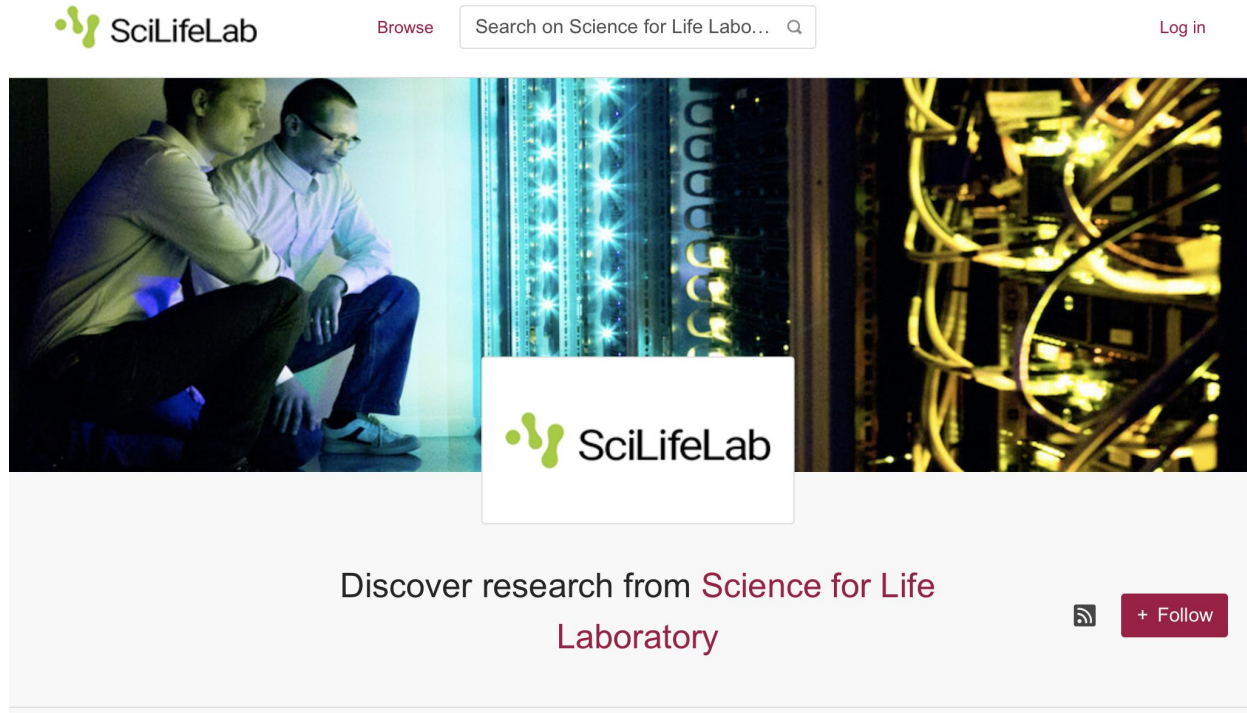
Confluence collaborative workspaces – share and edit documents, plan projects, manage tasks, etc.
<https://scilifelab.atlassian.net> / datacentre@scilifelab.se

Nextcloud – dropbox-style file sharing



Nextcloud – Dropbox style file sharing for larger data
<https://nextcloud.dckube.scilifelab.se> / datacentre@scilifelab.se

Figshare – general data publishing *demo to come*



Data publishing, DOI minting, meta data annotation, etc. <https://scilifelab.figshare.com>
Submission workflow: <https://www.scilifelab.se/community-pages/systems-data/repository>

User accounts



Getting an account

- Many SciLifeLab affiliated researchers or facility staff already have accounts. To check if you have an account, go to <https://scilifelab.figshare.com/account/forgotpassword>, type your email address and click submit. If you receive an email you already have an account and need to set up a password before you can log in. You can set up a password by following the instructions in the received email. ***Accounts are created for all SciLifeLab Fellows!***
- If you don't have an account, email datacentre@scilifelab.se and an account will be created for you.

How to login

- To login, go to the SciLifeLab Data repository and click on the red Log in button in the top right side of the screen. Enter your user name and password.

Upload data



There are several ways to upload your data:

- Through My data where you can drag and drop files of up to 5GB (default limit).
- Using the desktop uploader or the Figshare API, especially when working with large file/s – most web browsers can cope with about 5GB single file uploads.

Metadata

When uploading an item you will be asked to fill in a metadata form. The metadata form consists of both mandatory and non-mandatory fields.

- Mandatory metadata: title, author, subject category, item type, keywords, license, description and contact email.
- Additional metadata: funding, resource, references, publisher and access request email.

Upload data conditionally



There are a number of reasons why you may want to conditionally upload your files:

- There may be ethically or commercially sensitive data
 - You may not have the permission to make the data available
 - Your data may be stored somewhere else, but you want to link to it to get a DOI
 - You want to reserve a DOI for your data which you can include in your publication
-
- Embargo
 - Linked file
 - Meta data record
 - Reserve DOI
 - Sharing private data

Publish, edit and delete items



- Before publishing, consider attaching both a **readme.txt** file and a **manifest** file, containing a checksum, to the item. The readme.txt file should preferably contain the same metadata as are stated in the metadata form.
- When ticking **Publish**, you'll be prompted to check the **license** you've assigned and the terms of use. Your item will then be sent for review by SciLifeLab's institutional reviewers

Edit items

- You can go back and edit items after you've made them publicly available. Some changes may trigger a new version.

Delete items

- Click on the Delete item button in the metadata form to delete a private item.
- Published items are considered to be published permanently and can only be deleted in special cases.

Good behaviour for life science data



- Share your research data – preferably by submitting data to international repositories
- Annotate your data well
- Human data is often sensitive personal data and require stronger security and more careful handling
- Back up your data
- Keep track of data analysis and data management with the same detail as you would keep track of lab protocols in lab notebooks (and use an electronic notebook system!)
- Work reproducibly: <https://nbis-reproducible-research.readthedocs.io/>
- Set up data management plans
- Keep track of which versions of data and software that you have used to generate results that you publish
- Data sharing => more citations!

Publication database - *demo to come*



The screenshot shows the SciLifeLab Publications website. At the top, there is a navigation bar with links for Home, Publications, Facilities, About, and a Login button. The main header features the SciLifeLab logo and the title "SciLifeLab Publications". Below this, a paragraph states: "The publications in this database are the result of research assisted by the facilities of SciLifeLab. Publications due to researchers and fellows of SciLifeLab are available in a separate database at <https://publications-affiliated.scilifelab.se/>."

On the left side, there is a sidebar with a "Publications" section containing a list of years and the number of publications for each: 2019 (38), 2018 (541), 2017 (601), 2016 (632), 2015 (423), 2014 (374), 2013 (318), 2012 (243), 2011 (164), 2010 (139), and All (3473). Below this is a "Facilities" section and a "Journals" section.

The main content area is titled "The 10 most recent publications". It displays three publication entries, each with a title, authors, journal information, and a list of associated facilities. The first entry is "Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard" by Andrade P, Pinho C, Pérez i de Lanuza G, ..., Andersson L, Carneiro M, published in Proc Natl Acad Sci USA, 2018, 115(20):12032-12037. The second entry is "New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries" by Shrine N, Guyatt AL, Erzurumluoglu AM, ..., Tobin MD, Wain LV, published in Nat. Genet. 2019, 51(3):481-493. The third entry is "Rostania revised: testing generic delimitations in Collemataceae (Peltigerales, Lecanoromycetes)" by Košuthová A, Westberg M, Otálora MAG, Wedin M, published in MycoKeys 2019, 47(-):17-33.

Facilities:

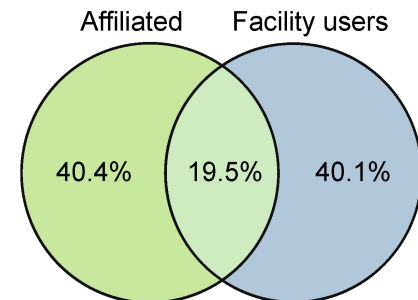
<https://publications.scilifelab.se>

Get curator accounts from Data Centre to report publications directly into the database:

datacentre@scilifelab.se

Affiliated researchers:

<https://publications-affiliated.scilifelab.se>



Overlap between facility and researcher publications



- Publish relational database on the web.
- Based on Sqlite.
- Allows a user to query (SQL) the contents.
- Allows download of data.
- The db can be created via the web interface.
- <https://dbshare.scilifelab.se/>
- Ask datacentre@scilifelab.se for more info, or an account to try it out.

DataGraphics



- Display plots (graphics) based on data in CSV or JSON format.
- Uses the Vega-Lite system for specifying the graphic.
- The data is semi-static; updateable as a whole.
- A plot always uses the current data; no need to update it explicitly.
- A graphic can be embedded in your own HTML page.
- <https://datagraphics.dckube.scilifelab.se/>
- Ask datacentre@scilifelab.se for more info, or an account to try it out.

datacentre@scilifelab.se



www.scilifelab.se/data



scilifelab.slack.com #datacentre



support@nbis.se