NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

The National Bioinformatics Platform Sweden

Sebastian Di Lorenzo
Sebastian.DiLorenzo@NBIS.se

https://github.com/Sebastian-D/Presentations

SciLifeLab Vision

"To be an internationally leading center that develops, uses and provides access to advanced technologies for molecular biosciences with focus on health and environment."

Solna

Uppsala

Facilities

Short-term support and infrastructure

Wide competence in bioinformatics, Assembly/Annotation, SysDev

Long-term support

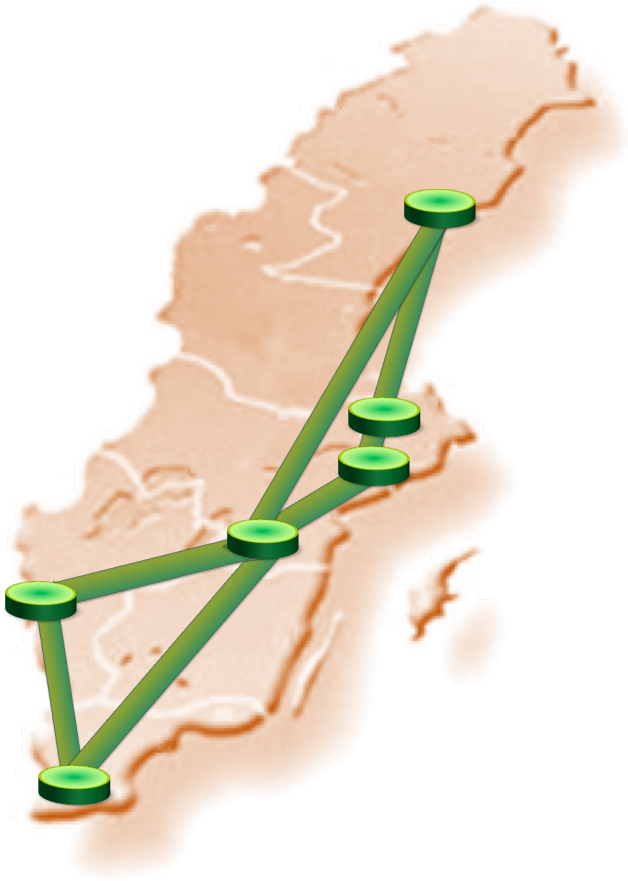
Large collaborative projects selected by scientific ranking

Systems biology

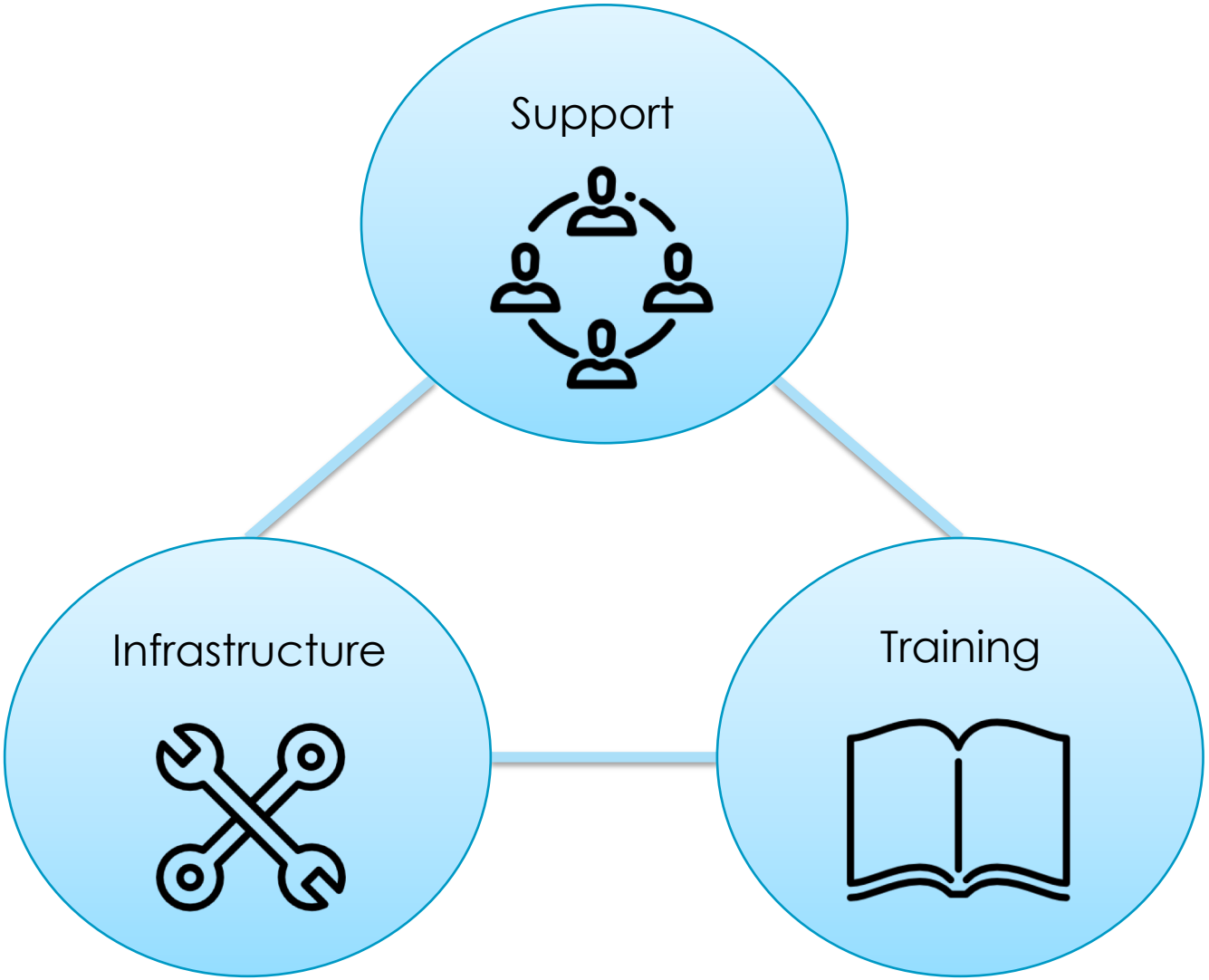
Network analyses and Integrative bioinformatics

Compute and storage

Computational and storage resources for bioinformatics, especially next-generation sequencing



Focus areas



Grant season

- **Free consultations**
  - Project planning
  - Grant writing
    - Bioinformatics tools
    - Reproducability
    - Data management
    - Storage
  - Data Management Plans (DMPs)

Data management

Will become a standard part of the research funding process

nature
International journal of science

EDITORIAL · 13 MARCH 2018

Everyone needs a data-management plan

They sound dull, but data-management plans are essential, and funders must explain why.

By 2019, all who receive grants from us must have a data management plan

As from spring 2019, if you are awarded a grant from the Swedish Research Council you must have a plan for how the research data generated within your project shall be managed.

You must not send in your data management plan to us when you apply for a grant, but your administrating organisation will be responsible for ensuring that a data management plan is in place when you start your project or corresponding, and that the plan is maintained.

DMPs - National guidelines

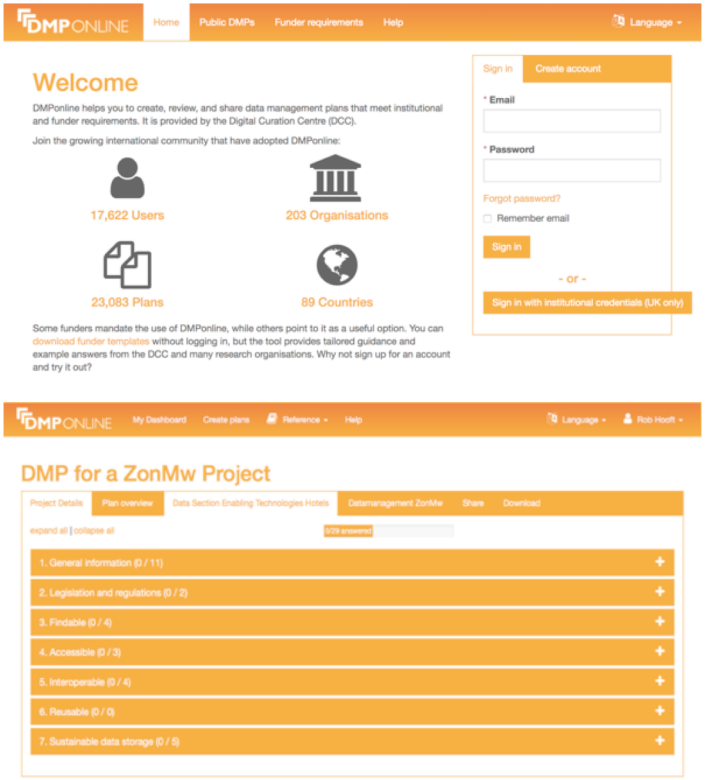
VR & SUHF (Association of Swedish Higher Education Institutions)

- **Central parts of a data management plan***
  1. Description of data – reuse of existing data and/or production of new data
  2. Documentation and data quality
  3. Storage and backup
  4. Legal and ethical aspects
  5. Accessibility and long-term storage
  6. Responsibility and resources

*Based on Science Europe's "Core Requirements for Data Management Plans"

DMP tools

DMPonline



https://dmponline.dcc.ac.uk/

ELIXIR Data Stewardship Wizard

https://dsw.scilifelab.se
https://ds-wizard.org/

Thanks for listening!

https://github.com/Sebastian-D/Presentations

VRs recommendations

1. Description of data – reuse of existing data and/or production of new data
   - How will data be collected, created or reused?
   - What types of data will be created and/or collected, in terms of data format and amount/volume of data?
2. Documentation and data quality
   - How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?
   - How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?
3. Storage and backup
   - How is storage and backup of data and metadata safeguarded during the research process?
   - How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?
4. Legal and ethical aspects
   - How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?
   - How is correct data handling according to ethical aspects safeguarded?
5. Accessibility and long-term storage
   - How, when and where will research data or information about data (metadata) be made accessible? Are there any conditions, embargoes and limitations on the access to and reuse of data to be considered?
   - In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?
   - Will specific systems, software, source code or other types of services be necessary in order to understand, partake of or use/analyse data in the long term?
   - How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded?
6. Responsibility and resources
   - Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?
   - What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data fulfil the FAIR principles?

Cheat sheet, don't show audience!

Short Term Support – an hourly fee

- Always open for applications
- Quick turn-over
- Work usually starts within 2 weeks from signed agreement
- Most projects 30-60 hours, stretching over 1-3 months
- 800 kr/h, time logged per hour used

Long Term Support – free

- Application rounds, 3 times/year
- Scientific committee
- Projects chosen for support based on scientific excellence
- 500 h of support, usually stretching over 1-2 years
- Hands-on involvement from research group mandatory

Partner project – 50% of an expert's salary

- For longer projects with a large bioinformatics component
- Always open for applications
- At least 2 years, at least 50%
- Hands-on involvement from research group recommended
- Decisions on case-by-case basis
- Expert might work at another site, will in general retain workplace located at NBIS hubs (regular visits with group can of course be arranged)

Training

- **Introduction to omics data analysis**
- 4 x Introduction to Bioinformatics using NGS data, 5 days (Manfred Grabherr)
- 1 x Omics data analysis, 10 days (Erik Fredlund, Lukas Orre)

- **Advanced topical courses**
- 2 x RNAseq, 3 days (Johan Reimegård)
- 1 scRNA, 2 days (Åsa Björklund)
- 1 x Genome assembly, 3 days (Henrik Lantz)
- 1 x Genome annotation, 3 days (Henrik Lantz)
- 1 x ChIP-seq 3 days (Agata Smialowska, Olga Dethlefsen)
- 1 x Proteomics, 4 days (Fredrik Levander)

- **Programming and reproducible research**
- 1 x Reproducible research, 2 days (Rasmus Ågren, Leif Wigge)
- 2 x Python, 2 days (Nina Norgren, Sergiu Netotea)
- 1 x R Summer School, 10 days: RaukR (Marcin Kierczak,Sebastian DiLorenzo)
  1 x R programming Foundations for Life Scientists, 5 days (Marcin Kierczak)
- 1 x Unix/Linux Tutorial for beginners, 3 days (Mihaela Martis)