

Open issues:

\* Do we have effective and useful national data management policies? Is there a case for collaboratively drafting a proposal to influence instating effective and useful national data management policies? Directed to ministries? Funding councils? (Joel)

\* Is there a need for (internationally?) consistent metadata for research data? Do we have any mechanisms for instating this? (Joel)

\* I don't know the difference between EUDAT B2SHARE, Zenodo, OwnCloud... Am I the only one? (Joel)

FYI:

\* BBMRI FAQ on new EU data protection law: <http://bbmri.de/documents/10184/13584/ef3a6ef1-d6f9-4245-b9d6-fa6d3eb72338.pdf>

Nonsense:

\* Useful cartoon illustrating the problems of open research data:

\* Data Sharing and Management Snafu in 3 Short Acts: [https://www.youtube.com/watch?v=66oNv\\_DJuPc](https://www.youtube.com/watch?v=66oNv_DJuPc)

= Agenda =

== Day 1 (18.05.2016) ==

""10:30 - 11:00 Welcome""

10:30 Welcome (Michaela Barth, NeIC) [[:Media:160518\_2ndNordic\_data\_services\_workshop\_caella.pptx|slides]]

10:40 Pooling Competences and resources (Rob Pennington, NeIC) [[:Media:DM\_workshop\_Pennington.pptx|slides]]

""11:00 - 12:30 Session 1""

11:00 CSC Storage Service Overview (Anssi Kainulainen, CSC)

11:30 SNIC Storage Service Overview (Dejan Vitlacil, SNIC)

12:00 KTH Records Management Office (KTH IT Archivist) (tabled)

12:10 Research Data Management in NBIS (Niclas Jareborg, NBIS) [[:Media:2016-05-18\_NBIS\_Research\_Data\_Management.pdf|slides]]

""12:30 - 13:30 Lunch""

""13:30 - 15:00 Session 2""

13:30 Welcome (Gudmund Høst, NeIC)

13:40 UNINETT Sigma2 Storage Service Overview (Andreas O Jaunsen, UNINETT Sigma2)

14:10 DeIC Storage Service Overview (Henrik Pedersen, DeIC)

14:40 Embedded librarians - Working in a research infrastructure (Monica Lassi, Lund University Library)

""15:00 - 15:30 Coffee break""

""16:00 - 17:00 Session 3""

16:00 NSD and Research Data (Atle Alvheim, NSD)

16:30 DDA and Research Data (Anne Sofie Fink Kjeldgaard, DDA)

""17:00 - 18:30 Social activity""

""19:00 - Dinner""

== Day 2 (19.05.2016) ==

""09:00 - 10:30 Session 4""

9:00 SND and Research Data [Data Pilot Activity] (Johan Fihn, SND)

9:45 FSD and Research Data (Helena Laaksonen, FSD)

""10:30 - 11:00 Coffee break""

""11:00 - 12:30 Session 5""

11:00 Ransomware and other threats to data storage and data services (Leif Nixon, Security Officer)

""12:30 - 13:30 Lunch""

""13:30 - 15:00 Session 6""

13:30 NDS and Research Data (Christine Kirkpatrick, National Data Service, USA)

14:00 National e-Infrastructure providers and National Data Service providers

... Do we provide complementary or competing services?

... Is there a room for collaboration?

... Building and designing services together?

15:00 Closing

Minutes:

== Welcome ==

(Michaela Barth, NeIC)

[[Media:160518\_2ndNordic\_data\_services\_workshop\_caela.pptx|slides]]

Round of presentations (clockwise around table):

\* Dejan Vitlacil, SNIC storage lead

\* Jens Larsson, NSC, Swestore coordinator

\* Mikael Borg, NBIS technical coordinator, providing bio data services for seden, part of ELIXIR

\* Niclas Jareborg, NBIS data manager

\* Michaela Barth, NeIC XO, Pool competencies focus area lead

\* Monica Lassi, Lund University library

\* Rob Pennington, NSCA, NeIC special advisor on data and HPC, Share resources focus area lead

\* Atle Alveim, Norwegian center for research data (formerly Social Science data services)

\* Johan Fihn, Swedish data services (SND)

\* Andreas Jaunssen, Sigma2 data services lead.

\* Anne Sofie Fink Kjeldgaard, National archive of Denmark (DDA)

\* Chris Ariyo, CSC, EUDAT service manager FI

\* Anssi Kainulainen, CSC, application architect at information infrastructure services -group

\* Christine Kirkpatrick, NDS, US

\* Joel Hedlund, NeIC XO, Tryggve project owner, Strengthen stakeholder dialogue focus area lead

\* Jessica Persson, KTH data archivist.

\* Henrik Peterson, chair forum for data management in Denmark

\* Gudmund Høst, NeIC dir

\* Jacko Koster, SNIC coordinator for international activities (prev dir)

\* Helena Laksonen, Acting director Social sciences data archives, information officer.

== Pooling Competences and resources ==

(Rob Pennington, NeIC)

[[Media:DM\_workshop\_Pennington.pptx|slides]]

NeIC does not operate any facilities

26 Mio people in Nordic region

National budgets are not getting better

Training Calendar: <https://neic.nordforsk.org/training/> upcoming trainings: [https://wiki.neic.no/wiki/Upcoming\\_trainings](https://wiki.neic.no/wiki/Upcoming_trainings)

NeIC training policy draft is accessible internally until approved, but can be made available on request to:

\* Joel [mailto:joel@nsc.liu.se]

\* Michaela [mailto:caela@kth.se]

\* Rob [mailto:rlpennington1@gmail.com]

Sharing of application experts expertise for advanced user support: People are not scalable!

== CSC Storage Service Overview ==

(Anssi Kainulainen, CSC)

Presentation slides: <https://kannu.csc.fi/index.php/s/3Blq80ovkYfrqgQ>

Main goal of IDA to keep data secure, focus not so much on sharing data

IDA also used for active data?

Metadata profile needs to declare if open for others, everyday storage and sharing in future plans for new IDA

Who is the data owner? Researchers or universities?

The project is the data owner, actual people can change. Policies for what happens when the projects ends

Effort going on to match metadata between Etsin and EUDAT B2Find to make data migration easier

Funding?

Contract with ministry: Paying customer is the ministry, agreements with the users

Services ordered by ministry, services defined in working groups

Open Science and research initiative has forum discussions.

Total budget of whole life cycle of data services?

-

% of users using IDA?

Compared to Taito with 1600 users yearly, 500 users is about 30%.

== SNIC Storage Service Overview ==

(Dejan Vitlacil, SNIC)

[[Please link any presentation you might have here]]

Major challenge: better communication with universities.

== KTH Records Management Office ==

(KTH IT Archivist)

[[Please link any presentation you might have here]]

(This item was tabled)

== Research Data Management in NBIS ==

(Niclas Jareborg, NBIS)

[:Media:2016-05-18\_NBIS\_Research\_Data\_Management.pdf|slides]]

\* Gothenburg, Lund and Umeå universities becoming part of SciLifeLab during 2016.

\* NBIS invites for drop in visits every week, check calendar at <http://nbis.se>.

\* SNIC Sens (bianca) first users in August 2016.

== Welcome (again) ==

(Gudmund Høst, NeIC)

(no presentation)

== UNINETT Sigma2 Storage Service Overview ==

(Andreas O Jaunsen, UNINETT Sigma2)

[[Please link any presentation you might have here]]

Data should not be duplicated unnecessarily requirement reduces the number of potential vendors for ongoing procurement

Final solution will be operated by SIGMA2 themselves: technical operation and advanced user support, support service in container with HPC services in metacenter containing altogether 40 people

Researchers who are not at NO institutions but who are collaborating with NO researchers can get access to the data infrastructure. It has not been structured to be an international data sharing environment.

== DeIC Storage Service Overview ==

(Henrik Pedersen, DeIC)

[[Please link any presentation you might have here]]

Apps store for data apps that are to be usable for researchers. Four are currently under consideration.

== Embedded librarians - Working in a research infrastructure ==

(Monica Lassi, Lund University Library)

[[Please link any presentation you might have here]]

== NSD and Research Data ==

(Atle Alvheim, NSD)

[[Please link any presentation you might have here]]

The project's home institutions (not the project) own the data and are legally responsible.

According to the law projects are not allowed to keep the data they have worked on for more than 3 resp. 5 years. But the data needs to be kept 10 years: deleted or otherwise anonymized at the researcher.

NSD runs a data privacy ombudsman service, a service offered to their “member” institutions.

== DDA and Research Data ==

(Anne Sofie Fink Kjeldgaard, DDA)

[[Please link any presentation you might have here]]

Researchers and students need to register first with a signature on a paper to be able to use the service.

Keywords provided by archive, classification based on a social science oriented thesaurus.

European languages thesaurus based in London enables to search across several languages. Output is an xml file.

Secure data distribution system and environment needed in the future.

== Day 2 (19.05.2016) ==

== SND and Research Data [Data Pilot Activity] ==

(Johan Fihn, SND)

[[Please link any presentation you might have here]]

Atle Alvheim on our fatalistic view:: Not building a place to die for the data, it's all about the services, should focus on what function we want the data to fulfil

== FSD and Research Data ==

(Helena Laaksonen, FSD)

[[Please link any presentation you might have here]]

Aila usage from universities outside Finland a little bit below 10%

== Ransomware and other threats to data storage and data services ==

(Leif Nixon, Security Officer)

[[Please link any presentation you might have here]]

Phalanx 2.0 still not public available, (Phalanx 0.6 has leaked) It is not illegal to write malware.

Ebury has not leaked yet either: 2-factory identification could have made Ebury attacks impossible

Cdorked.A + Boaxxe + Glupteba + Ebury -> all about Porn site ads

Money flow investigation: worth a Lamborghini in Russia

FBI needs venue to get involved: crime has to happen in the geographic area.

Criminal deported to US waiting for sentence.

Don't run java applets in your browser!

What can you do: Defense still stays the same:

- \*Keep backups: make sure backups are protected: Don't have your backup mounted all the time (e.g write once, introduce delays in replication and have snapshots if possible)
- \*Control authority: Nobody should be able to delete all your data at once
- \*Be up to date: They can't exploit a patched vulnerability

Ransomware on a colleague's computer can encrypt common shared documents

Threats on different OSes typically proportional to marketshare.

Billion \$ question: kernel on all Linux compromised?

Some browser exploits can grab passwords out of eg WinSCP profile storage. Don't use it. <- seriously. it saves passwords unencrypted in the profile.

\* (no not browser exploits, the whole computer was corrupted)

\*\* You sure? That's a win machine.

\*You can also log in remotely to win machines and have a password...

\*\* Actually, yes, the win machine was rooted, but a major contributing factor in this case is that WinSCP saves passwords unencrypted. Don't use it.

Q (joel):

\* Compiler corruption to trojan kernel builds?

\*\* Not likely.

\* Isn't DNS queries to random servers unusual?

\* Amateurs using sophisticated tools made by someone else? Exfil servers have more than one consumer?

\* How do you replace an exfil host?

\* If ISP and LE are in, can't you just shut down the compromised hosts?

\* What was the browser exploit?

\* 2f auth enough to dodge Ebury etc? Future?

\*\* Yes.

\* Do you need to \*buy\* an exploit kit? Can't you just browse ads until you get one?

\* How to protect backups?

\*\* Enterprise backup systems with limited access.

\*\* Don't keep backup mounted all the time.

\*\* Write-once solutions, to random filenames.

\*\* Delay in replication.

\*\* Snapshots.

\* <http://www.welivesecurity.com/2014/03/18/operation-windigo-the-vivisection-of-a-large-linux-server-side-credential-stealing-malware-campaign/>

== NDS and Research Data ==

(Christine Kirkpatrick, National Data Service, USA)

[[Please link any presentation you might have here]]

\* NDS Workbench Demo: <https://www.youtube.com/watch?v=Wz-45JGk4tg>

\*

== National e-Infrastructure providers and National Data Service providers ==

Possible discussion topics:

\* Do we provide complementary or competing services?

\* Is there a room for collaboration?

\* Building and designing services together?

Are there people who should be invited to this type of meeting?

Should security aspects something that should be worked on together or talked about more?

Is there an opportunity for another information exchange at NeIC 2017?

\* Very useful to attend the forum. It might be good to bring in university libraries. Is there someone from a university who can describe what they need? Someone from a high level who can describe what they need?

\* Mention of describing or defining the roles with university libraries or users. How do they handle the data?

\* Follow up on the lack of formulated policies that specify that this is a valuable activity in a heterogeneous environment. It should include the academic side and the technology side. There is enormous diversification in the field and there isn't a single institution managing the services. The major problem for users is finding out what the services are they need to place the data or find the data to be used. Organizing the services to manage the data services should be decentralized. The current technologies in use is a tiny fraction of what is available.

\* Possibly have representatives from funding agencies at the next meeting so that they can hear about the problems.

\* Data management should be advocated as a policy because it is a profitable thing that can be done and there is value in the data that should be understood.

\* How do you engage the range of stakeholders in all of the countries? Each country is different - DK has RDM Forum. There was a recent workshop in NO that brought a number of stakeholders together.

\* Should we meet again in this environment and make it slightly larger?

\* Perhaps this type of meeting that could come into the NeIC conference as a co-located activity to expose it to the wider community.

\* This might be challenging with the funding agencies to have the people who will be able to speak authoritatively attend.

\*\* They should hear what is being done and the challenges in presenting these to the funding agencies.

\*\* The funders want to hear from the researchers who will benefit not necessarily from the e-infrastructure providers.

\*\* US Big Data hubs - four regions and each region should focus on specific topics. The funding is for bringing people together. Can a group be gotten together in the Nordic region that includes the domain users?

\* The e-science action plan 2.0 has five relevant tasks - create a forum within NeIC that includes national data services, e-infrastructures and produce a document to respond to this as a position paper..

\*\* Engage with the university CIOs as well. How many universities in the Nordics that might be engaged?

\*\* Could representatives from this group take an initiative and have an impact?

\*\* Do some background work and then plan to have a workshop in June 2017 to produce the position paper?

\*\* Example - do a survey on the view of the RC of DM or check the web sites for each country? Draw an outline of what is available or relevant in each country?

\*\* There is nothing in common in the Nordics on the topic of data. RIs might be quantified in terms of productivity.

\*\* NeIC would be happy to facilitate a new activity.

\*\* Compile a list of capabilities and talents in each of the Nordic countries - benchmarking effort? Can NeIC work with the group on this?

\* Include the CIOs may be important if they have to provide research computing support.

\* Try to organize another event in 12 months.

\* Move forward with the idea of a Nordic Forum for Research Data Management.

\* Training - is there something that might be done with the different countries? Information exchange on different training resources, such as youtube videos?

\*\* Finnish training material is generally in Finnish.

\*\* Train the trainer idea - can this be described better or more extensively?

\* Any topics that weren't covered here that could be included? Two data initiatives from each country at the meeting.

\*\* Publishers have a need to attach data to the publications and help on reproducibility.

\* Topics for the next meeting might be more focused on specific issues after this meeting which was general.

\*\* For NeIC 2017 - select topics and open for submissions by participants? How long should this be?

\* There are several initiatives that may be developing silos in the different countries due to institutions needing to take responsibility? Is building ERICs building silos?

\*\* Could have an open brainstorming meeting in a month or two to decide on what to do next?

\*\* How to give the users a good offer for solving their problems?

\* National research councils are usually financing single infrastructures that may be topic or domain specific or nationally focused and these aren't being shared.

- \*\* Should be collaborating more on sharing the services or technologies.
- \* EUDAT has a good concept and the services are still in development. Strength should be that they are part of one infrastructure that flows together but the links are still to be implemented, which will be interesting.
- \*\* A generic API that can use any of the necessary services - API is still being worked on with EUDAT.
- \* Lund has people who are the bridging element between data librarians and researchers to make it more discoverable/available.
- \* Common topic this year was Data Management Plan, so the group is moving along similar pathways.
  
- \* Conference in Bergen - iASSIST2016 end of May/first days of June.
- \*\* Perhaps this type of activities should be considered for dissemination like training is disseminated.
- \*\* Data librarians are in attendance at this conference.

-----