

# **NeIC 2019 - Nordic Infrastructure for Open Science**

Tuesday, 14 May 2019 - Thursday, 16 May 2019

Tivoli Hotel & Congress Center

## **Book of Abstracts**



# Contents

Welcome by NeIC chairman . . . . .	1
The Danish ambitions towards future Research Infrastructure Collaboration . . . . .	1
A regional approach towards Open Science . . . . .	1
Nordic Collaboration for Social Impact . . . . .	1
Bringing Open Science in the US into Practice . . . . .	1
Open Science: is the Research Data Alliance a help or a hindrance? . . . . .	2
Improved Observation Usage in Numerical Weather Prediction (iOBS) . . . . .	2
Conference dinner . . . . .	3
Shaping up the Nordics for EOSC . . . . .	4
Lifeportal . . . . .	4
Repurposing Climate Data . . . . .	5
Integrated Nordic-Baltic Genebank Information Management System . . . . .	5
Glenna . . . . .	5
EOSC-Nordic Technical session . . . . .	6
Introduction to the Rahti container cloud . . . . .	7
JupyterHub for research facilities . . . . .	7
An Introduction to Deep Learning for Natural Language Processing . . . . .	8
Reimagining research computing . . . . .	9
Open Science with Sensitive Data . . . . .	10
An Introduction to Deep Learning for Natural Language Processing . . . . .	11
Reimagining research computing . . . . .	11
Open Science with Sensitive Data . . . . .	11
Transportation social activities . . . . .	12

Social activities . . . . .	12
Security in the Nordics . . . . .	12
Building and Managing Linux Containers for Centralized and Distributed Systems . . . . .	13
FAIR and GO FAIR . . . . .	13
FAIRness in practise: How to achieve it? . . . . .	14
National coordination of open access to research data in Sweden . . . . .	14
FAIR Across –Implementation of FAIR into research practice . . . . .	14
Open discussion on turning FAIR into practice . . . . .	15
Research perspectives on AI, transparency, privacy, law. . . . .	15
Closing remarks . . . . .	15
NeIC presents Nordic project highlights . . . . .	15
FAIR . . . . .	16
Large scale Nordic Language Processing Laboratory . . . . .	16
ELIXIR Cloud & AAI . . . . .	16
NeIC Tryggve - cross-border services for human data . . . . .	17
Bridging RDA and Danish Research Communities . . . . .	18

**Plenary / 12**

## **Welcome by NeIC chairman**

**Corresponding Author:** sp@adm.dtu.dk

**Plenary / 43**

## **The Danish ambitions towards future Research Infrastructure Collaboration**

**Author:** John Renner Hansen<sup>1</sup>

**Co-author:** Gitte Kudsk<sup>2</sup>

<sup>1</sup> *Chairman DeiC board*

<sup>2</sup> *DeiC*

**Corresponding Authors:** dekan@science.ku.dk, gitte.kudsk@deic.dk

**Size of poster:**

**Plenary / 13**

## **A regional approach towards Open Science**

**Corresponding Author:** arne.flaoyen@nordforsk.org

**Plenary / 14**

## **Nordic Collaboration for Social Impact**

The Nordic e-Infrastructure Collaboration (NeIC) offers a framework where national actors can join forces to co-create new services, define common operational frameworks and pool the e-infrastructure competencies. NeIC supports excellent research through partnerships of national actors. However, the impacts of our collaboration reach much farther. My presentation is biased towards social impact, but if you follow closely you may find hints at future directions that we could take together.

**Size of poster:**

**Bringing Open Science in the US into Practice / 31**

## **Bringing Open Science in the US into Practice**

**Author:** Christine Kirkpatrick<sup>1</sup>

<sup>1</sup> *San Diego Supercomputer Center, Univ of California San Diego*

**Corresponding Author:** christine@sdsc.edu

With dozens of government agencies and foundations funding research at over 200 universities and hundreds more institutes and businesses, the United States comprises a challenge to comprehensive open science offerings. Approaches by various government agencies to require and incentivize open access to data will be mentioned, as well as platforms and services that enable data sharing and discovery. Models working in the EU, are being replicated in the US and providing a basis for increased awareness and value in open data. The role and challenges of including industry or private research will also be discussed.

**Open Science: Is the Research Data Alliance a help or a hindrance? / 24**

## **Open Science: is the Research Data Alliance a help or a hindrance?**

**Author:** Hilary Hanahoe<sup>1</sup>

<sup>1</sup> *Research Data Alliance*

**Corresponding Author:** hilary.hanahoe@rda-foundation.org

(Title given by organisers: Insights on Open Science & EOSC from an RDA perspective)

Within the complex, international landscape of open science and open data, the research data landscape is highly fragmented, by disciplines or by domains. When it comes to cross-disciplinary activities, the notions of “building blocks” of common data infrastructures and building specific “data bridges” are accepted metaphors to approach data complexity and enable data sharing.

The Research Data Alliance (RDA) develops solutions, specifications and best practices enabling data to be shared across barriers through focused Working Groups and Interest Groups, formed of data professionals from all around the world.

This presentation will address where the RDA community stands within the realm of open science and, specifically, the European Open Science Cloud. Does it support or just add another element of complexity?

**Improved Observation Usage in Numerical Weather Prediction (iOBS) / 21**

## **Improved Observation Usage in Numerical Weather Prediction (iOBS)**

**Author:** Jørn Kristiansen<sup>1</sup>

<sup>1</sup> *MET Norway*

**Corresponding Author:** jornk@met.no

Observations from the “Internet of things” (IoT), such as intelligent cars, phones, buildings and personal weather stations (PWS), including commodity weather sensors, provide detailed information on local to hyper-local meteorological phenomena. This NordForsk infrastructure project (iOBS) will accommodate an increasing amount and diversity of observation data, and provide a system of harmonised data pooling and merging. The targeted breakthrough and measurable benefit of iOBS is the effective assimilation of diverse observations in regional high-resolution NWP models for the delivery of reliable and accurate weather forecasts and warnings for the benefit of operations, business and society. The basis will be the current operational NWP model, AROME-MetCoOp and/or the very recent addition of a nowcasting suite. At the same time, there is currently a significant and unnecessary diversity at the different National Meteorological Institutes in formats, file structures and

(local) software used for observation handling and pre-processing. This fragmented data handling introduces redundancies, errors and missing observations, and the consequence is that valuable information is lost. iOBS will therefore introduce the Scalable Acquisition and Pre-Processing system (SAPP) for a joint observations handling.

The project will enable use of high-resolution and high-frequency observations. This requires to improve, develop and implement timely quality control (QC) algorithms for a massive amount of private observations of surface pressure. To our knowledge, if successful this will be the first time private pressure observations are assimilated in an operational NWP system.

The observation data flow will be built in parallel on two future generation e-infrastructures: MET Norway's PPI and Glenna-2. PPI provides flexibility, scalability in computing data storage capacity and full end-to-end data integrity to meet modern requirements on data consistency. PPI offers the benefits of both building on existing operational solutions, run as an operational environment and act as a reference to the cloud service. Glenna-2 will make effective use of hybrid environments combining specialized HPC resources and for example container technology with the more flexible cloud delivery model. Having two e-infrastructures solutions offers redundancy and flexibility, addressing the needs and requirements of Nordic (and beyond) research and operations.

The benefits of this 2-year project include:

- Improved NWP forecast quality from increased number of observations used in data assimilation
- Improved QC algorithms for pre-processing private observations
- Reduced cost for software maintenance and development
- Improved conditions for Nordic research collaboration on both novel technologies and handling of different observation types
- Knowledge transfer across scientific disciplines and technological solutions
- Redundancy and flexibility by using both a cloud based research infrastructure (Glenna-2) and a proven operational infrastructure (PPI)
- Raise awareness of benefits of public-private partnerships, e.g. our QC will inform data manufacturers about their data quality

The project partners are CSC, FMI, MET Norway and SMHI.

**Gemyse Restaurant, inside Tivoli Gardens (please bring NeIC2019 confirmation email to enter) / 38**

## Conference dinner

The dinner takes place at Gemyse restaurant inside the famous and magical theme park, Tivoli Gardens. To join the dinner you must select attendance when registering to the conference. You can look forward to a delicious three-course gourmet dinner in beautiful surroundings. The required entrance ticket to Tivoli Gardens is then included.

Welcome!

Contact info:

Google maps link  
Bernstorffsgade 5  
1577 Copenhagen

+45 88 70 00 00  
gemyse@nimb.dk

`<table> <tr><td><a target="blank" href = "https : //indico.neic.no/event/18/images/24 - Hovedindgangen.jpg" >< imgsrc = "https : //indico.neic.no/event/18/images/30 -`

*Hovedindgangen.jpg*” alt = ”Tivoli” >< /a >< /td >< td >< atarget = ”\_blank” href = ”https : //indico.neic.no/event/18/images/25 – GemyseRestaurant.jpg” >< imgsrc = ”https : //indico.neic.no/event/18/images/31 – GemyseRestaurant.jpg” alt = ”Gemyse” >< /a >< /td >< /table >

### Size of poster:

### Plenary / 19

## Shaping up the Nordics for EOSC

**Author:** Lene Krøl Andersen<sup>1</sup>

<sup>1</sup> NeIC

**Corresponding Author:** lene.krol.andersen@deic.dk

*Shaping up the Nordics for EOSC* is based on the work behind the EOSC-Nordic project proposal coordinated by NeIC, which was submitted to the EC during autumn 2018. *Shaping up the Nordics for EOSC* aims to facilitate the coordination of EOSC relevant initiatives within the Nordic and Baltic countries and exploit synergies to achieve greater harmonisation at policy and service provisioning across these countries, in compliance with EOSC agreed standards and practices. The project brings together a strong consortium of 24 complementary partners including e-Infrastructure providers, research performing organisations and expert networks, with national mandates and experience with regards to the provision of research data services, and a unique capacity to realise the outcomes of the EOSC design as outlined by the EOSC Implementation Roadmap.

### Plenary / 20

## Lifeportal

**Authors:** Sabry Razick<sup>1</sup>; Nikolay Vazov<sup>1</sup>

<sup>1</sup> University of Oslo

**Corresponding Authors:** sabry.razick@usit.uio.no, n.a.vazov@usit.uio.no

Lifeportal(lifeportal.uio.no/) is a web-based interface developed for researchers who do not have advanced computer science expertise but need to perform resource-consuming computational analyses. Lifeportal promotes open science by enabling the users to share and reuse the results of these analyses, workflows and data among their collaborators or entire workgroups within one single platform.

Lifeportal is built on the galaxy platform (galaxyproject.org) and it is customized to fit the needs of the researchers and students. The unique features of the Lifeportal are :

- HPC backend
- Open ID Connect
- Project and user management module We will show perform an analysis involving different software using the Abel HPC cluster and display results using only a web-browser. Then we will show how these data and analysis pipeline could be shared freely with other Lifeportal users, making the science reproducible and methods reusable.



Plenary / 30

## Repurposing Climate Data

**Author:** Anne Fouilloux<sup>1</sup>

<sup>1</sup> *University of Oslo, Norway*

**Corresponding Author:** [annefou@geo.uio.no](mailto:annefou@geo.uio.no)

Advances in the development of climate models and associated data viewers and processing tools is achieving unprecedented maturity in the environmental scientific community. This was accompanied by the standardization of model output formats (conventions for Climate and Forecast metadata), the availability of open databases (i.e., the Earth System Grid Federation), and often of the climate model codes themselves.

Applying such **FAIR** principles virtually makes it possible to re-run climate model runs or undertake other experiments. On the one side such new opportunities should attract interest from other communities such as social and human sciences. On the other side, the climate models, viewers and processing tools are generally far too complex for non-specialists and computationally demanding thus hindering cross-disciplines transfer.

In this presentation we will show how climate models can be run out-of-the-box, without much effort, using an online web platform. We will also show how climate model outputs can be visualized or how deep-learning techniques can be applied using the same web portal.

Plenary / 25

## Integrated Nordic-Baltic Genebank Information Management System

**Author:** Jan Svensson<sup>1</sup>

**Co-authors:** Anna Palmé<sup>1</sup>; Karolina Aloisi<sup>1</sup>

<sup>1</sup> *Nordic Genetic Resource Centre*

**Corresponding Authors:** [anna.palme@nordgen.org](mailto:anna.palme@nordgen.org), [karolina.aloisi@nordgen.org](mailto:karolina.aloisi@nordgen.org), [jan.svensson@nordgen.org](mailto:jan.svensson@nordgen.org)

The Nordic and Baltic genebanks are responsible for conservation of plant genetic resources for food and agriculture. The e-infrastructure used by genebanks is termed Genebank Information Management System (GIMS). Implementation and development of a new Nordic Baltic integrated GIMS with functionalities that allows for incorporation of more data (phenotype/genotype) will be of great benefit for breeders and researchers using plant genetic resources. Efficient use of genetic resources is dependent on an informative database which allows for simple to complex queries, from Boolean searches to more complex queries using combined Boolean searches together with filtering for phenotypic (for example, morphology, disease resistance, yield, quality parameters) and geographic information. In the future there will also be a need to integrate genotypic (genomics) data on the collections. The aims are to fully integrate all information on clonal material from primary collections to clonal archives, develop batch tools for registration of material (including pictures, passport- and phenotype-data), deploy tools to support seed/clone health information (phytosanitary documentation), set up direct links to FAO and ITPGRFA for reporting on PGR, direct export to European (EURISCO) and Global (Genesys PGR and GBIF) databases, provide advanced viewing and filtering methods for phenotypic data, develop capabilities to integrate geographic information, increased ability for Boolean searches across more database tables, prepare for future genotypic (genomic) data on collections, and a “one-stop-shop” for researcher to find and order material from all Nordic-Baltic genebanks.

**Plenary / 11****Glenna****Author:** Dan Still<sup>1</sup><sup>1</sup> CSC

The Glenna2 project aims to provide added value to the Nordic national cloud and data-intensive computing initiatives by supporting national cloud initiatives to sustain affordable IaaS (Infrastructure as a Service) cloud resources through financial support, knowledge exchange and pooling competency on cloud operations. The national cloud platforms support the projects aim to develop an internationally leading collaboration and technological platform for data-intensive computing and through collaboration the project facilitate researchers ability to have access to, analyze, present and share large data sets.

The project also aims, by leveraging the pooled competency, to take responsibility for assessing future hybrid cloud technology and communicate the findings to the national initiatives and also, by pooling national cloud application expert support, create a Nordic support channel for cloud and big data.

**Size of poster:****Open Science / 44****EOSC-Nordic Technical session****Author:** Maria Francesca Iozzi<sup>1</sup>**Co-author:** Ilja LIVENSON<sup>2</sup><sup>1</sup> UNINETT/Sigma2<sup>2</sup> UT.ee

Chair: Maria Francesca Iozzi, Co-chair: Ilja Livenson

EOSC-Nordic is a project recently granted by the EC in the framework of the INFRAEOSC-5b call. It aims at improving coordination of EOSC relevant initiatives within the Nordic and Baltic countries and boosting harmonisation of governance and service management policies across the region in compliance with EOSC agreed standards and practices.

The project brings together a strong consortium of 24 partners including e-Infrastructure providers, research performing organisations and expert networks, with national mandates with regards to the provision of research services and open science policy and wide experience of engaging with the research community and mobilising national governments, funding agencies, international bodies and global initiatives and high-level experts on EOSC strategic matters. Demonstrators are bound to user communities present in the Nordic and Baltic region and cover a variety of scientific domains –climate, computational linguistic, bio-diversity, personalise medicine –and technical challenges, for example cross borders AI/ML platforms, distributed data management and data analysis, cross-border research with sensitive data.

In the present session, you will learn more about the EOSC-Nordic and its main technical objectives and planned activities. A particular focus will be put on the technical strategies for enabling cross-borders research and building an interoperability framework that will be able to make e-infrastructure services discoverable in the region and in the EOSC.

**Size of poster:**

**Workshops II / 22****Introduction to the Rahti container cloud****Author:** Risto Laurikainen<sup>1</sup>**Co-author:** Yacine Khettab<sup>2</sup><sup>1</sup> *CSC - IT Center for Science*<sup>2</sup> *CSC - IT Center for Science Ltd.***Corresponding Authors:** risto.laurikainen@csc.fi, yacine.khettab@csc.fi

CSC has a new cloud platform called Rahti. It is based on OpenShift - Red Hat's distribution of Kubernetes. It is a generic cloud platform that is suitable for a wide range of use cases from hosting web sites to scientific applications. What differentiates it from previous cloud platforms such as CSC's cPouta is the ease with which applications can be managed, scaled up and made fault tolerant.

This workshop will comprise two parts: an overview of Rahti and its features and a practical hands-on part where you can try Rahti yourself.

**Be sure to bring your laptop if you would like to take part in the hands-on exercises!** Optionally, you may also install the oc command line tool beforehand for some of the exercises:

<https://github.com/openshift/origin/releases/tag/v3.11.0> (scroll to the bottom of the page and download a version of "openshift-origin-client-tools" for your operating system).

In the overview part, we'll cover topics like:

- What Rahti is and what you can do with it
- What problems Rahti solves and how it compares to previous cloud platforms
- How to get access to Rahti

In the hands-on part, you get to:

- Start a web application based on a ready-made template
- Run a mock scientific computing job
- Deploy a web application with just one command based on a source code repository

Training accounts in Rahti will be provided for the hands-on session.

**Size of poster:****Workshops I / 26****JupyterHub for research facilities****Authors:** Thor Wikfeldt<sup>1</sup>; Sabry Razick<sup>2</sup>; Radovan Bast<sup>3</sup>; Richard Darst<sup>4</sup><sup>1</sup> *KTH/NeIC*<sup>2</sup> *University of Oslo*<sup>3</sup> *UiT*<sup>4</sup> *Aalto University*

**Corresponding Authors:** richard.darst@aalto.fi, radovan.bast@uit.no, kthw@kth.se, sabry.razick@usit.uio.no

Jupyter notebooks combine the accessibility of an interactive web-frontend, the reproducibility of a laboratory notebook, and the collaborative potential of a cloud-based deployment. The accessibility and interactivity lowers the barrier for researchers to prototype, write, and share data analysis pipelines, and the literate programming approach of Jupyter makes it particularly simple to reproduce, reuse, and adjust notebooks by colleagues and peers.

Jupyter has another use: providing access to remote resources via JupyterHub. Many typical JupyterHub deployments have used cloud-based resources for one-off purposes, but there is also good support for JupyterHub as an interface to HPC clusters and other pre-existing research facilities. JupyterHub can provide a stepping stone for light computing on existing clusters - as well as a more user friendly interface for preparation and visualization for existing power users.

**In this workshop, we will demonstrate the use of JupyterHub and provide guidance so that attendees can set up their own JupyterHub deployments.** There will be a show-and-tell of Jupyter itself and existing JupyterHub deployments. We will go over the basic requirements and practical implementation for a JupyterHub setup. The workshop includes discussion about the difference between traditional batch and interactive workloads, and how the parameters of HPC systems can be tuned to interactive uses. At the conclusion of the workshop, participants will be well prepared to begin deployment of JupyterHub to their own facilities and a Nordic JupyterHub community will begin.

### Pre-workshop

Prerequisites: since we do not go into depth about Jupyter notebooks themselves, we will share links to talks/lessons on basic Jupyter notebooks in an updated abstract so participants can learn and experiment in advance.

- Jupyter: wikipedia
- Jupyter notebooks: video
- JupyterHub: brief description, video

### Workshop outline

1. **Introduction to workshop (5 min):** *Radovan Bast*
2. **Brief introduction to Jupyter (5 min):** What is Jupyter and why is it cool? *Thor Wikfeldt*
3. **JupyterHub  $\neq$  x,  $\forall$  x (10 min):** What is JupyterHub and why? *Richard Darst*
4. **JupyterHub from a sysadmin point of view (20 min):** What does a sysadmin need to know to quickly set up a JupyterHub deployment? What are the challenges and solutions in operating JupyterHub and integrating Jupyter and HPC resources? *Richard Darst*
5. **Example deployments (20 min)** *Csaba Anderlik, Gergely Sipos, Richard Darst*
6. **Panel Discussion (30 min):** Q&A - where do we go from here?

**Size of poster:**

Workshops II / 53

## An Introduction to Deep Learning for Natural Language Processing

**Authors:** Andrei Kutuzov<sup>1</sup>; Stephan Oepen<sup>1</sup>

<sup>1</sup> *University of Oslo*

We will provide an introduction to the application of Artificial Neural Networks to a variety of tasks in Natural Language Processing (NLP), i.e. enabling computers to ‘make sense’ of human language. The tutorial will cover example problems in document classification, sentiment analysis, and sequence labeling; we will introduce common architecture variants, such as deep feed-forward networks, convolutional neural networks, and recurrent neural networks. A basic recollection of linear algebra and Python syntax may be useful, but there should be something for everyone in this tutorial either way. We have yet to decide whether there will be time and opportunity for hands-on exercises.

## Workshops I / 29

# Reimagining research computing

**Authors:** Radovan Bast<sup>None</sup>; Richard Darst<sup>1</sup>; Sabry Razick<sup>2</sup>; Thor Wikfeldt<sup>3</sup>

**Co-author:** Anders Follin<sup>4</sup>

<sup>1</sup> *Aalto University*

<sup>2</sup> *University of Oslo*

<sup>3</sup> *KTH/NeIC*

<sup>4</sup> *LUNARC*

**Corresponding Authors:** radovan.bast@uit.no, kthw@kth.se

In modern times, computation power is becoming more and more important. However, at the same time, the rest of the world is becoming consumerized: while the general expectation is that information technology is easier to use, the design of high-performance computing (HPC) systems has not kept up with modern developments in computer usability. There are many historical artifacts of how HPC systems are set up: HPC systems are often optimized for data transfer over scp, while users often prefer solutions where remote drives are mounted. We expect computations to fit into nice “rectangular” boxes of number of cores × time × memory, while with modern data science workflows, the time and memory can be unknown at the start of a job, and, in particular, interactive usage leads to highly intermittent CPU and memory requirements. Why is knowing Linux shell scripting a requirement for every job when we want our facilities to be usable by anyone? How can we empower users to have more control over their software stack?

In this workshop, we will explore the largest usability barriers in HPC systems, existing solutions, and create a joint vision of a modern HPC system. The first talks will be presentations on vision and usability from invited speakers from both HPC and human-computer interaction (HCI). After that, there will be brainstorming sessions (guided, in small groups, unconference, or panel discussions) where we identify the biggest pain points. Then, there will be group discussions in a speed-blogging format to create a shared vision document which will be the result of this workshop. After this workshop, there should be additional Nordic infrastructure cooperation to improve the accessibility, and possibly standardization, of large computational resources beyond those who traditionally use them.

**“Homework”:** This is an interactive workshop, so please come prepared. Talk to people at your institution and/or other meeting at NeIC. Poll the people around you: what are the biggest issues with using your institution’s computational facilities? Issues can be both general and specific, e.g. “all files have to manually be transferred, but due to the use of ssh proxy hosts there it is difficult from outside the campus network” or “it is easier to pay Amazon than pay us”.

### Workshop outline

- **Unmanned Vehicles, Remote Visualization and Interactive HPC - Rethinking the use of HPC resources** [30 min, Jonas Lindemann, LUNARC (presenter), Anders Follin, LUNARC]

- **Examples of accessibility improvements at various sites** [20 min, Richard Darst (Aalto University), Sabry Razick (University of Oslo)]
- **Unconference introduction** (10 min)
- **Unconference** (30 min)
- break, discuss with others during this time
- **Unconference continued** (40 min)
- **Discussion, presentations by groups, and panel follow-up** (40 min)
- **Concluding remarks** (10 min)

**Size of poster:**

**Open Science / 28**

## Open Science with Sensitive Data

**Author:** Antti Pursula<sup>1</sup>

<sup>1</sup> CSC

**Corresponding Author:** antti.pursula@csc.fi

Many scientific fields are using, or would like to use, personal or sensitive data in the research. Such fields include for example genomics, health, social sciences and language research. The sensitive data that has been cleared for secondary use, should be properly managed and made findable under the same principles than non-sensitive research data. This naturally needs to be done under strict ethical and legal compliance and via secure IT services. However, providing secure e-infrastructure for large cross-border research projects dealing with sensitive data is still in great demand and remains in some extents an unsolved challenge. Moreover, the emphasis on open science and FAIR data by the science communities and policy-makers increase the demand for professional research data management in connection to sensitive data.

This workshop will discuss the current status, opportunities and challenges of secure e-infrastructure services from various angles, and tries to form conclusions as well as inspire action for supporting open science with sensitive data. The topics for the short presentations and panel discussions are selected to highlight different sides of the topic. Topics include for example the following: openness in a sensitive data landscape, experiences from Trygve project both from service provider and user perspective, secure processing of distributed data, impact of NeIC sensitive data activity, as well as Research Data Management and sensitive data.

The workshop programme includes short talks followed by panel discussion on a few selected topics. We plan to use online tools for one channel for addressing questions to speakers as well as enable online surveys as an option for speakers to interact with the audience.

### Agenda

#### 13:30 - 15:00 First session

Chair: Henric Zazzi, KTH

13:30 Trygve developments for secure cloud federation: Antti Pursula, NeIC

13:45 Researcher perspective - the NorTwinCan consortium: Jacob Hjelmberg, SDU

14:00 Sensitive data archiving in Nordics and Federated EGA: Niclas Jareborg, NBIS

14:15 European 1 M genomes declaration: Malin Eklund, VR

14:30 TSD secure service and GDPR: Gard Thomassen, USIT Oslo

14:45 Discussion

15:00 - 15:30 coffee break

**15:30 - 17:00 Second session**

Chair: Antti Pursula, NeIC

15:30 EOSC-hub data policies: Rob Baxter, EPCC

15:45 Perspectives of EOSC-Nordic and EOSC-Hub on sensitive data: Francesca Iozzi, SIGMA2

16:00 Nordforsk Nordic Commons initiative: Juni Palmgren, KI

16:15 RDA activities on FAIR sensitive data: Heidi Laine, CSC

**16:30 Panel discussion** (chair: Tomasz Malkiewicz, NeIC)

Panelists: Malin Eklund (VR), Gudmund Høst (NeIC), Juni Palmgren (KI) and Gard Thomassen (USIT)

17:00 Wrap-up: Antti Pursula, NeIC

**Size of poster:****Workshops II / 54****An Introduction to Deep Learning for Natural Language Processing****Authors:** Andrey Kutuzov<sup>1</sup>; Stephan Oepen<sup>1</sup><sup>1</sup> *University of Oslo***Corresponding Authors:** andreku@ifi.uio.no, oe@ifi.uio.no

Continuation of session

**Workshops I / 46****Reimagining research computing****Author:** Richard Darst<sup>1</sup><sup>1</sup> *Aalto University*

Continuation of session.

**Size of poster:****Open Science / 45****Open Science with Sensitive Data**

Many scientific fields are using, or would like to use, personal or sensitive data in the research. Such fields include for example genomics, health, social sciences and language research. The sensitive data that has been cleared for secondary use, should be properly managed and made findable under the same principles than non-sensitive research data. This naturally needs to be done under strict ethical and legal compliance and via secure IT services. However, providing secure e-infrastructure for large cross-border research projects dealing with sensitive data is still in great demand and remains in some extents an unsolved challenge. Moreover, the emphasis on open science and FAIR data

by the science communities and policy-makers increase the demand for professional research data management in connection to sensitive data.

This workshop will discuss the current status, opportunities and challenges of secure e-infrastructure services from various angles, and tries to form conclusions as well as inspire action for supporting open science with sensitive data. The topics for the short presentations and panel discussions are selected to highlight different sides of the topic. Topics include for example the following: openness in a sensitive data landscape, experiences from Tryggve project both from service provider and user perspective, secure processing of distributed data, impact of NeIC sensitive data activity, as well as Research Data Management and sensitive data.

The workshop programme includes short talks followed by panel discussion on a few selected topics. We plan to use online tools for one channel for addressing questions to speakers as well as enable online surveys as an option for speakers to interact with the audience.

### **Agenda**

#### **13:30 - 15:00 First session**

Chair: Henric Zazzi, KTH

13:30 Tryggve developments for secure cloud federation: Antti Pursula, NeIC

13:45 Researcher perspective - the NorTwinCan consortium: Jacob Hjelmberg, SDU

14:00 Sensitive data archiving in Nordics and Federated EGA: Niclas Jareborg, NBIS

14:15 European 1 M genomes declaration: Malin Eklund, VR

14:30 TSD secure service and GDPR: Gard Thomassen, USIT Oslo

14:45 Discussion

15:00 - 15:30 coffee break

#### **15:30 - 17:00 Second session**

Chair: Antti Pursula, NeIC

15:30 EOSC-hub data policies: Rob Baxter, EPCC

15:45 Perspectives of EOSC-Nordic and EOSC-Hub on sensitive data: Francesca Iozzi, SIGMA2

16:00 Nordforsk Nordic Commons initiative: Juni Palmgren, KI

16:15 RDA activities on FAIR sensitive data: Heidi Laine, CSC

**16:30 Panel discussion** (chair: Tomasz Malkiewicz, NeIC)

Panelists: Malin Eklund (VR), Gudmund Høst (NeIC), Juni Palmgren (KI) and Gard Thomassen (USIT)

17:00 Wrap-up: Antti Pursula, NeIC

**Size of poster:**

60

## **Transportation social activities**

**Social activities / 39**

## **Social activities**

**Workshops I / 40**



## Security in the Nordics

This workshop will identify common needs to share and develop joint security measures among Nordic e-infrastructures. The workshop focus on identifying requirements and solutions for security compliance to protect the infrastructures and sharing of data. The workshop will cover fields of potential joint interests, such as vulnerability management, security assessments, development of security skills, and ways to share critical information on security. The participants are also requested to contribute with suggestions for joint security initiatives.

The target audience for the workshop is security professionals, service managers and persons responsible for external relations and liaisons at Nordic e-infrastructures.

### Workshops II / 32

## Building and Managing Linux Containers for Centralized and Distributed Systems

**Author:** Abdulrahman Azab Mohamed<sup>1</sup>

<sup>1</sup> *University of Oslo*

Linux containers, with the build-once-run-anywhere approach, are becoming popular among scientific communities for software packaging and sharing. Docker is the most popular and user friendly platform for running and managing Linux containers. Singularity is a platform for deploying lightweight containers for HPC systems. Kubernetes is a portable orchestration system for managing containerised workloads. This hands-on tutorial workshop will cover the following:

- Overview of the Linux containers technology
- Docker: Installation, building and managing Docker containers
- Singularity: Installation, building and running singularity containers, and creating singularity containers from Docker containers
- Containers for HPC: using Docker and Singularity containers in HPC job scripts using HTCondor
- Container Orchestration: Introduction to Swarm/Kubernetes and Hands-on

**Size of poster:**

### Open Science / 48

## FAIR and GO FAIR

**Author:** Erik Schultes<sup>1</sup>

<sup>1</sup> *International Science Coordinator, GO FAIR International Support and Coordination Office*

The 15 FAIR Principles have found unusually rapid uptake among a broad spectrum of stakeholders, from research scientists who make data, to e-infrastructures who distribute data, to science funders who track impact of data. Erik will describe the FAIR Principles, their relation to Open data, and review example implementations. This discussion, and these examples will be presented in the context of the International GO FAIR Initiative. GO FAIR is a voluntary community of stakeholders devoted finding consensus on standards and solutions that comprise an emerging Internet of FAIR Data and Services.

**Open Science / 49****FAIRness in practise: How to achieve it?****Author:** Benjamin Pfeil<sup>1</sup>

<sup>1</sup> *Leader of the Bjerknes Climate Data Centre, Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research*

The introduction of the FAIR data management principles in 2016 was a milestone in the field of science. While the terms Findable, Accessible, Interoperable and Re-Usable are well known –the practical implementation remains a challenge and needs the engagement of everyone involved: data managers, long-term repositories and scientists. There are several questions: How to achieve FAIRness for long-tail data? How to achieve it for data from external funded projects? How to ensure that Big Data is not detached from the scientific community? What are the needs to achieve FAIRness? The Bjerknes Climate Data Centre, hosted at the University of Bergen, has a unique specialization: Handling data from the Research Infrastructure ICOS (Integrated Carbon Observation System), implementing services for Copernicus Marine Environmental Monitoring Service (CMEMS) but also handling long-tail data from the scientific community and doing data management for smaller external funded projects. This talk will highlight the various needs and implementation strategies.

**Open Science / 50****National coordination of open access to research data in Sweden****Author:** Sanja Halling<sup>1</sup>

<sup>1</sup> *Senior Research Officer, Swedish Research Council*

The Swedish Research Council has a task from the government of coordinating the national work of introducing open access to research data. Good data management and the FAIR principles are crucial for open access to research data, and as such have been the focus of the coordination work. Criteria for assessing the FAIRness of research data have been produced by the Swedish Research Council and can also be used to support researchers in data management in accordance with FAIR. The Swedish Research Council has also developed the metadata platform RUT (Register Utiliser Tool) to facilitate register-based research supporting the FAIR principles.

**Open Science / 51****FAIR Across –Implementation of FAIR into research practice****Author:** Anne Sofie Fink<sup>1</sup>

<sup>1</sup> *Head of Digital Service, Danish National Archives*

The project 'FAIR Across' took off in informal discussions among representatives from Danish universities, university libraries, the Royal Library and the Danish National Archives (DNA) in the National Forum for Research Data Management. The starting point was the shared view that the FAIR data principles had to be presented to researchers in new ways in order to be implemented into research practice.

The project objectives were to help researchers across a broad range of disciplines understand the FAIR principles and implement actual changes into their day-to-day research practice to make their research data more FAIR. The project conducted an extensive series of interviews with researchers

regarding FAIR and developed a number of well-received promotional products to help lowering the barriers for adoption.

**Open Science / 52**

## **Open discussion on turning FAIR into practice**

**Author:** Anders Sparre Conrad<sup>1</sup>

<sup>1</sup> *Copenhagen University*

**Corresponding Author:** asc@kb.dk

**Research perspectives on AI / 35**

## **Research perspectives on AI, transparency, privacy, law.**

Artificial Intelligence can solve new problems, but AI also brings along new ethical and legal issues, such as: Can I trust the AI model? Can I explain it to others? Is it biased, unfair or even illegal? What is algorithmic transparency or algorithmic fairness? Is privacy good or bad for AI research or vice versa? I will pinpoint some key AI research challenges, of which some are solvable and others seem to have no definitive answer.

**Size of poster:**

**Closing remarks / 36**

## **Closing remarks**

**Corresponding Author:** gudmund.host@nordforsk.org

**Poster / 57**

## **NeIC presents Nordic project highlights**

**Author:** Michaela Barth<sup>1</sup>

<sup>1</sup> *NeIC, KTH*

**Corresponding Author:** caela@kth.se

Through the Nordic e-Infrastructure Collaboration (NeIC) together the Nordic countries are tackling e-infrastructure challenges beyond singular national capabilities. Specific aspects of two current NeIC projects are highlighted on this poster.

Dellingr, NeIC's cross-border resource sharing project, contributes towards establishing a framework

Within the context of the Nordic Large Hadron Collider (LHC) Computing Tier-1 the effective high

**Size of poster:**

A0

**Open Science / 37****FAIR****Author:** Anders Sparre Conrad<sup>1</sup><sup>1</sup> *Copenhagen University***Corresponding Author:** asc@kb.dk

The provision of research data in accordance with the FAIR principles could be seen as a cornerstone of Open Science and will be a major deliverable of the European Open Science Cloud (EOSC). In this session we will explore various aspects of the implementation of FAIR in the Nordic countries, covering perspectives from the researcher to the policy-maker. Topics can include the roles of the national policies compared to university and funder policies, cooperation between universities and service providers, roles of national and international networks and cooperation, outreach to user communities, scalability and long-term sustainability for FAIR data and other implementation aspects of FAIR.

**Size of poster:****Poster / 59****Large scale Nordic Language Processing Laboratory****Author:** Bjørn Lindi<sup>1</sup><sup>1</sup> *NTNU***Corresponding Author:** bjorn.lindi@ntnu.no

Our vision is to implement a virtual laboratory for large-scale NLP research by

- creating new ways to enable data- and compute-intensive Natural Language Processing research by implementing a common software, data and service stack in multiple Nordic Centres
- pooling internationally competitive, data-intensive research and experimentation on scale that would be difficult to sustain on commodity computing resources
- enabling internationally competitive, data-intensive research and experimentation on scale that would be difficult to sustain on commodity computing resources

**Size of poster:****Poster / 55****ELIXIR Cloud & AAI****Authors:** Shubham Kapoor<sup>1</sup>; Susheel Varma<sup>2</sup><sup>1</sup> *CSC*

<sup>2</sup> *EMBL-EBI*

**Corresponding Authors:** shubham.kapoor@csc.fi, susheel.varma@ebi.ac.uk

The ELIXIR Cloud Platforms are a collection of cloud services offered by national ELIXIR nodes to serve research use cases from different ELIXIR communities. ELIXIR Cloud platform leverages all four ELIXIR Compute Platform 2019 -2023 Work Programme pillars (AAI, Data Storage & Transfer Services, Hybrid Cloud and Container Orchestration Services), including building on internal cross-platform collaborative efforts within ELIXIR namely: Tools (Bio.tools, OpenEBench, Biocontainers), Interoperability (CWL, Bioschemas, identifiers.org), Data (Core Data Resources) and Training (Workshop As A Service). ELIXIR Cloud platform focuses on developing its cloud services in a federated manner such that they are compliant to relevant global standards or specifications (e.g. GA4GH, ISO, etc.)

**Size of poster:**

A0

**Poster / 56**

## NeIC Tryggve - cross-border services for human data

**Authors:** Antti Pursula<sup>1</sup>; Abdulrahman Azab<sup>2</sup>; Jonas Hagberg<sup>3</sup>; Niclas Jareborg<sup>3</sup>; Josefine Nordling<sup>4</sup>; Ali Syed<sup>5</sup>; Henric Zazzi<sup>6</sup>

<sup>1</sup> *NeIC / CSC*

<sup>2</sup> *University of Oslo*

<sup>3</sup> *NBIS*

<sup>4</sup> *CSC*

<sup>5</sup> *Computerome*

<sup>6</sup> *KTH*

**Corresponding Authors:** hzazzi@kth.se, josefine.nordling@csc.fi, jonas.hagberg@nbis.se, alisyed@dtu.dk, azab@usit.uio.no, antti.pursula@csc.fi

Research in biomedical sciences aims ultimately at curing diseases and improving quality of life. Successful research on the field requires the use of human data of various types and from various sources. However, working with human data requires added security measures to ensure privacy protection for the research subjects. Nordic e-Infrastructure Collaboration NeIC and the ELIXIR nodes in four Nordic countries (DK, FI, NO, SE) have done long-standing collaboration to develop secure IT infrastructure in order to support cross-border collaborative research utilizing human data in the Nordic countries.

The NeIC Tryggve project is focusing currently on three areas: federated archiving of sensitive data, computation with sensitive data across secure platforms, and implementation of use cases. The federated archiving of sensitive includes technological development for Federated EGA activity. In the computation across platform, the vision is to create a Nordic secure platform where each node can send and accept workflows across Tryggve sites, under strict security and utilising standard workflow execution frameworks. The use cases are driving the development and highlight several of the developments mentioned above.

The available secure computing environments in the participating countries provide both remote desktops or infrastructure level access to the resources. The Tryggve project gives support in accessing these systems and additionally develops tools and legal framework that make their joint utilization easier for transnational research teams. The outcomes of the project are available from the project's web site <https://neic.no/tryggve>.

**Size of poster:**

A0

Poster / 58

## Bridging RDA and Danish Research Communities

**Authors:** Anne Sofie Fink Kjeldgaard<sup>1</sup>; Zaza Nadja Herbert-Hansen<sup>1</sup>

<sup>1</sup> *Danish National Archives*

**Corresponding Authors:** asf@sa.dk, nadhan@fiskeristyrelsen.dk

Presentation of the Danish Research Data Alliance Node

Zaza Nadja Herbert-Hansen and Anne Sofie Fink, Danish National Archives, project management for DK- RDA-Node

The Research Data Alliance (RDA) builds the social and technical bridges to enable the open sharing and re- use of data. The RDA Vision is that researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

The poster aims to present the Danish National Research Data Alliance Node.

Objective for the national node is to:

- Encourage and support the uptake of RDA recommendations and outputs in Denmark
- Foster active engagement in RDA activities such as RDA working groups and interest groups by Danish researchers and research support organisations
- Coordinate with relevant Nordic and European organisations on RDA related results and activities
- Use of existing national network for RDM (National Forum for RDM) and DeIC (Danish e-Infrastructure Consortium)
- Danish RDA ambassador supports the node

The expectations for our node activities are to:

- Increase knowledge about RDA and use of RDA in Denmark
- Encourage synergy between national activities and international activities on RDM, Open data, FAIR data and services etc. synthesized and utilized by the Danish RDA node
- Support Danish researchers, data professionals etc. in RDA partaking
- Coordinate with other Nordic RDA national nodes and research infrastructures

Sustainability for our national node will be that RDA stays a resource for Danish research communities supporting sharing and re-use of research data in the future.

**Size of poster:**