

Til:

Gard Thomassen,
USIT/ITF

Fra:

Fagråd for eInfrastruktur

Egenevaluering av eInfrastrukturmidler 2017-20

Ordningen med eInfrastrukturmidler ble opprettet i vedtak i universitetsstyrets møte 20.-21. juni 2016 med virkning fra budsjettåret 2017. Fagråd for eInfrastruktur ble satt til å administrere denne og har fordelt tilsammen 44 millioner kroner fordelt i perioden 2017-20. Hensikten med midlene var å utvikle eInfrastrukturressurser, -kompetanse og -støttetjenester for å fremme universitetets forskning.

I universitetsrådets vedtak forutsettes det at ordningen skal evalueres etter 4 år. Fagrådet har startet dette arbeidet med en egenevaluering av ordningen og fagrådets arbeid i denne perioden, se [Fagråd for eInfrastruktur i 2016-20](#).

I den anledning har fagrådet startet et arbeid med å innhente nærmere informasjon om bruken av eInfrastrukturmidlene fra de som har fått tildelt midler i 2017-19. Fagrådet ønsker at det redegjøres for hvordan midlene er brukt, hvilke resultater som er skapt og at gjøres noen vurderinger av effekten av bruken av midlene sett i forhold til hensikten med eInfrastrukturmidlene.

I årene 2017-19 ble det brukt 16,170 millioner kroner til hub/node-organisert støtte for IT i forskning. Tre av de hub/nodene som fikk tildeling i 2017 søkte om forlenget tildeling i 2020 og 2021 og rapporterte i den anledning ved å svare på nedenstående spørsmål.

Tildeling til AI/ML/DL-hub/node:

Dere fikk 3,5 millioner kroner fordelt på 1,5 millioner i 2018 og 2,0 millioner kroner i 2019 til en hub/node for AI/ML/DL og ber dere svare på spørsmålene fagrådet ønsker å få svar på til egenevalueringen.

1. Hva er de tildelte midlene brukt til?

Angi hvordan midlene tildelt i perioden 2018-19 er benyttet (i hele 1000 kr). Oppgi eventuell egenfinansiering og om midler er fordelt til noder.



Formål	2018	2019	Sum
Lønn, frikjøp	0 kr	785.691 kr	785.691 12kr
Anskaffelser, investeringer	486.550 kr	778.443 kr	1.264.993 kr
Kompetanseutvikling	68.855 kr	56.211 kr	125.066 kr
Andre kostnader	2.759 kr	3.114kr	5.872 kr
SUM	558.163kr	1.623.459 kr	2.181.622 kr
Egenfinansiering	300.000 kr	200.000 kr	500.000 kr
Fordelt til noder	0 kr	0 Kr	0 kr

2. Hva slags nettverk er bygd opp?

Redegjør for hvilke noder som inngår i hub/noden, om det har vært samarbeid med andre hub/noder og hvilke andre som har vært involvert i virksomheten.

1.1. Oversikt over nodene

Main partners were USIT and the University Library (UB). While project leadership was formally located at USIT, USIT and UB collaborated equally together without imposing unnecessary hierarchy. Instead of establishing other formal nodes, the project engaged with different users and groups of users via different channels (chat room, email list, web page) and events (workshops, project status meetings, AI lounge).

USIT was responsible for building up the technical infrastructure, provided support to users, co-organized workshops and the AI lounge, and met with different users and user groups to discuss their needs.

UB co-organized workshops and the AI lounge, and met with different users and user groups to discuss their needs.

1.2. Samarbeid med andre hub/noder

The project didn't formally work together with other projects at the hub/node level, but rather encouraged and engaged in direct interactions with users and IT staff from all UiO departments and was constantly establishing new contacts using different dissemination channels (AI lounge, email list, chat, web page, workshops).

AI lounge: We organised 11 events with various UiO internal and external speakers drawing audiences from 5 to 50 participants.

1.3. Involvering av IT-organisasjonen sentralt (USIT) og lokalt

USIT's division ITF had the lead role in the project. Other sections of USIT were involved in hardware procurements, setting up the hardware and invited to participate in regular AI lounge events.

To establish connections to other IT staff at UiO, lokal-IT was informed about upcoming events such as AI lounges. This resulted in good discussions with local-IT about requirements and experiences made at the department level.

We hosted two internal seminars with scientists (November 2018 and June 2019) to inform about project and its status as well as to discuss way forward. This is a good format to exchange status and ideas about what services are provided and what needs to be done to improve their usability for users.

1.4. Andre involverte i hub/noden

USIT: Anne Schad Bergsaker, Gard O Sundby Thomassen, Ole Widar Saastad, Fredrik Magnussen, Sabry Razick, Bjørn-Helge Mevik, Knut Waagen, Thomas Röblitz, Rune Hamre

UB: Andrea Gasparini, André Walsøe, Ahmed Abdi Mohammed

3. Hvilke resultater er skapt?

Beskriv hvilke resultater i form av kompetansebygging, eInfrastrukturressurser og -tjenester, tekniske løsninger, møteplasser og samarbeidsarenaer med mer som er utviklet i regi av hub/noden i 2017-19

1.5. Kompetanseutvikling

An example of the competence building achieved already in spring 2019 is visible through the technical report "Towards AI services à la carte: Status and Future Work at University of Oslo" (see [PDF](#)).

We organized five workshops to train users:

- 2019 January 11th: Full-day Research Bazaar workshop titled «Exploring Research Data with Artificial Intelligence and Design Thinking» jointly organized by UB and USIT. About 20-25 participants. UB provided the science case and conducted the training while USIT provided the infrastructure for hands-on exercises.

- 2019 March 20: Full-day workshop about using the MS Azure cloud infrastructure for AI tasks. About 20 participants. USIT and UB hosted the workshop while training was provided by Microsoft & Leading Learning Partner Association.
- 2019, November 19: Techno-day, where iris.ai, Keenious.com, and yewno.com were presented.
- 2020 January 8: half-day research bazar about using the ai-based research tool IRIS.ai. About 22 participants.
- 2020 February 12th: Full-day workshop about using the Google Cloud Platform. About 30 participants. USIT hosted the workshop while training was provided by Google.

Through all the activities on hardware infrastructure, organizing the AI lounges, meeting users and developing new services employees at USIT and UB built up competence for future activities in this field. Particularly:

- USIT staff obtained valuable knowledge on designing a local infrastructure to support research employing AI, that is, what kind of services are needed and how they need to interplay with other services being available locally or externally. USIT has gained experience in different hardware choices, sizing hardware, how to set them up and how to operate them in a stable and efficient way.
- USIT has explored different hardware choices (different GPUs from NVIDIA and AMD) and investigated means to integrate machines in a virtualized platform (nrec.no). All together these efforts provide us with greater flexibility in choosing hardware and richer means to fit operations and service types being offered better to the needs of users in research and teaching.
- USIT has also explored the use of the hardware in training our own neural network for classifying images from the museum image database. This was done both to improve our own experience with using the hardware to be able to provide better user support, but also to test the capabilities of the hardware and potentially create a model that can help classify images in the extensive photo databases collected by the university museums over many years, as such classifications are prone to human error.
- By frequently meeting with scientists and lecturers we developed a better understanding of what kind of resources, software frameworks and support services users need. Where possible we actively supported teaching in classes such as the course Jus5671 (law faculty, Prof. Malcom Langford, spring 2019 and spring 2020) and the course IN5400 (department of informatics, Prof. Anne Solberg, spring 2019).
- Together with NRK we developed a prototype for audio-to-text transcription. Such capabilities are available from public cloud providers but require that the audio to be transcribed is sent to the provider. Sensitive data, however, must not be sent to external

providers, hence we started developing a prototype. The goal was not only to provide such functionality to our users, but to also gain hands-on experience in a typical deep learning project.

- By organizing the AI lounges, UB helped in forming an arena for bringing everyone interested in AI - researchers, IT staff and students – together to get to know each other, to exchange ideas, and to foster new collaborations.

1.6. eInfraressurser og -tjenester

The AI HUB contains 3 servers with 4 NVIDIA RTX 2080 Ti GPUs each. These servers have been in production since December 2018. The servers run Ubuntu 16.04 as operating system and standard software packages for machine and deep learning have been installed on them. Every UiO member is eligible to use these servers. We have seen somewhat varying loads on the servers – while fully loaded in some periods, they are only lightly loaded in other periods. Generally, user feedback about these resources was very positive. However, it was noted that the overall capacity should be increased. Temporarily we run Jupyter hub services on all three of them to support the course IN5400 (see below). Making these servers available early in the project resulted in some issues such as too small local disk and difficulties in integrating the machines into USIT's server infrastructure. However, we could also gain knowledge which will help us in the future to operate such machines more effectively. In addition to these three servers, we also procured 2 servers for testing purposes. One is configured with an NVIDIA T4 GPU, while the other one has 2 AMD Vega64 GPUs. These servers were mainly used to test different means to setup and use GPUs in a virtualized platform. Particularly, the machines were integrated via vGPU support and passthrough support in the NREC. While first tests were promising, this direction needs continued effort before GPUs under the NREC could be put into general production.

We provided different support for lectures at the law faculty (course Jus5671, Prof. Malcom Langford, spring 2019 and spring 2020) and at the department of informatics (course IN5400, Prof. Anne Solberg, spring 2019). We anticipate more requests for such kind of support in the future.

1.7. Møteplasser, samarbeidsarenaer

To kick off the project we organized a workshop titled “AI-based data-driven science” on September 7, 2018 (see [workshop page](#) for details). The workshop draw participation by researchers and IT staff from all over UiO and served as a good start to understand use cases and requirements and also to establish contacts to potential users.

AI lounge: organized 11 events to let everyone with interest in the field learn something new, meet to exchange ideas, and possibly start new collaborations.

List of AI lounges

- November 19, 2019: Smart Knowledge Services

- October 22, 2019: Bio-Inspired Robotics @ UiO
- October 15, 2019: Big Data - Intelligent Chances
- September 24, 2019: Big Data - Intelligent Chances
- August 27, 2019: Real world machine learning
- May 28, 2019: Automated Acoustic Data Processing with Deep Learning
- March 26, 2019: Image processing and Deep Learning
- February 26, 2019: Natural Language Processing at NRK and IFI
- January 29, 2019: Hands-on with the new GPU servers of the AI HUB
- December 18, 2018: Machine learning for predicting antibody recognition in silico
- November 20, 2018: Status of AI activities at UiO

Meetings with individual researchers/research groups to inform about the project, to learn about their needs, to possibly help them getting started with AI and to facilitate their use of the available resources:

List of meetings

- AI services at Intervensjonsklinikken, January 29, 2019
- Resource needs for Greiff lab and Bioinformatics at IFI, January 29, 2019
- Student administration, September 27, 2019

1.8. Annet

The project partners participated in several external events (some to build up competency and some to present the work of the project):

- 2019 NAIS Symposium, Trondheim, May 27/28, 2019, poster «Towards AI Services à la Carte: Status and future work at University of Oslo»
- UiB Symposium on AI and ML, Solstrand, March 25/26, 2019, presentation “AI hub-node project at UiO”
- Geilo Winter School, January 20-25, 2019 Learning from Data
- Fantastic Futures, Oslo, December 5, 2018
- AIM2 North Symposium, Oslo, November 7, 2018
- Geilo Winter School, January 21-26, 2018 Practical Artificial Intelligence

In 2018, we developed and coordinated an NFR INFRASTRUKTUR proposal called NORFAML (Norwegian Facility for Applied Machine Learning) with partners UiO, UiB, UiT, Sigma2 and SINTEF Digital to build a national infrastructure for applying machine learning and deep learning in research. Besides the five main consortium partners the proposal included several additional associated partners from academia and IT industry.

Presentation of the project at:

- 4th Turning the Tide of Antimicrobial resistance (TTA) meeting on “Tuberculosis & Machine learning”, November 20, 2019, presentation “Big and sensitive data – status and future outlook for TSD and AI computing resources at USIT”
- Bioinformatics meeting, Department of Informatics, University of Oslo, November 19, 2019, presentation “Towards AI services à la carte: Status and future work at University of Oslo”
- Workshop “AI-based data-driven science”, September 7, 2018, presentations “AI hub-node project”, “NFR project proposal” and “Use of artificial intelligence at the Oslo University Library”

4. Hvilken effekt har resultatene av hub/nodens virksomhet hatt?

Beskriv kort hvilke effekter virksomheten har hatt for forskere og forskningsgrupper of for forskningen ved universitetet.

AI HUB resources have been in use since December 2018 by about 50 students (master/PhD) and researchers (PostDoc, Prof). Some papers have already been published using the AI HUB machines.

We supported two courses: one with access to the AI HUB machines via Jupyter Notebook and one with lectures about using MS Azure.

In general, activities where we have gathered internal and external experts, has been crucial to support the development of the NFR proposal.

5. Hva er de viktigste lærdommene og erfaringene fra arbeidet?

Med henblikk på framtidig utvikling av hub/node-virksomheten, beskriv kort hva som er de viktigste lærdommene og erfaringene – positive såvel som negative – å ta med seg fra hub/nodens virksomhet disse årene.

The work has shown that with relatively little effort needed resources, support for teaching new technologies and establishing a forum for university-wide exchange can be achieved even in a short time provided that there is critical mass in both interested parties/people and funding. It is easy to bootstrap an experimental service, but it is much harder to move even parts of it into the

portfolio of generally offered services. It is important to anchor new services well within the main units which shall operate these, and similarly a continuous dialogue with future users is needed to develop and offer the service that is actually needed. As the university is a large organization, it is likely that someone else in some unit may be doing similar work. The work of particularly infrastructure projects may benefit from reaching out to relevant (service) units very early in the project (or potentially integrate them in some way into the project) to allow for effective collaboration.

6. Hva er hub/nodens exit-strategi?

Skal resultatene av virksomheten videreføres i noen form og i tilfelle hva skal videreføre, hvordan er det tenkt skje og hvem vil ha ansvar for drift og videreutvikling.

USIT will continue to operate the AI HUB resources and possibly integrate them into a new local infrastructure supporting researchers. As demand for such types of resources will likely increase in the future, there is a need for future funding satisfying needs centrally at UiO. Also, the way the resources are offered may (servers with interactive access) require further development to support different needs in the best way (interactive access, Jupyter notebooks, batch system plus additional services such as development support with running continuous integration (CI) tasks on GPUs). It is very important to align the central infrastructure for AI with what is available at the department level both in terms of how resources are operated, what user environment is provided and what support is offered to enable researchers always use the best fitting resource and possibly scale up/down their work easily.

The AI lounge has proven to be a good format to attract speakers and enable exchange of ideas and should therefore be continued as well. Both USIT and UB should continue this event series in the future too. Possibly the AI lounge would benefit from being coordinated with other seminars or events in the field.

In addition, one should look into the possibility to include some of the HUB activities in new digital scholarship center (DSC).

Both the operation of the AI HUB resources including basic support and running AI lounge events could serve as a good means to establish new contacts and further discuss needs with users.

7. Hva er de viktigste lærdommene og erfaringene?

Infrastrukturmidler eller tilsvarende vil bli videreført i en eller annen form. Beskriv kort hva som er de viktigste lærdommene og erfaringene å ta med seg i videre arbeid med denne formen for finansiering av tiltak.

The programme enables researchers and IT experts to work closely together on emerging topics which might otherwise not be possible (due to lack of time, due to lack of critical mass, due to lack of funding to acquire necessary hardware). The success of the programme might be increased

- if the programme and its funded projects would be known more publicly,
- if the projects were enabled to reach out to relevant partners and users more easily,
- if the exchange among the different projects funded through the programme would be facilitated (or even required),
- if there is a default option to carry on some of the work of a project beyond the end of the originally envisioned project duration (at a reduced cost/budget) with an emphasis on increasing the reach of project.