# Minutes – UiO AI hub-node project kickoff

## Date

November 6th 2018, 09:00-12:00

## Location

Kristen Nygaards hus, Gaustadalléen 23 a, https://www.uio.no/om/finn-fram/omrader/gaustad/ga04/ in room 2410

## Agenda

09:00 - 09:10 Welcome, goals of the kickoff

09:10 - 09:30 Overview of AI-hub node project (goals, project period, status), results of the September AI workshop and intro to NORFAML proposal

09:30 - 10:20 Mapping use cases vs competences; and prioritization of outcomes for the coming 6-12 months

10:30 - 11:30

- What can we learn from others (providing assistance in ML, structured way in providing assistance)?
- What is the best working mode (continuous joint work ("sitting together"); just a meeting at start, then separately)?
- For establishing a forum: How do we involve all interested?
- Existing IT hardware and gaps?
- Awareness of legal issues (GDPR)?

11:30 - 12:00 Wrapping up, follow-up meeting

## Participants

- Ole Saastad, USIT/ITF
- Sabry Razick, USIT/ITF
- Milen Kouylekov, USIT/ITF
- Gard Thomassen, USIT/ITF (until about 10:30)
- Thomas Röblitz, USIT/ITF
- Ahmed Mohammed, UB
- Andrea Gasparini, UB
- Geir Kjetil Sandve, IFI/Bioinformatics
- Vegard Antun, Mathematics
- Kristin Skar, IFI/Head of IT
- Håvard Kvamme, Mathematics/BigInsight
- Geir Storvik, Mathematics/Data Science master
- Charles Martin, IFI/ROBIN
- Manuela Zucknick, OCBE
- Tony Håndstad, Medical Genetics
- Nermin Zecic, OUS (from ca. 10:00)

## Minutes

### Welcome, goals of the kickoff & Overview of AI-hub node project (09:00 – 10:45)

The status and goals for the project as well as near term (6-12 months) tasks and milestones were presented.

**Three main activities were outlined with the following tasks and milestones:**

1. Basic infrastructure: Hardware & services
   - Procure and setup hardware (3 servers with NVIDIA RTX 2080 Ti, 128 GiB RAM): 12/2018
   - Procure and setup 1-2 servers with alternative GPU options (e.g., AMD, Huawei): 06/2019
   - Establish access to UH-IaaS and apps.sigma2.no: 03/2019
   - Establish access to external Clouds: 09/2019
2. Competence built up
   - Joint work on use cases (e.g., support in teaching or research projects): continuous activity with short term projects (~ 3 months), first results expected 03/2019
   - Best practice guides: continuous activity, first results expected 09/2019
3. Connecting people, teaching users
   - AI lounge (casual atmosphere, talks, discussions, hands-on on any topic related to AI at UiO), regular event starting once a month: first Nov 20, 2018, 15:00-16:00, Status of AI activities at UiO
   - Courses could be integrated into AI lounge or HPC course week, seeking collaboration with Carpentry to offer high-quality courses: first "course" offered as part of Research Bazaar (Jan/2019)
   - Web page as a starting point for everyone interested in AI to learn what resources are available, what events happen, who has expertise, etc: continuous activity, first results to be expected after first AI lounge
   - Email lists
     - Announcements: itf-ai-announcements@usit.uio.no
     - Getting support: itf-ai-support@usit.uio.no

**Main questions (Q&A)/comments (C) were:**

- Q: What does "how to apply machine learning" means? A: IT staff shall through joint work gain experience in advising users (researchers, lecturers) in technical aspects applying ML. They won't advise on scientific aspects. The "service" would span basic support as well as advanced user support comparable to what IT staff is capable and doing for parallel computing (HPC) today.
- C: Simply advising on what software frameworks to use, and overview of basic parameters would be a good start. In addition, advise on whether GPUs or simply more CPUs should be used would be good too. Generally, beginners should focus on using high-level frameworks.
- Q: What is role of non-USIT participants (in this meeting)? A: Need guidance from actual users on what is important (resources, competences) and when this should be available. Also, need contacts to discuss requirements in detail, and jointly work on use cases to solve problems and to gain experience.
- C: On hardware choices/strategy: ITF tries to procure cost-effective hardware. Support to efficiently use these via popular software frameworks such as TensorFlow is a must. Budget is

limited, so not all requirements will be met. Will also try to gain experience on alternatives to NVIDIA.
- C: On external clouds: Will benefit scenarios with highly volatile capacity demands (e.g., teaching). Strong interest in ITF to provide access, however, basic legal work needs to be done. Hope, to be able to offer something within 2019.
- C: ITF will try to make existing resources (at UiO and Sigma2) available as easy as possible (e.g., with pre-configured containers/VM images). Existing resources at end of 2018 will be: 3 servers with NVIDIA RTX 2080 Ti cards, NIRD toolkit (apps.sigma2.no) with NVIDIA V100. In 2019 more resources will become available.
- Q: At whom are courses aimed at? A: Start with beginners ("training on how to use available systems and frameworks to get started"). Later may add more advanced courses.

**Mapping use cases vs competences; and prioritization of outcomes for the coming 6-12 months (10:55-11:45)**

A list of tasks grouped into four categories – Use Case, Data Organization, Model Learning and Deployment – were shown. The participants discussed which tasks are important and how the project should engage with users to obtain experience in supporting these tasks.

Below, the major discussion points are listed:

- Q: Where should we start? A: Integrating data is not the most important point. Easy access to data is important.
- Q: How big is the data? A: TiBs.
- Q: How to start? A: Something users cannot do (easily) without USIT.
    - Support for teaching environment.
    - Containers pre-configured and maintained with needed software frameworks (e.g., Pandas, scikit-learn). Need more documentation on how to use containers.
    - Some systems (e.g., Fram) seem not well-suited for team work.
    - Need advice on details how to use frameworks.
- C: PhD students need to go out of UiO to get things done (GPU and processing power).
- On data organization:
    - C: Reportedly, a huge fraction of overall time spent on using machine learning is spent on working with data.
    - C: It's hard to get access to data.
    - C: Preprocessing seems not to require much time though.
    - C: Data preparation and data quality could be topics for courses.
    - C: Better understanding of the data (domain knowledge) is important.
- Operations of resources:
    - C: Ensure that unused resources are not blocked by users.
    - C: Would be good to know how much of a resource a user is using.
    - C: Depending on demand vs capacity, there needs to some mechanism to allocate resources to users/projects.
    - C: Goal is to have the resources under OpenStack-based UH-IaaS which provides the greatest flexibility. Then, easy to use pre-configured images need to be offered and maintained.
    - C: Staged setup: when running fine on test cases, get access to larger production environment.
- C: Good visualization tools may help in understanding data (and results).

- C: Jupyter notebook looks nice but is messy.
- C: In a shared environment, reproducibility is a concern.
- C: DevOps and good practices may help to guide users (and admins).
- C: Sometimes answers to questions seem to change. Would like to have one place we can point students to.
- C: Knowledge of "what can go wrong?" could be helpful in falling into (un)known traps.
- Q: Do you depend on low-level frameworks such as CUDA? A: No, high-level frameworks, e.g., TensorFlow, are sufficient.
- C: Two groups of users – one using existing methods via high-level frameworks, one developing new methods. Latter uses languages such as R to demonstrate that a method works, but usually not implement it for efficiency. In some cases, only a small part of the code would need to be implemented for efficiency and such implementation could be called from a higher-level language (however, this seems like an advanced topic).
- C: Porting code from high-level languages (e.g., R) to ML frameworks might be good use case.
- C: Could be interesting to offer "ML as a Service".
- Q: Will there be support for Spark? A: Little needs from users. Some support on Abel (as module) and NIRD toolkit (apps.sigma2.no). Could connect users to group at Uni Research Computing to get advice.
- Q: Why does the project invest in hardware? A: Still cheaper for sustained loads than Clouds. May also be better to customize/tailor environments (at higher personnel costs though). Important for retaining competence in-house.

Many good comments and questions. However, it seemed too early to define a concrete use case or task to start with. Hence, the best way forward seems to setup individual meetings and start a discussion. ☺

### Wrapping up & follow-up meeting (11:45-12:00)

Main actions for the project:

- Start individual conversions with potential users in research and education.
- Maintain information about resources, events and activities at a web page.
- Pursue tasks & milestones as suggested.

Minutes will be provided. The participants showed interest in meeting again after 6 months to review progress of the project and to provide advice on its future course.