

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.

Tildeling til tjeneste for data- og koderøkting:

Dere fikk tilsammen 2,94 millioner kroner til tjeneste for data- og koderøkting fordelt på 1,94 millioner i 2018 og 1,0 million i 2019 og ber dere svare på spørsmålene fagrådet ønsker å få svar på til egenevalueringen.

1. Hvordan er de tildelte midlene brukt?

1.1. Disponering av midler

Formål	2018	2019
Lønn, frikjøp	2 940 000 kr	1 000 000 kr
Anskaffelser, investeringer	kr	kr
Kompetanseutvikling	kr	kr
Andre kostnader	kr	kr
SUM	2 940 000 kr	1 000 000 kr
Egenfinansiering	kr	kr



2. Hvilke resultater i form av eInfrastrukturressurser er skapt?

Når det gjelder bruk av midlene tildelt for 2018 vises det til rapporten (vedlagt) sendt inn i 2019 som redegjør for frikjøpet av 2 stillinger. Begge er nå (2020) fast jobb på USIT i Underavdeling for IT i forskning (ITF).

Frikjøpet i 2019 har i stor grad gått til lønnsmidler for en ansatt som har en doktorgrad i matematikk og som bistår i alle faser av forskningen: innsamling av empirisk materiale, datarøktning, programvareutvikling, metodeutvikling og bidrag (med statistikk-kunnskaper) under artikkelskriving.

1.2. Kompetanseutvikling

Utviklet og holdt kurs i statistikk-pakken Stata og Python. Opplæring i kjøring av Python-kode utviklet av USIT på eget forskningsmateriale.

Bistått kurs i SPSS og R gitt av kolleger. En erfaring vi har gjort oss i forbindelse med kursvirksomheten er at forskere og studenter med lite forkunnskap trenger mye assistanse i begynnelsen, før de i det hele tatt kan ta verktøyene det kurses i i bruk.

1.3. Tjenester, verktøy for forskningen

Som det framgår av listen over forskningsprosjekter nedenfor under pkt 3., er de fleste av verktøyene og tjenestene som er utviklet knyttet til Python, PostgreSQL og R både innenfor og utenfor TSD.

1.4. Annet

Kursoversikt:

- Introductory course in Stata held three times in 2018, twice as regularly each semester, and once for PhD students at Aker Hospital. 23 signed up in spring (in Norwegian), 10 in fall (in English). Additionally a few participants from OUS and OsloMet. Detailed slides with screenshots of menus etc. were distributed to everyone signed up. Topics: Data management, reproducible syntax, descriptive statistics, some estimation, help resources. Faculty of Medicine has recently switched PhD education from SPSS to Stata, hence increased interest is expected, see course in [spring 2019](#) and [autumn 2019](#).
- Developed [introductory programming course in Python](#). Aimed towards humaniora/social scientists. Focus on processing widely used text data formats (.txt, .csv, .html, .xml, .json), e.g. extracting and structuring information from large amounts of such data. More than 90 people signed up. Slides and example code was provided to participants. Held three times in January 2020, in English and Norwegian.

3. Hvordan er forskere, forskningsmiljøer eller andre involvert?

1.5. Forskere og forskningsmiljøer

Her følger en liste (på engelsk) med de forskningsprosjektene "frikjøpet" har assistert/deltatt i:

- Asimina Michailidou at ARENA Center for European Studies. Harvesting and processing large amounts of text data from newspapers and parliaments. Data harvested from web sites in various formats (pdf, html, json). Structured and uploaded to searchable database. Topical classification based on keyword search (Various Python tools). Exploration of alternative approaches to classification
- Bjørn Høyland at Department of Political Science. Structuring text data from pdf-files on EU parliament voting. Continued from 2018: Scraping and structuring of additional data types with Python and Apache Tika.
- IDART study lead by Thomas Kristiansen, OUS: Developed data management solution for sensitive data in TSD. Pseudonymisation, data joining, data cleaning, automatically including new data (Stata scripts). Co-author on submitted journal paper.
- Museum of Cultural History/USIT: Develop prototype deep learning model for image classification. Separating images of excavation sites from images of artifacts (Tensorflow, Python)
- Nortras at Department of Psychology, Hanne Strømme: Preliminary meetings about new interface to Nettskjema data in TSD. This started up properly in 2020 as an independently financed project.
- National Center for Suicide Research and Prevention, Martin Øverlien Myhre: Preliminary meetings about data management in TSD (PostgreSQL, R). This started up properly in 2020 as an independently financed project.
- For the project Clinical course of neck pain in chiropractic patients, lead by Nina Kørpe Vøllestad, Department of Health Sciences: Assistance with Stata programming, data analysis. Co-authored one submitted journal paper.
- Flora Nah Asah at Scholarly Health Informatics Learning, Institute of Informatics: Statistical analysis of questionnaire data on usability of a data collection tool in Ghana. Co-authored one conference paper.
- Einar Strømnes, Museum of Natural History: Performed statistical analysis of monthly parasite count data in fish.
- Lise Ulvestrand at Department of Psychology: Advice on Nettskjema/TSD and data collection.
- Kaja Knudsen Bergo, Faculty of Medicine: Advice on outliers
- Camilla Udjus, Faculty of Medicine: Advice meeting on statistical analysis (Small longitudinal study)
- Rikke Myrer Lund at Department of Psychology : Advice meeting on statistics for Work- and Organisational Psychology (SPSS)
- Jenny Victoria Tran, Faculty of Medicine: Advice meeting on statistics (trends in count data)

- Sølvi Taraldsen, Faculty of Medicine: Advice on Nettskjema data in TSD. Help with SPSS performance issues in TSD
- Kristin Torgersen, Faculty of Medicine: Debugging Stata code
- Claudia Barth, Faculty of Medicine: Advice meeting on statistics. Count data with inflated zeros.
- Sissel Jørgensen, Department of Chemistry: Help implementing statistics model in Python, and general advice on statistical software
- Para-NASPP eStroke app: Project coordination TSD/app-development. Independently financed.
- Anders Meland at Flymedisinsk Institutt. Advice meeting on statistics (SPSS, general descriptive statistics)
- Tone Sjøheim, OUS: Advice on setting up data collection system, TSD
- Johanna Damaris, Department of Culture Studies and Oriental Languages: Meeting about processing and analysis of handwritten texts and metadata from Trolldomsarkivet
- Jannicke Borch Myhre, Faculty of Medicine: Advice on data collection for Norkost 4. Continued as independently financed project in 2020.

1.6. Andre involverte

1.7. Målgruppe og brukere

Forskere og høyeregradsstudenter ved UiO.

4. Hvilken effekt har resultatene av tiltaket hatt?

1.8. Hva er effekten av tiltaket?

Kompetanseheving og muliggjøring av spydspissforskning.

1.9. Hvor mange brukere?

Python-kurset som ble utviklet og gitt i januar 2020 fikk 90 påmeldinger. Stata-kursene, som går jevnlig gjennom året, har mellom 8 og 15 påmeldte.

5. Hva er de viktigste lærdommene og erfaringene?

Midler av denne typen, som gis over en kortere periode med konsentrert fokus, er viktige for å få i gang forskningsmiljøer/-prosjekter som selv måtte ha opparbeidet seg den nødvendige kunnskapen om metoder og verktøy, noe som isolert sett hadde gått på bekostning av utforskingen av selve forskningsspørsmålene.

Selv om man alltid, ideelt sett, skal starte med forskningsspørsmålene, er det mer og mer vanlig å skjele til verktøy og metoder, for å, på et tidlig stadium i forskningen, gjøre seg tanker om hvilke spørsmål man kan håpe å få svar på innenfor tiden satt av til forskningsprosjektet.