

Sigma2s strategi

May 2018

UNINETT



Sigma2s visjon

«To maximize the impact and return of scientific research by providing a predictable, sustainable and cost-efficient e-infrastructure»

- UNINETT Sigma2 AS administrerer den nasjonale infrastrukturen for beregningsvitenskap i Norge og tilbyr tjenester innen tungregning og datalagring. Arbeidet er organisert i infrastrukturaktiviteter, finansiert av Norges forskningsråd og samarbeidspartnere. Virksomheten drives ikke-kommersielt. Sigma2 er et datterselskap av UNINETT og har hovedkontor i Trondheim.
- UNINETT Sigma2 tilbyr tjenestene tungregning og datalagring til individer og grupper involvert i forskning og utdanning ved norske universiteter og høyskoler, samt andre organisasjoner og prosjekter med offentlig finansiering.
- Oppdraget er gitt av Norges Forskningsråd, i samarbeid med Universitetet i Bergen, Universitetet i Oslo, Universitetet i Tromsø Norges arktiske universitet og Norges Teknisk-Naturvitenskapelige Universitet. I tillegg koordinerer selskapet den norske innsatsen i europeiske forskningsinfrastrukturer som Partnership for Advanced Computing in Europe (PRACE) og European Collaborative Data Infrastructure (EUDAT).

To maximize the impact and return of scientific research
by providing a predictable, sustainable and cost-efficient e-infrastructure

Provide advanced compute and
data services

Offer Sigma2 infrastructure for
other research related services

Integrator for cloud based and
international research services
and common components

- Procure, operate and develop a critical national e-infrastructure
- Compute services
- Data services (storage & data-management)
- Advanced User Support
- Training
- Support services

- Offer storage for other research data services
- Offer compute resources and storage for research-near public sector (Marine, Met, Health, ...)

- Provide researchers with an easy access to other cloud based services
- Integration of Service Platform with other cloud services, both commercial and sector internal, to enable a hybrid architecture.
- Participation in international e-infrastructures (e.g. NeIC, PRACE, EUDAT, EOSC, EuroHPC)

Innhold

- Bakgrunnsinformasjon for teknologi, omgivelser og lignende.
- Strategiske tema
- Detaljer på strategiske tema

Former High level objectives

- Procure, operate and develop a critical national infrastructure
- Promote e-infrastructure to new research communities
- Lead and coordinate participation in international cooperation for e-infrastructure
- Provide an attractive and sustainable e-infrastructure for all research communities, with the following characteristics:
 - High reliability and availability
 - Cost effectiveness
 - Predictable access
 - Interoperability within the national e-infrastructure (Notur/NorStore) and between national and international infrastructures (e.g. PRACE, EUDAT)
- Provide services for data analytics of large datasets (Big Data)

Omgivelsene og markedet har endret seg

- Tjenesteorganet er blitt opprettet og har fått ansvar for områder som også til dels dekker Sigma2
- HPC, lagring og analyse smelter i større grad sammen
- Sky som leveransemodell og European Open Science Cloud
- Større behov for forskningssinfrastruktur; (Klima, astro, biologi/medisin og helse, fluiddynamikk)
- Åpne data for gjenbruk
- Internasjonalt og nordisk samarbeid
- Nye kundegrupper og interessenter
 - Større ambisjoner for helserelatert forskning på sensitive data (dataen er eller blir sensitive)
 - Nye forskningsmiljøer jobber med big data gjennom nye analysemetoder (f.eks. samfunnsvitenskaplige fag)
 - Forvaltningsoppgaver (f.eks. Næring- og fiskeridep)

Teknologiutvikling

- **Big data / Analytics** alias "the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes"
 - Already piloteted framework for data-analytics. Planned as part of the Service plattform
- **Datalake** alias "storing data within a system or repository, in its natural format"
 - Relevant if Sigma2 changes the archiving policy towards data not connected with publications and possibly unstructured.
 - Next step in the development of the data-centric model. Should differentiate on cost, performance, traditional storage and object storage
- **Cloud technology** (self service, scalability, cooperation)
 - Very relevant. The Service Plattform already designed for this. Container technology is already tested on Abel through our PRACE commitments.
- **Remote access** - technologies to create a better user experience when resources are acced remotely or in the cloud
- **AI - Machine Learning (AI/ML)** alias gives computer systems the ability to "learn with data, without being explicitly programmed."
 - Will be relevant to consider HW and services for AI/ML
- **Quantum Computing**
 - Not considered mature for production environment for 3-5 years

Sektoren må ha tilgjengelig et tjenestetilbud som dekker ulike typer prosjekter

- Kjernekundene til Sigma2 er brukere med behov for kapasitetskrevenende og I/O-intensive lagrings- og beregningstjenester
- Kundegruppe
 - Forskere i UH-sektoren, institutter
 - Internasjonale prosjekt (PRACE) (in-kind ?)
 - Kommersiell forskning
 - Nordiske prosjekt (NeIC)?
- Sigma2 må ha i tankene at kundene betjenes av andre tjenesteleverandører for ulike behov (samarbeide med andre tjenesteleverandører)
- Samordning gjennom KDTO

Sigma2 ønsker å bidra til å realisere et helhetlig tjenestetilbud innen e-infrastruktur for sektoren

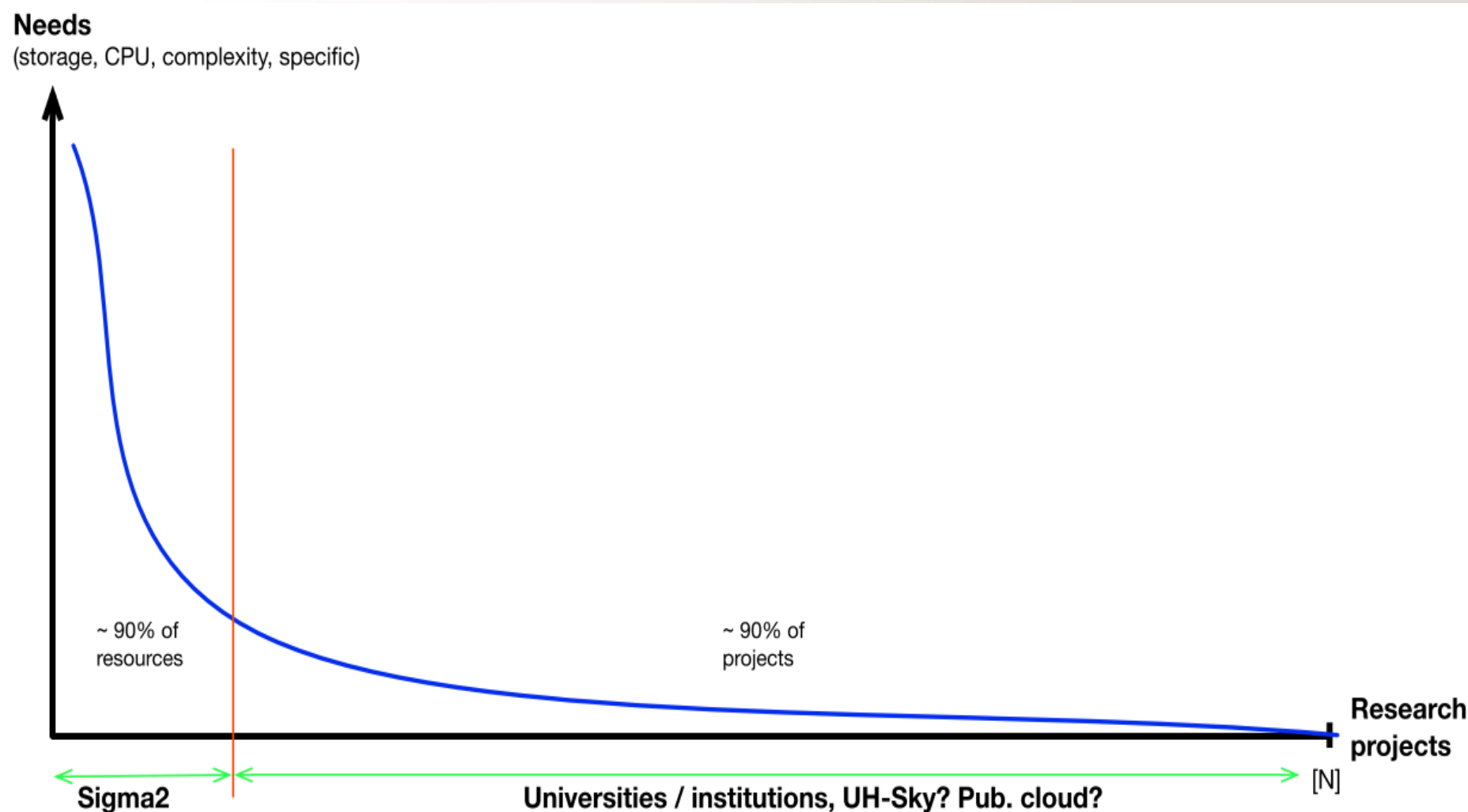
3 strategiske tema for Sigma2:

Videreutvikle dagens tjenestetilbud

Tilby infrastruktur for andre tjenester

Tilby integrasjoner mot skytjenester og felleskomponenter/-tjenester i sektoren

Skisse potensielt målbilde samlet tjenestetilbud



Sigma2s tjenester er i dag, og bør i fremtiden være, fokusert mot de aller mest krevende forskningsprosjektene

Samtidig bør det samlede tjenestetilbudet innen beregning og lagring være harmonisert slik at hverdagen for forskerne blir enklest mulig

Sigma2 ønsker derfor å både stille sin infrastruktur til rådighet for andre tjenester i sektoren, og være en muliggjører for effektiv og trygg bruk av kommersielle skytjenester som supplement.

Strategisk tema 1: Videreutvikle dagens tjenestetilbud

- Nåværende tjenestetilbud skal møte eksisterende behov og videreutvikles til å møte nye behov hos eksisterende brukere og nye brukergrupper
- Nåværende og fremtidig tjenestetilbud:
 - Beregningstjenster
 - Datatjenester
 - Avansert brukerstøtte
 - Opplæring
 - Støttetjenester

1: Videreutvikle dagens tjenester

Beregningstjenester

Status

- Vi står midt i steget mellom to organisasjonsmodeller - den gamle 4-site-modellen med fire maskiner og 4 drifts- og brukerstøttetjenester og 2-maskinmodellen med felles driftsorganisasjon, brukerstøtte og applikasjonsforvaltning
- Dagens beregningstjenester leveres i hovedsak av ressurser som er varianter av samme type (klynge med interconnect, kommandolinje, x86)
- Sigma2 e-infrastruktur er i liten grad brukt av UH-sektoren utenom BOTT

Prioriterte utviklingsområder

- Konsolidering på ny HW-plattform (A- og B-seriene og tjenesteplattform) med en operasjonell driftsorganisasjon og applikasjonsforvaltning aller viktigst
- Tilpasning til bio-last (kommer ifm. At Abel og Stallo fases ut)
- Styring av lastfordeling basert på kostnadseffektivitet
- Tilby containerteknologi på HPC, for eksempel HTCondor - utvikling koordinert med Tjenesteplattform. Dette muliggjør også avlastning av jobber mot skyplattformer.
- Følge opp UH utenom BOTT

1: Videreutvikle dagens tjenester

Lagring og datahåndtering

Status

- Arkitekturen er lagt om til en datasentrisk arkitektur som legger til rette for en utvidet tjenesteportefølje
- Dagens tjenesteportefølje:
 - NIRD project storage
 - Archive
 - NIRD Service Plattform
 - Data Management Planning (DMP) tool - easyDMP
- TSD - administrative and allocation procedures to be implemented

Prioriterte utviklingsområder

- Data Management Planning
 - Videreutvikling av easyDMP og integrasjonen med andre norske tjenester og strategisk markedsføring. Videre arbeid i regi av EOSC-hub og i samarbeid med OpenAIRE.
 - Etablering av nasjonalt forum for datarøkting
- Legge til rette for gjenbruk av åpne data, f.eks. gjennom videreutvikling av mekanismer til å tillate hosting av ofte brukte datasett. Mulig "user-scenarios":
- Deling av data internt og på tvers av miljøer
- Legge til rette for ny verktøy på en effektiv måte gjennom videreutvikling av NIRD Tjeneste Plattform:
 - || Prosjektspesifikke portaltjenester
 - High Throughput Computing (HTC)
 - Visualisering
- Videreutvikle ARKIV
 - skal følge FAIR-prinsippene
 - Arkitektur og integrasjoner
- Avklare behov for sertifisering (ISO27001, Data seal of approval)

1: Videreutvikle dagens tjenester

Opplæring

Status

- En rekke ulike aktører bidrar med variert spekter at opplæring: Tilbudet er bra, men fragmentert.
- 2016-2017 preget av intern opplæring ifbm. innfasing av ny infrastruktur. Parallelt, stort fokus på Pooling Competencies i NelC-regi.
- 2018: Incentiver for **intensivert satsing** på kurs og opplæring mot eksterne brukere:
 - NIRD og FRAM er satt i produksjon
 - Innfasing av tjenester på Sigma2 Service Platform i fremdrift.
 - Applikasjonsforvaltning får fart på sitt arbeid
 - Mottatt Evalueringskriteriene
 - Videre bidrag på nordisk nivå krever nasjonal portefølje
- Tilbakemeldinger i User survey demonstrerer et stort og variert behov for kursing og opplæring

Prioriterte utviklingsområder

- Etablere en **nasjonal strategi for opplæring**, inkludert etablering av en *helhetlig* portefølje av kurs, som dekker brukernes behov.
- Kursporteføljen skal bestå av:
 - Introduksjonskurs
 - Applikasjonsspesifikke kurs
 - Tjenesteorienterte kurs
- Virkemidler:
 - Etablere samarbeid med Software Carpentry og Code Refinery
 - Videreutvikle nasjonal training gruppe, skape møteplasser.
 - Utvikling av generisk og avansert training materiale for bruk i Metasenteret.
 - arrangere spesialiserte brukerkurs
- Muliggjøre bruk av digitale undervisningsformer
- Når:
 - 2018: Ferdigstille generisk HPC- Curriculum
 - 2019: Ferdigstille avansert HPC- Curriculum

1: Videreutvikle dagens tjenester

Formidling

Status

➤ <<Kommer når mer av strategien er klar>>

Prioriterte utviklingsområder

1: Videreutvikle dagens tjenester

Avansert brukerstøtte (AUS)

Status

- Økning i antall AUS-prosjekter etter en stillere periode
- De aller fleste AUS-prosjekter har godt utbytte
- Etablert nytt regime med kontinuerlig utlysning

Prioriterte utviklingsområder

- Vurdere nytt AUS regime
 - Ikke rigide kategorier
 - Tett involvering Applikasjonsforvaltning
 - Ressursfordelingskomiteens rolle - kan AUS-søknader behandles av metasentrene?
- Innføre bruker-liaisons, tilknytte oss spesielle kompetanseressurser i brukermiljøene
- Prioriterte satsinger for å utvikle tjenester som møter avanserte brukerbehov

1: Videreutvikle dagens tjenester

Støttetjenester

Status

Intro støttetjenester (Andreas), KPIer osv.

Dagens støttetjenester:

- Søknadsmotor for HPC, lagring og brukerkonto
- Selvbetjening for passordendring / reset
- Automatisert, selvstyrt kobling til regne og lagringsressurser (metadoc)
- Automatisert innsamling av accounting fra HPC / lagring (SGAS)
- Tjenesteovervåkning av Sigma2s anlegg (Zabbix)

Prioriterte utviklingsområder

- Støttetjenester for tjenesteplattformen
- Tettere integrasjon mot databaser med god datakvalitet (NFR, CRISTin)
- Tettere integrasjon mot ressursene vi betjener, eksempelvis levere informasjon gjennom LDAP til Fram/NIRD
- Adapter/API mot skytjenester (tildeling/accounting)?
- Integrasjon mot profilering av HPC-prosjekter

1: Videreutvikle dagens tjenester

Arbeidsprosesser og organisasjonsutvikling

Status

- Ikke etablert porteføljestyring
- Prosjekt-rammeverk fra UNINETT-konsern
- (Programmer)
- Tjenesteutvikling iht UNINETT Tjenestelivsyklus
- Sigma2 har hatt alt fokus rettet mot å få kjøpt inn og etablert nye og fremtidsrettede regneløsninger, og har i begrenset grad brukt ressurser på å få etablert en strategi, herunder kompetansestrategi. Dette må fremover få tillagt større ressurser.
- På samme måte er det i begrenset grad etablert metodikk for kvalitetssikring av drift og tjenesteleveranser. Dette har (**dog ikke?**) gitt utslag av svake leveranser eller driftsproblemer.

Prioriterte utviklingsområder

- Etablert porteføljestyring for prosjekter?
- Kompetansestrategi som del av strategiprosess.
- Driftsprosesser i henhold til anerkjente rammeverk f.eks. FITS?
- Hele metasenter må inkluderes i utviklingstekning

1: Videreutvikle dagens tjenester

Utvikling av samspillet med kundene

Status

- ⑩ Kommunikasjonsstrategi og interessentanalyse for selskapet utarbeidet i 2016. Oppdateres jevnlig.
- ⑩ Innføring av bidragsmodell skaper behov for å definere kunderolle på nytt. Skal betalende sluttbrukere prioriteres høyere?
- ⑩ Forvaltning av interessentanalysen for svak: mangler systematisk og strukturert oppfølging og etterlevelse av prioriterte mål og kundegrupper.
- ⑩ Viktig å nå de miljøene som potensielt har behov for Sigma2, men som ikke bruker oss i dag.
- Utfordring å få universitetene til å definere egen kunderolle. Må adresseres!

Prioriterte utviklingsområder

- Opprette arena for dialog mellom Sigma2 og universitetene hvor relevante personer fra universitetenes ledelse deltar (2018)
- Stimulere til at det etableres kunderepresentant hos alle universiteter
- Bedre forankring av strategi og samspill mellom lokale aktiviteter og Sigma2s aktiviteter.
- ⑩ Utarbeide plan for systematisk og strukturert oppfølging av kundegruppene (ref. Identifiserte kunder i interessentanalysen)
- ⑩ Systematisere rapporteringen av aktiviteten (inngår som KPI i Evalueringskriteriene)

1: Videreutvikle dagens tjenester

Utvikling av brukerhåndtering

Status

- Bruker = sluttbruker som mottar support
- For svak brukerorientering - kommunikasjonsstrategi mot alle forskere?
Får noen mer oppmerksomhet enn andre?
- Viktig å nå de miljøene som potensielt har behov for Sigma 2 men som ikke bruker oss i dag.
- Ikke systematisk arbeid med forskningsinfrastrukturer
- Generelt er det behov for å definere og måle verdiskapning for brukerne

Prioriterte utviklingsområder

- Kartlegge og analysere alle nåværende brukere fra ulike disipliner
- Identifisere potensielle nye brukergrupper, spesifisere hvem de er og hvordan tilnærme oss disse (konkretiseres i kommunikasjonsstrategien)
- Bruke Metasenteret for å nå nye brukergrupper (Iboende utfordring lokalt vs felles)
- Måle verdiskapning fra forskning: Publikasjoner på nivå 1 og 2 ?

1: Videreutvikle dagens tjenester

Kunder utenfor sektoren

Status

- Få kunder og liten aktivitet

Prioriterte utviklingsområder

- Kartlegge og analysere de mest aktuelle (best tilpassede) kundegruppene
- Produktspesifikasjon (hva kan vi tilby, og til hvilke vilkår)
- Markedsføring

Strategisk tema 2: Tilby infrastruktur for andre tjenester i sektoren

- Det har vært etterspørsel etter ren infrastrukturtjeneste på lagring, der Sigma2 leverer lagringsplass som andre kan bygge en tjeneste rundt. F.eks:
 - UiT sin interne tjeneste for lagring av forskningsdata (DataVerse) - UiO kan også tenke seg å bruke denne tjenesten
 - BIBSYS/BIRD
 - NSD - Objectstorage
 - NTNU - Generell lagring
- Sigma2 infrastruktur skal kun tilbys om det er mer kostnadseffektivt sammenlignet med det andre leverandører kan tilby

2: Tilby infrastruktur

Lagring

Status

- As of today, the allocation of the NIRD storage goes mainly through RFK
- Projects are evaluated according to scientific merit and usage pattern
- Use of storage as backend for other services is not allowed
- As of today, we do not offer storage for services not attached to a specific research project.
- No "low-cost" storage at the moment in NIRD

Prioriterte utviklingsområder

- Evaluate the offer of "low-cost" storage : what user scenario might be solved? Would need self service solution.
- Evaluate mechanisms for micro [Portals] and macro allocation (access granted and managed by communities, local institutions, ect).
- Evaluate principles to deliver storage to service providers for research (policies, LSA, cost model, FAIRness etc..)
- Explore possibilities for stronger integration/coordination with other data services in order to maximize/optimize the offer and user experience at national level

2: Tilby infrastruktur

HPC

Status

- Tilbyr HPC ressurser for CERN grunnforskning (WLCG)
- Differensierer ikke forvaltningsoppgaver fra forskning (for eksempel Havforskningsinstituttet)

Prioriterte utviklingsområder

- Ny og effektiv infrastruktur for WLCG som står godt sammen med øvrige CPU-ressurser (mulighet for utnyttelse av ledig kapasitet)
- Kartlegging (av eksisterende og mulig) bruk ifm. forvaltningsoppgaver (mulig inntektskilde fra andre dep.)

Strategisk tema 3: Leverer integrasjon mot skytjenester

- Sky-tjenester/plattformer kan være gode alternativ for forskere under visse forutsetninger
 - Trenger ikke flytte store datamengder
 - Trenger ikke raskt nettverk/høy parallellitet
 - Trenger ikke «spesialisert» HW
- Stor fordel om det kan være smidig/transparent overgang for forskere til bruk av skytjenester
- Sky betalingsmodell er velegnet ved stor variasjon i behov, og plutselige behov
- OBS: kostbar nedlastning om det blir svært store datamengder

3: Levere integrasjon mot skytjenester

Tjenesteplattform

Status

- Legacy services to be deployed on the NIRD Service Platform in March 2018
- AppStore (self service and deployment portal):
 - Owncloud (sharing and sync of data - «DropBox»)
 - Spark (data analytics)
 - Jupyter Notebook (collaboration tool)
 - R Studio/Shiny (statistical analysis)

Prioriterte utviklingsområder

- Integration of Service Platform with other cloud services, both commercial and sector internal, to enable a hybrid architecture:
 - Procure cloud resources
 - Increasing computing resources and add scalability
 - Offer new analytics services from 2nd and 3rd parties
 - Facilitate the sharing of software, /workflows
- Host for UNINETT services?

3: Leverer integrasjon mot skytjenester

HPC

Status

- Er gjort lite på området så langt, men Norge har kommet langt på skyintegrasjon i PRACE-sammenheng (container-teknologi)
 - Kostundersøkelser
 - WLCG
 - Foreslått oppstart ifm. B1-anskaffelse

Prioriterte utviklingsområder

- Identifisere egnede beregninger på skyplattform(er)
- Videreutvikle løsninger/tjenester basert på containers
- Sammenheng/samspill Tjenesteplattform og HPC
- Mulig pilot, for eksempel AWS (gjennom Géant avtale) for egnet miljø/miljøer (WLCG?).

4: Metasenteret/univ. sin rolle

Tjenesteproduksjonsmodell

Status

- Dagens tjenesteproduksjonsmodell involverer kun deler av IT-organisasjonene i BOTT og får i liten grad utnyttet spesialistene. (Applikasjonsforvaltning)
- Gjennom å spille på ressursene til BOTT kan man finne ressurser til utvikling. (TSD)

Prioriterte utviklingsområder

- Bruke andre miljøer enn Sigma2s metasentre til basis drift og tjenesteproduksjon?
- Spille på ressurser hos BOTT til å utvikle og videreutvikle tjenester.
 - F.eks.: UiO ønsker å utvikle tjeneste/platform for AI, og at Sigma2 deretter tar over driftsansvaret for tjenesten

Begrepsapparat

- Tjenesteplattform
- Datasentrisk
- ...