

Deling av data

UNINETT Fagdag

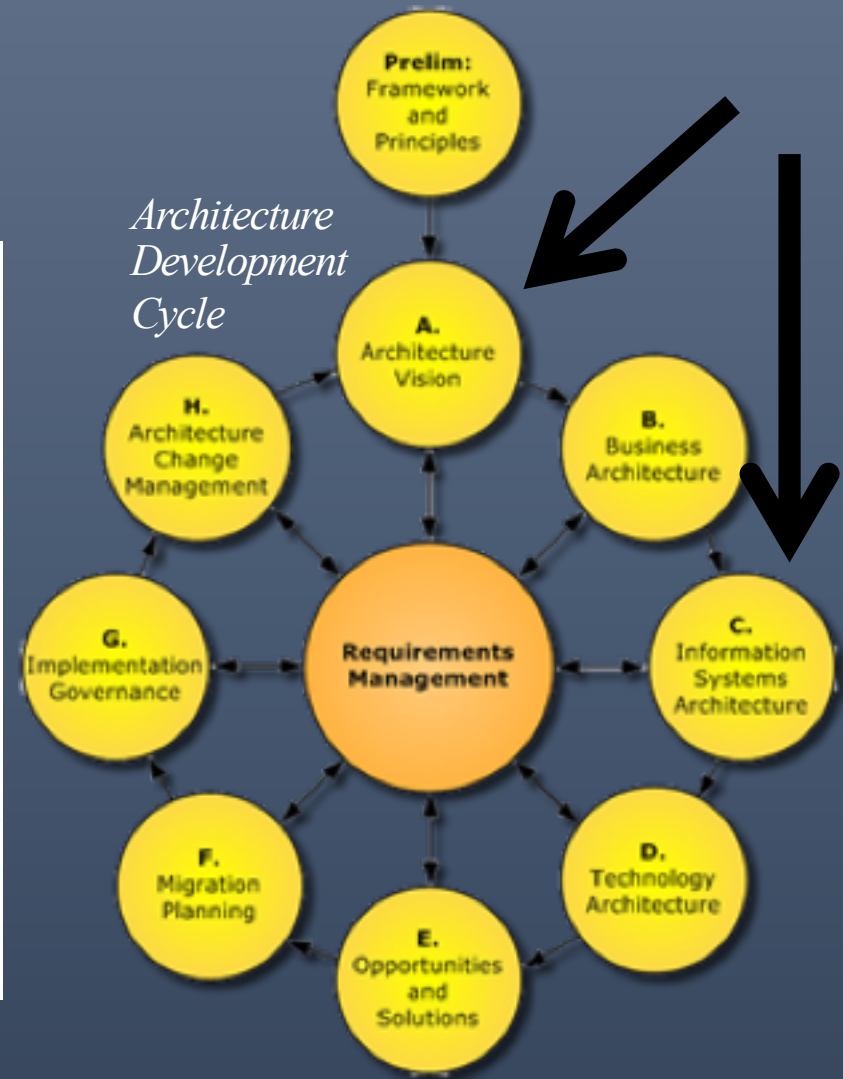
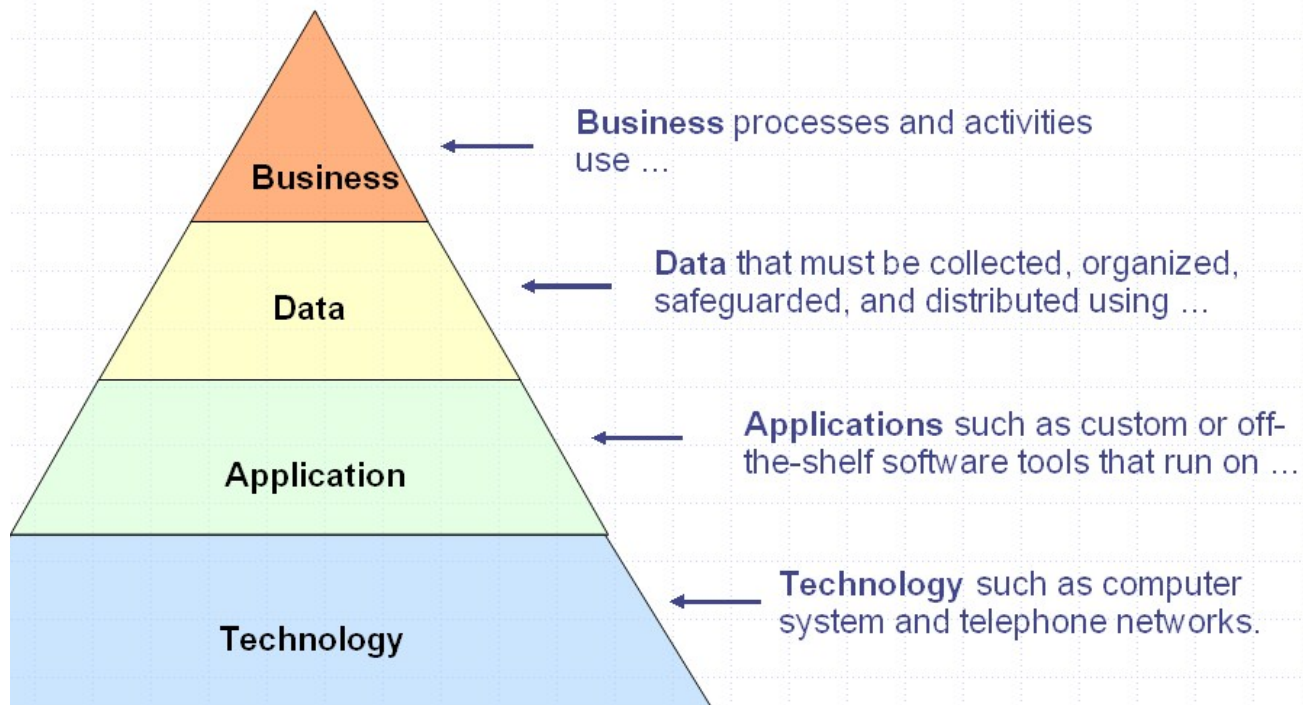
12. April 2018

Jan Erik Garshol

Hvorfor ønsker vi å dele data?

- Kun utføre registrering- og vedlikeholdsjobben en gang
 - Definert master og definerte slaver
- Slippe å lage synkroniseringsløsninger (som alltid feiler)
 - Sanntidsoppdrag eller ved behov duplisering av data (pull/push)
- Felles data gir felles oppfatning av virkeligheten – i ulike systemer
 - Informasjonsmodell som beskriver hvordan data samspiller
- Ulike løsninger men samme data
 - Ulike grensesnitt/applikasjoner for ulike brukergrupper – men samme behov

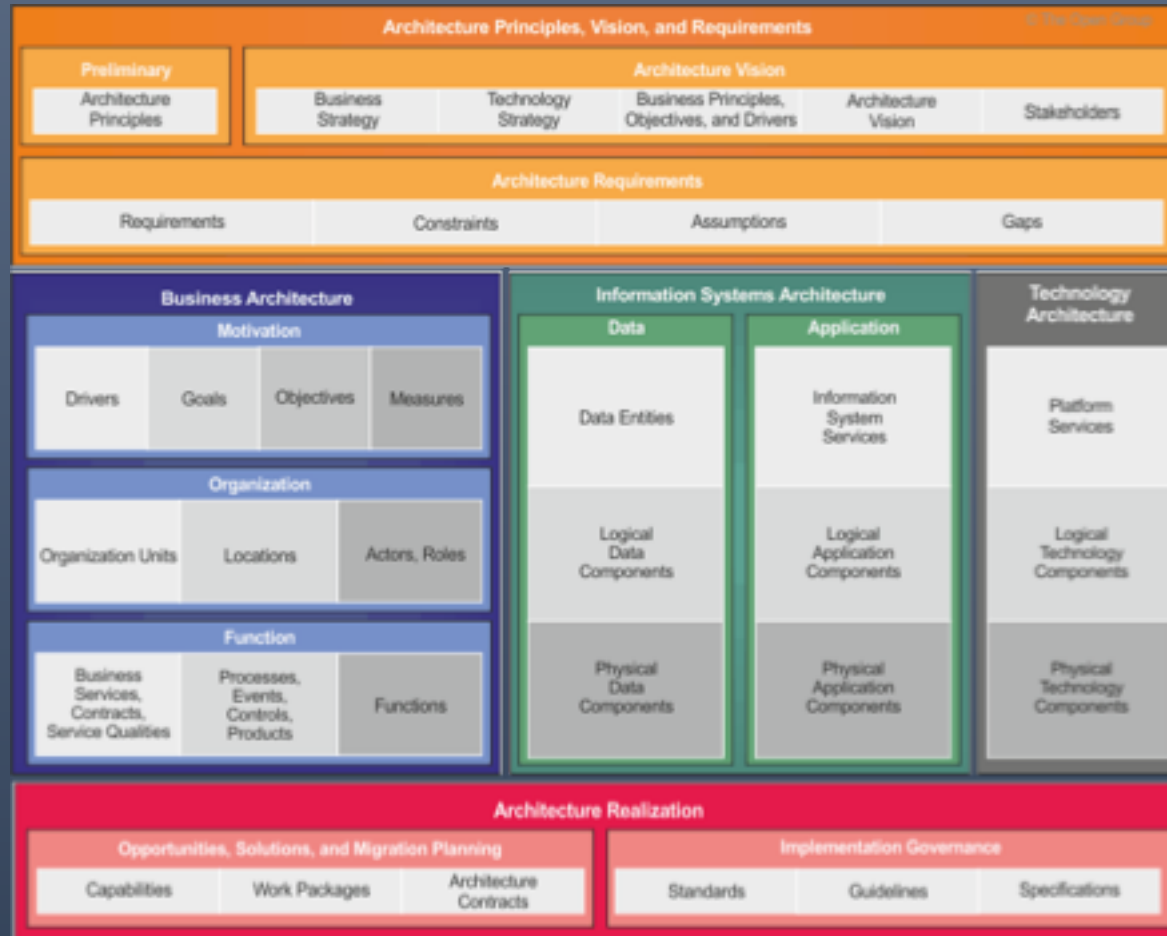
Behov -> data -> løsning



By Niles E Hewlett, PMP CEA, Enterprise Architecture Team, USDA-OCIO -
The USDA Enterprise Architecture Program, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=5458141>

<http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>

TOGAF 9 - Content Metamodel Overview



<http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap33.html>

Felles IKT-arkitekturprinsipper for universitets- og høgskolesektoren

- Forfattere:

- | | | |
|----------------------------|---------|--------|
| • Heidi Bergh-Hoff | UNINETT | → KDTO |
| • Carl-Fredrik Sørensen | NTNU | |
| • Jan Erik Garshol | BIBSYS | → KDTO |
| • Bård Henry Moum Jakobsen | UiO | |
| • Geir Magne Vangen | FSAT | → KDTO |
| • Ørjan Dypvik Pettersen | HiN | |
| • Johnny Hansen | UiT | |

- Ferdig 03.09.2015

- Tilgjengelig på <https://hdl.handle.net/11250/2473356>

Sektoren 7 arkitekturprinsipper – Om data

- Tjenesteorientering
 - Autoritativ kilde data er gjort tilgjengelig gjennom gjenbrukbare applikasjonstjenester
- Interoperabilitet
 - Den felles informasjonsarkitekturen realiseres i felles datamodell
 - Betydningen av data er klart definert i informasjonsarkitekturen
 - Alle data har én autoritativ kilde der all oppdatering skjer
- Tilgjengelighet
 - Det er lett å finne fram til og få aksess til data man har tilgang til
 - i rett prosess til rett tid, som oftest elektronisk, 24x7, globalt
 - Data har en utpekt dataeier og definerte rutiner for forvaltning
 - Krav til datakvalitet er spesifisert

- Sikkerhet

- Informasjon klassifiseres og tilgang til data styres for å ivareta det definerte sikkerhetsnivået
- Det legges til rette for å oppfylle regulatoriske krav for sammenstilling, oppbevaring og sikker avhending av data

- Åpenhet

- Tilgang til data er åpen som "default". Tilgangsbegrensninger innføres ved behov. Data er formatert i henhold til åpne standarder og tilgjengelig gjennom åpne grensesnitt

- Flexibilitet

- Autoritative datakilder må tilby nødvendig flexibilitet innenfor sin semantiske interoperabilitet for å muliggjøre ulik bruk der det er nødvendig, inklusiv å kun bruke deler av autoritative data

- Skalerbarhet

- Data hentes direkte fra autoritativ kilde såfremt det er praktisk mulig

Dette krever en felles informasjonsmodell

- Hvilke prosesser skal støttes?
- Hvor likt blir like prosesser realisert - i ulike organisasjoner?
- Hvilke system er master for hvilke data – og for hvem?
- Hvilke vokabular/terminologi skal brukes i hvilke systemer?
- Felles autoritetsregister – som faktisk blir brukt

- Kun oppdatere masterdata – aldri slave-data!

F

- **FINDABLE:**

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

A

- **ACCESSIBLE:**

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

I

- **INTEROPERABLE:**

- I1. (meta)data use a formal, accessible, shared, and broadly applicable lang. for knowledge represe.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

R

- **RE-USABLE:**

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

Nasjonalt Vitenarkiv

- Kunnskapsdepartementet har gitt KDTO i oppdrag å utrede hvordan et nasjonalt vitenarkiv skal realiseres.
- KDTO vil etablere et prosjekt, som har leveransefrist 1. oktober 2018
 - Mandat er under utarbeiding
- Det vil bli etablert en referansegruppe med personer nominert av alle forskningsutførende institusjoner (ikke bare UH-sektor)

Nasjonalt Vitenarkiv

- Bakgrunn for arbeidet er Brekke-utvalget, 16 juni 2016: «Nasjonale retningslinjer for åpen tilgang til forskningsresultater»
- Leveranse er en rapport til Kunnskapsdepartementet, som gir en plan for hvordan vitenarkiv kan realiseres
- Etablering av et Nasjonalt Vitenarkiv vil sannsynligvis erstatte Brage og Cristin vil bli berørt
 - Alle vitenskapelige artikler skal lagres i NVA, senest på publiseringstidspunkt.

Undervisningsdata

- «En rekke norske institusjoner har utviklet og også tilgjengeliggjort ressurskilder både for egenprodusert digitalt innhold og innhold som er hentet fra andre kilder.

En felles løsning for tilgang vil gjøre det mulig å forvalte læringsressurser sentralt for å stimulere til økt produksjon og deling av læringsressurser, samt gjøre åpne digitale læringsressurser for høyere utdanning tilgjengelige.»

Punkt 5.2.6 i «*Digitaliseringsstrategi for universitets- og høyskolesektoren 2017-2021*»

- Slike løsninger blir gjerne omtalt som «*Learning Objekt Repository*» (LOR), når innholdet er læringsressursers.

DLR – en realisering av et LOR

- Etablert i samarbeid med NTNU.
- Valgt av BI, etter en internasjonal RFI-prosess
- OsloMet og HVL m.f. ønsker å ta i bruk denne løsningen.
- Integrert i ulike LMS (Canvas, Blackboard Learn, Open edX, It's Learning)
- Integrerer ulike kilder av resurser (for eks. Mediasite) og gir bruker tilgang til eksterne data, i en felles løsning.

Forskningsdata

- Nasjonal strategi for tilgjengeliggjøring av forskningsdata definerer 3 grunnprinsipper:
 - Forskningsdata skal være så åpne som mulig, så lukkede som nødvendig
 - Forskningsdata bør håndteres og tilrettelegges slik at verdiene i dataene kan utnyttes best mulig
 - Beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene
- Det finner i dag flere ulike nasjonale infrastrukturer som tilbyr lagring og publisering av forskningsdata:
 - NIRD (Sigma2), NORDI (NSD), Bird (KDTO), NMDC, DataverseNO (UiT)

Bibliografiske data

- BIBSYS Konsortiet har alle bibliografiske data (innholdsfortegnelsen til «alle» vitenskapelige bibliotek i Norge) i Alma
- Oria tilbyr gjenfinning av alle publisert resurser
 - Global indeks, aggregert av ulike internasjonale aktører
 - Nasjonale kilder blir lagt til for å komplementere
- Felles Autoritetsregister er et generisk system som søker å tilby mange ulike masterdata til mange ulike system
 - Person(Forfatter), Institusjon, Emne, Verk, Standard tittel m.f.
 - Cristin-prosjekt?

Semantisk web

- Aggregerer ulike kilder og normerer metadata for å tilby en helhet
- Tilbyr søk på tvers av alle kilder og alle typer data
- Tilbyr mulighet til å navigere i en graf, etter søk
- Gir navigasjon tilbake til master kilde

- Forsøker å bli en inngangsport til en felles (bibliografisk) informasjonsmodell, som opprettholder peker til masterdata og gir internasjonal kontekst til nasjonale data
 - DOI/handle, ISNI/VIAF, ORCID m.f.