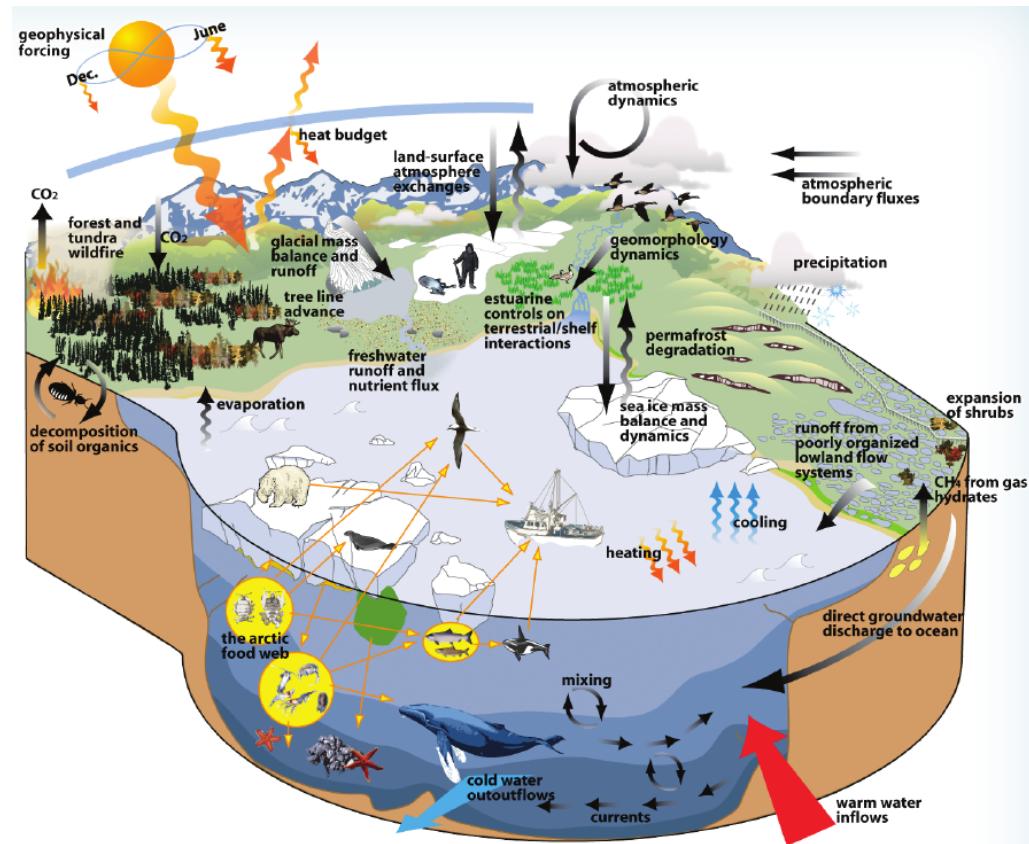


Dynamiske data, utfordringer, muligheter og sporing av beslutninger

Øystein Godøy, Trygve Halsne

Hva er dynamiske data?

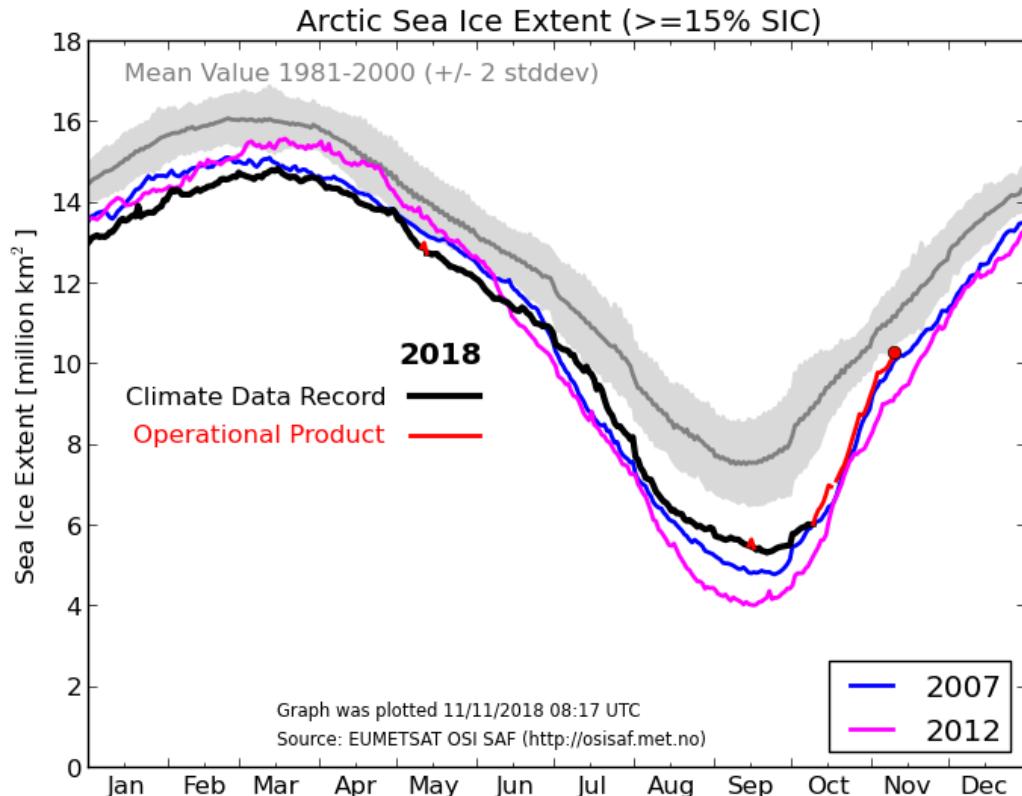
- Data som beskriver tidsavhengige prosesser
 - Data oppdateres kontinuerlig
- Typiske generisk datatyper
 - Tidsserier på faste eller bevegelige stasjoner
 - Tidsserier av profiler på faste eller bevegelige stasjoner
 - Trajektorier og trajektorier av profil
 - Tidsserier av griddede data
- Typiske kilder
 - In situ målinger
 - Fjernmåling
 - Simulerte
 - Analyserte



Roberts, A. and coauthors 2010. A Science Plan for Regional Arctic System Modeling, International Arctic Research Center Technical Papers 10-0001

Formål med dynamiske data

- Beskrive forhold i
 - fortid
 - sanntid
 - fremtid

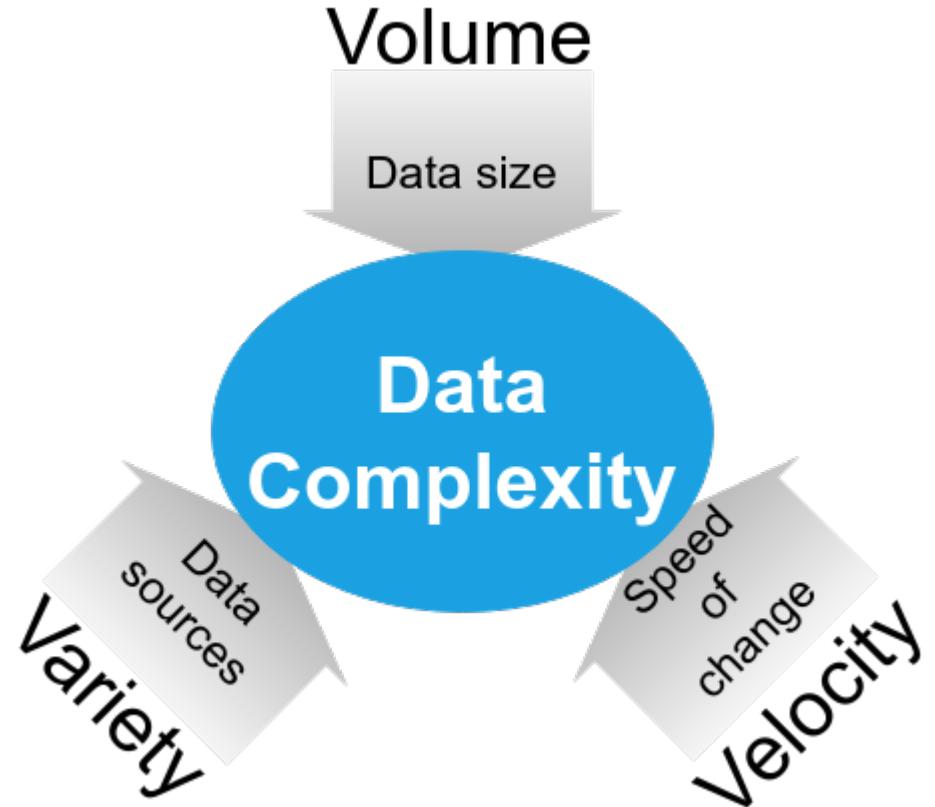


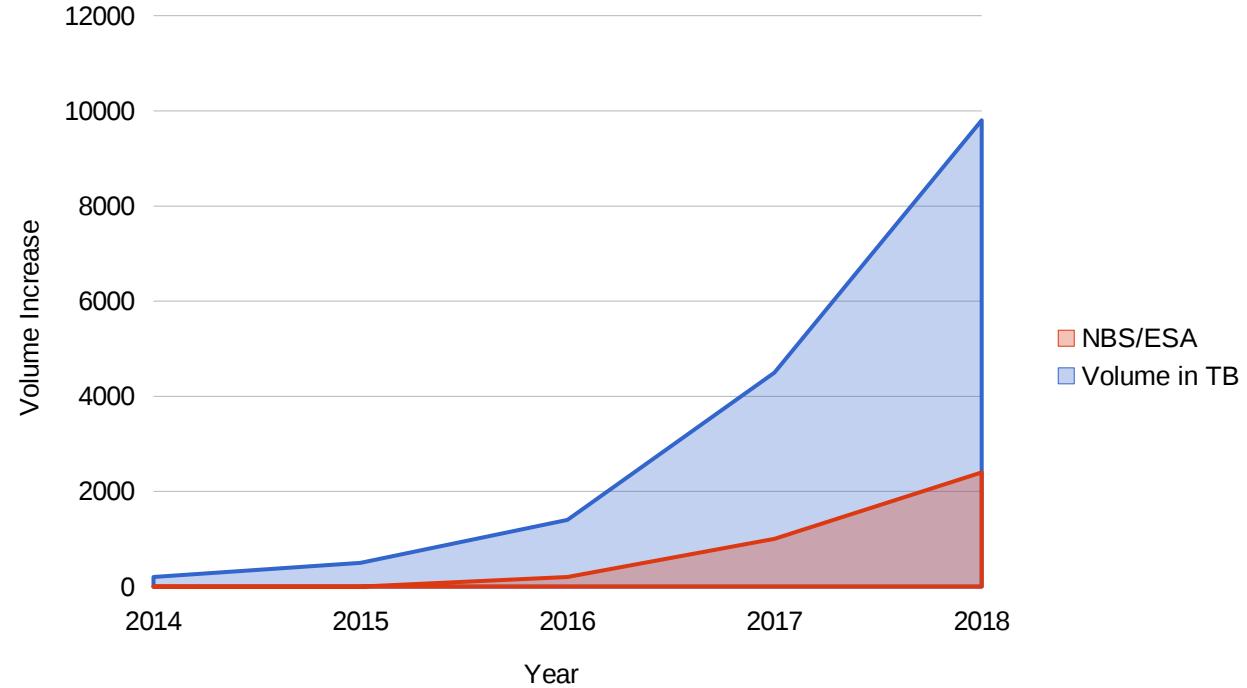
Om vi er heldig



Utfordringer med dynamiske data?

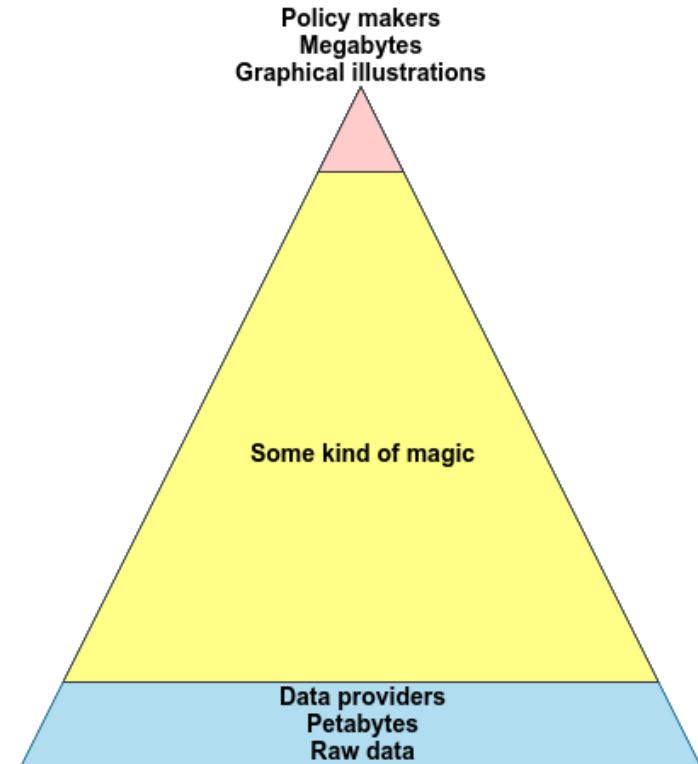
- Kontinuerlige datastrømmer
 - Innsamling og kvalitetssikring
- Flere dimensjoner
 - Noen data har flere tidslinjer
- Datavolum
 - Numeriske simuleringer
 - Fjernmålte data
 - Antall målepunkter
- Heterogene data
 - Ulike observasjonssystemer
- Omløpstid
 - Flere og flere systemer leverer minutdata
- Leveranser til ulike nasjonale og internasjonale nettverk
 - Kostnader





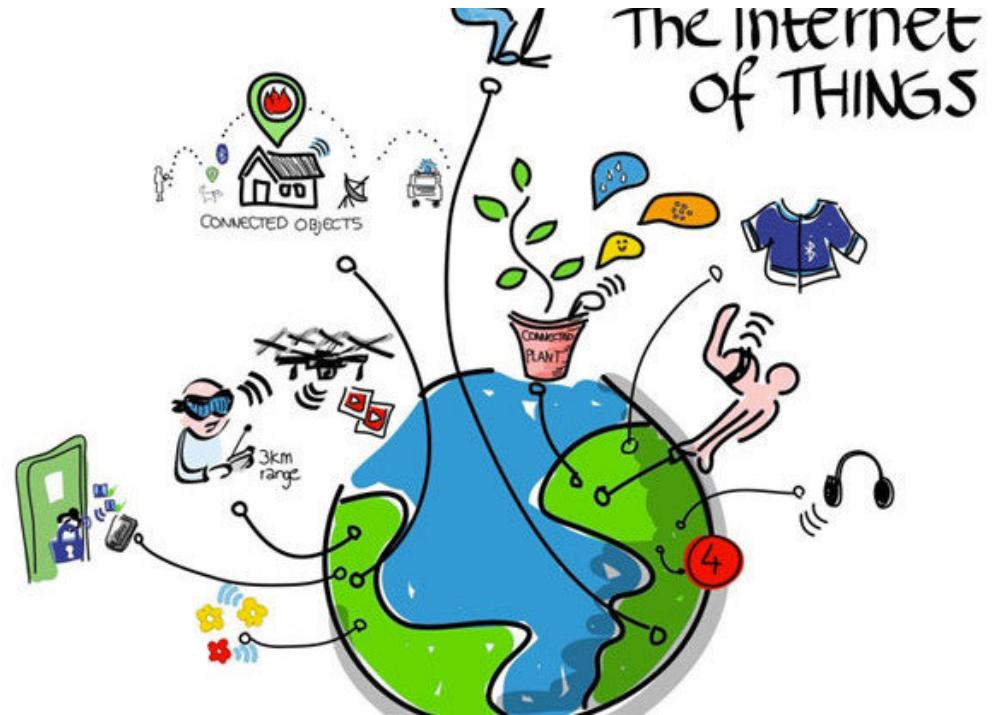
Utfordringer med økende datavolum

- Når volumet med tilgjengelig informasjon vokser
 - Hvordan sikres effektiv gjenfinning av relevant informasjon?
 - Hvordan sikres effektiv forståelse av informasjonen som er funnet?
 - Hvordan sikres effektiv anvendelse av informasjonen?
- Må se på alternative måter å jobbe med data



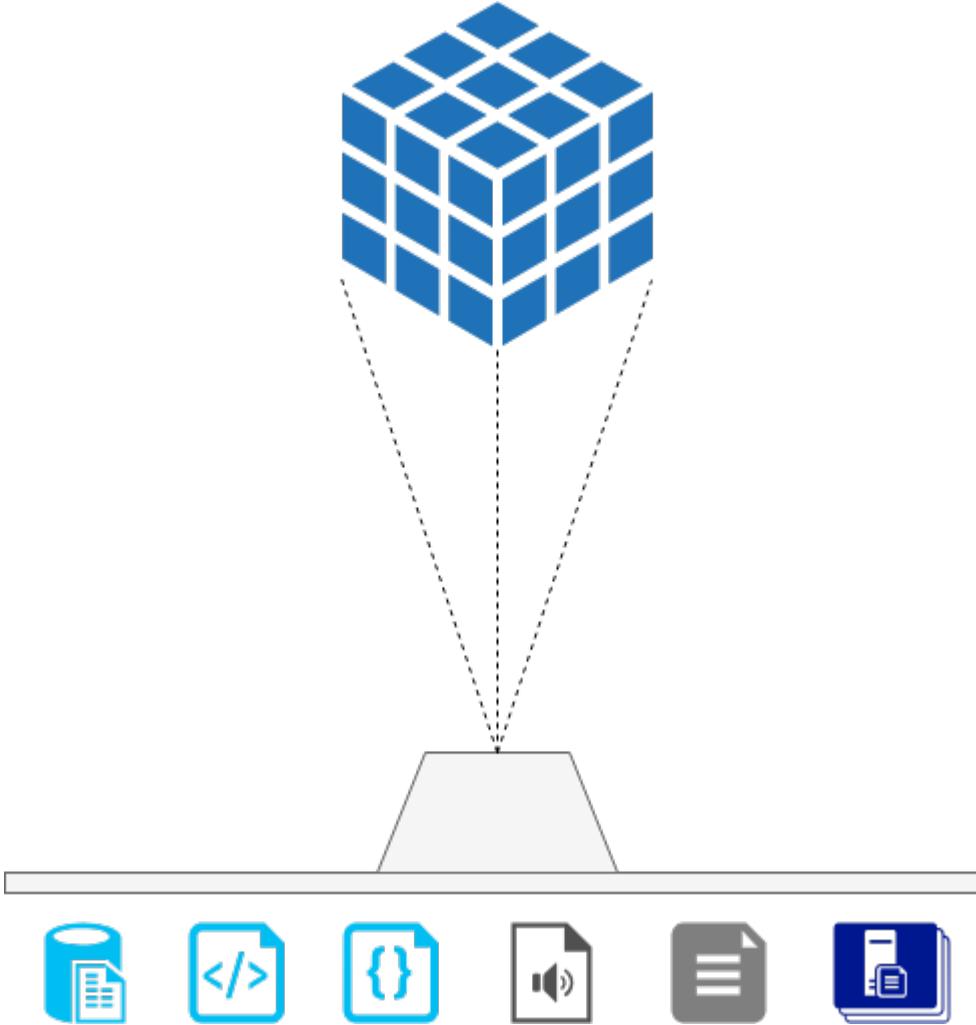
Saker som skal løse alle våre problemer...

- Big data
- Machine learning
- Cloud technology
- Containers
- «A new programming language»

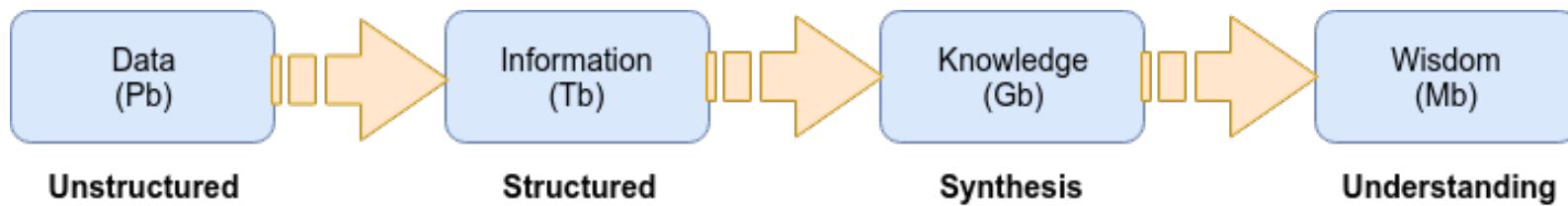


Beware the Internet of Things hype

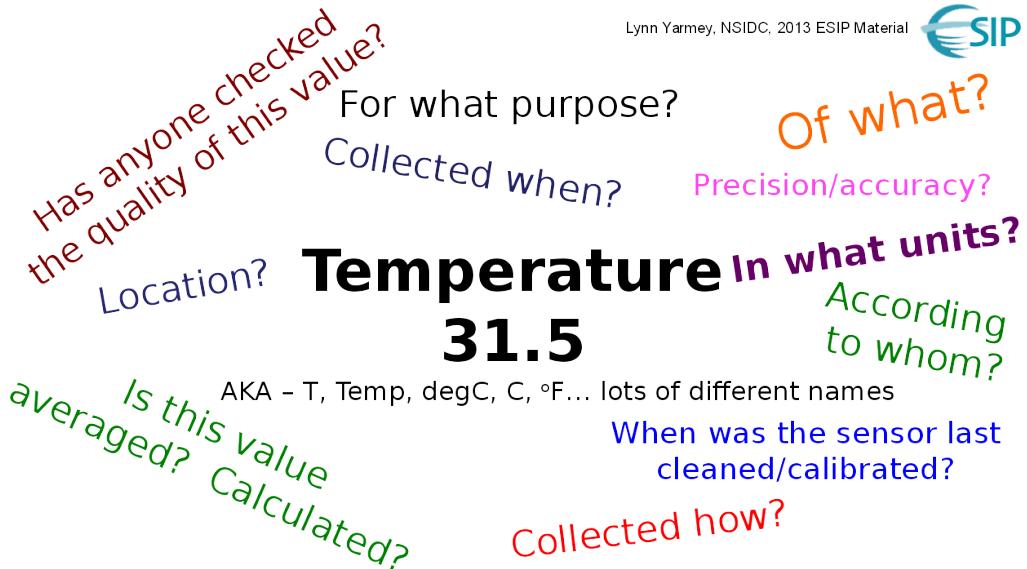
<https://www.computerworld.com/article/3116987/internet-of-things/beware-the-internet-of-things-hype.html>



Kunnskapsuthenting



Må sette data i en sammenheng



- Må få data til å «snakke sammen»
 - Standardisert dokumentasjon og strukturering
 - F.eks. Climate and Forecast convention
 - Sporbare data
 - PID
 - Versjonering
 - Både nasjonalt og internasjonalt
- Hva er et tall?
 - Bruksmetadata er påkrevd for effektiv anvendelse av dataene
 - Omfatter også beskrivelse av f.eks. observasjonsinnhenting
 - F.eks. WIGOS

Types of metadata

Type	Purpose	Description	Examples
Discovery	Used to find relevant data	Discovery metadata are also called index metadata and are a digital version of the library index card. It describes who did what, where and when, how to access data and potential constraints on the data.	ISO19115 GCMD DIF
Use	Used to understand data found	Use metadata are describing the actual content of a dataset and how it is encoded. The purpose is to enable the user to understand the data without any further communication. It describes content of variables using standardised vocabularies, units of variable, encoding of missing values, map projections etc.	Climate and Forecast Convention BUFR GRIB
Configuration	Used to tune portal services for datasets for users.	Configuration metadata are used to improve the services offered through a portal to the user community. This can be e.g. how to best visualise a product. This information is maintained by the GCW portal and is not covered by discovery or use metadata standards.	
Site	Used to understand data found	Site metadata are used to describe the context of observational data. It describes the location of an observation, the instrumentation, procedures etc. To a certain extent it overlaps with discovery metadata, but more so it really extends discovery metadata. Site metadata can be used for observation network design.	WIGOS OGC O&M



```
File Edit View Bookmarks Settings Help
netcdf radflux_bjornoya {
dimensions:
    time = UNLIMITED ; // (3847970 currently)
    strlen25 = 25 ;
variables:
    double time(time) ;
        time:long_name = "time of the observation" ;
        time:short_name = "time" ;
        time:standard_name = "time" ;
        time:units = "seconds since 1970-01-01 00:00:00 UTC" ;
        time:axis = "T" ;
    char stationid(strlen25) ;
        stationid:long_name = "name and/or stationnumber used as identifier" ;
    float latitude ;
        latitude:long_name = "latitude" ;
        latitude:short_name = "latitude" ;
        latitude:standard_name = "latitude" ;
        latitude:units = "degree_north" ;
        latitude:valid_min = -90.f ;
        latitude:valid_max = 90.f ;
    float longitude ;
        longitude:long_name = "longitude" ;
        longitude:short_name = "longitude" ;
        longitude:standard_name = "longitude" ;
        longitude:units = "degree_east" ;
        longitude:valid_min = -180.f ;
        longitude:valid_max = 180.f ;
    float ssi(time) ;
        ssi:long_name = "shortwave irradiation at the surface" ;
        ssi:short_name = "ssi" ;
        ssi:standard_name = "surface_downwelling_shortwave_flux" ;
        ssi:_FillValue = -999.f ;
        ssi:units = "watts/meter2" ;
        ssi:cell_method = "time: mean (last minute)" ;
    float ssisenstamp(time) ;
        ssisenstamp:long_name = "temperature of the surface shortwave irradiation sensor" ;
        ssisenstamp:short_name = "ssisenstamp" ;
        ssisenstamp:_FillValue = -999.f ;
        ssisenstamp:units = "degC" ;
        ssisenstamp:cell_method = "time: mean (last minute)" ;
    float dli(time) ;
        dli:long_name = "difference between downward atmospheric longwave irradiation and emitted CGR4 irradiance" ;
        dli:short_name = "dli" ;
        dli:standard_name = "surface_net_downward_longwave_flux" ;
        dli:_FillValue = -999.f ;
        dli:units = "watts/meter2" ;
        dli:cell_method = "time: mean (last minute)" ;
    float disenstamp(time) ;
        disenstamp:long_name = "temperature of the surface longwave irradiation sensor" ;
        disenstamp:short_name = "disenstamp" ;
        disenstamp:_FillValue = -999.f ;
        disenstamp:units = "degC" ;
        disenstamp:cell_method = "time: mean (last minute)" ;
    float battery(time) ;
        battery:long_name = "minimum battery voltage" ;
        battery:short_name = "battery" ;
        battery:_FillValue = -999.f ;
        battery:units = "V" ;
        battery:cell_method = "time: min (last minute)" ;
:
```

Bjørnøya : ncdump

```
File Edit View Bookmarks Settings Help
// global attributes:
    :Conventions = "CF-1.0" ;
    :history = "2008-10-23 creation\n",
    "2016-01-01 revision" ;
    :title = "Downwelling surface radiative fluxes at Bear Island" ;
    :abstract = "Downwelling surface radiative fluxes observed at the meteorological station at Bear Island in the Barents Sea. Measurements are made using Kipp and Zonen CMP21 and CGR4 pyranometers and pyrgeometers. Daily maintenance is performed by the meteorological personnel at the station. Data are averaged over the last minute and the time is set to UTC. This data set has been collected with support from the Norwegian Research Council. The quality control focuses on the radiative parameters, thus sensor temperatures may contain errors." ;
    :topiccategory = "Climatology/Meteorology/Atmosphere" ;
    :keywords = "Radiative Flux" ;
    :gcmd_keywords = "Atmosphere > Atmospheric Radiation > Shortwave Radiation\n",
        "Atmosphere > Atmospheric Radiation > Longwave Radiation" ;
    :area = "Barents Sea" ;
    :activity_type = "Land station" ;
    :PI_name = "<DØY>stein God<F8>y" ;
    :contact = "o.godoy@met.no" ;
    :institution = "Norwegian Meteorological Institute" ;
    :url = "http://www.met.no/" ;
    :product_name = "radiative fluxes" ;
    :Platform_name = "Bj<F8>rn<F8>ya" ;
    :project_name = "IAOOS-Norway/IPY-THORPEX" ;
    :start_date = "2008-04-01 13:14 UTC" ;
    :stop_date = "2015-12-16 12:50 UTC" ;
    :distribution_statement = "Restricted to IAOOS-Norway" ;
    :southernmost_latitude = 74.5166667 ;
    :northernmost_latitude = 74.5166667 ;
    :westernmost_longitude = 19.01666667 ;
    :easternmost_longitude = 19.01666667 ;
    :quality_statement = "Quality controlled" ;
    :nco_opemp_thread_number = 1 ;
data:
time = 1207055640, 1207055700, 1207055760, 1207055820, 1207055880,
1207055940, 1207056000, 1207056060, 1207056120, 1207056180, 1207056240,
1207056300, 1207056360, 1207056420, 1207056480, 1207056540, 1207056600,
1207056660, 1207056720, 1207056780, 1207056840, 1207056900, 1207056960,
1207057020, 1207057080, 1207057140, 1207057200, 1207057260, 1207057320,
1207057380, 1207057440, 1207057500, 1207057560, 1207057620, 1207057680,
1207057740, 1207057800, 1207057860, 1207057920, 1207057980, 1207058040,
1207058100, 1207058160, 1207058220, 1207058280, 1207058340, 1207058400,
1207058460, 1207058520, 1207058580, 1207058640, 1207058700, 1207058760,
1207058820, 1207058880, 1207058940, 1207059000, 1207059060, 1207059120,
1207059180, 1207059240, 1207059300, 1207059360, 1207059420, 1207059480,
1207059540, 1207059600, 1207059660, 1207059720, 1207059780, 1207059840,
1207059900, 1207059960, 1207060020, 1207060080, 1207060140, 1207060200,
1207060260, 1207060320, 1207060380, 1207060440, 1207060500, 1207060560,
1207060620, 1207060680, 1207060740, 1207060800, 1207060860, 1207060920,
1207060980, 1207061040, 1207061100, 1207061160, 1207061220, 1207061280,
1207061340, 1207061400, 1207061460, 1207061520, 1207061580, 1207061640,
1207061760, 1207061820, 1207061880, 1207061940, 1207062000,
1207062060, 1207062120, 1207062180, 1207062240, 1207062300, 1207062360,
1207062420, 1207062480, 1207062540, 1207062600, 1207062660, 1207062720,
1207062780, 1207062840, 1207062900, 1207062960, 1207063020, 1207063080,
1207063140, 1207063200, 1207063260, 1207063320, 1207063380, 1207063440,
1207063500, 1207063560, 1207063620, 1207063680, 1207063740, 1207063800,
1207063860, 1207063920, 1207063980, 1207064040, 1207064100, 1207064160,
1207064220, 1207064280, 1207064340, 1207064400, 1207064460, 1207064520,
```

Bjørnøya : nc dump

Bjørnøya : ncview — Konsole

File Edit View Bookmarks Settings Help

```
-rw-rw-r-- 1 steingod steingod 2,6M mai      1  2015 radflux_bjornoaya-201504.dat
-rw-rw-r-- 1 steingod steingod 1,2M mai      1  2015 radflux_bjornoaya-201504.nc
-rw-rw-r-- 1 steingod steingod 2,6M juni     1  2015 radflux_bjornoaya-201505.dat
-rw-rw-r-- 1 steingod steingod 1,2M juni     1  2015 radflux_bjornoaya-201505.nc
-rw-rw-r-- 1 steingod steingod 2,4M juli    1  2015 radflux_bjornoaya-201506.dat
-rw-rw-r-- 1 steingod steingod 1,1M juli    1  2015 radflux_bjornoaya-201506.nc
-rw-rw-r-- 1 steingod steingod 2,4M aug.   1  2015 radflux_bjornoaya-201507.dat
-rw-rw-r-- 1 steingod steingod 1,1M aug.   1  2015 radflux_bjornoaya-201507.nc
-rw-rw-r-- 1 steingod steingod 2,5M sep.   1  2015 radflux_bjornoaya-201508.dat
-rw-rw-r-- 1 steingod steingod 1,2M sep.   1  2015 radflux_bjornoaya-201508.nc
-rw-rw-r-- 1 steingod steingod 2,4M okt.   1  2015 radflux_bjornoaya-201509.dat
-rw-rw-r-- 1 steingod steingod 2,2M nov.   1  2015 radflux_bjornoaya-201510.dat
-rw-rw-r-- 1 steingod steingod 2,6M des.  1  2015 radflux_bjornoaya-201511.dat
-rw-rw-r-- 1 steingod steingod 1,2M des.  1  2015 radflux_bjornoaya-201511.nc
-rw-rw-r-- 1 steingod steingod 1,4M jan.   1  2016 radflux_bjornoaya-201512.dat
-rw-rw-r-- 1 steingod steingod 616K jan.   1  2016 radflux_bjornoaya-201512.nc
-rw-rw-r-- 1 steingod steingod 15M aug.  16  2016 radflux_bjornoaya-2016.cdl
-rw-rw-r-- 1 steingod steingod 21M aug.  16  2016 radflux_bjornoaya-2016.dat
-rw-rw-r-- 1 steingod steingod 9,4M aug. 16  2016 radflux_bjornoaya-2016.nc
-rw-rw-r-- 1 steingod steingod 103M okt. 14  2016 radflux_bjornoaya.nc
-rw-rw-r-- 1 steingod steingod 73M aug.  29  2013 radflux_Bjørnøya.nc
-rw-rw-r-- 1 steingod steingod 32M aug.  29  2013 radflux_Bjørnøya.tgz
-rwxr-xr-x 1 steingod steingod 322 juni   6   2008 radobs_collection_status.txt*
drwxr-xr-x 2 steingod steingod 4,0K juni  25  2013 tmp/
drwxr-xr-x 2 steingod steingod 4,0K juni  25  2013 tmp/
steingod@tuba:/disk1/data/radflux/Bjørnøya$ ncdump radflux_bjornoaya.nc | m
steingod@tuba:/disk1/data/radflux/Bjørnøya$ ncview radflux_bjornoaya.nc
Ncview 2.1.6 David W. Pierce 29 Oct 2015
http://meteora.ucsd.edu:80/~pierce/ncview_home_page.html
Copyright (C) 1993 through 2015, David W. Pierce
Ncview comes with ABSOLUTELY NO WARRANTY; for details type `ncview -w'.
This is free software licensed under the Gnu General Public License version 3; t
ype `ncview -c' for redistribution details.

Warning: Cannot convert string "-*-helvetica-*~r-*~-14-*~-*~-*~-*" to type F
ontStruct
Note: 43208 missing values were eliminated along axis "time"; index= 1 2 3 4 5
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 4294967298..
```

No variable selected

Review 3.1.6 David L. Rizzo - 29 Oct 2015

*** SELECT A WORKTABLE TO START ***

Current: x=20-May-2008 10:10:08, u=904,478

Quit **->1** **<<** **<** **||** **>** **>>** **Edit** **?** **Delay:** **█** **Opts**

3gauss **Inv P** **Inv C** **Mag X1** **Linear** **Axes** **Range** **Bi-lin** **Print**

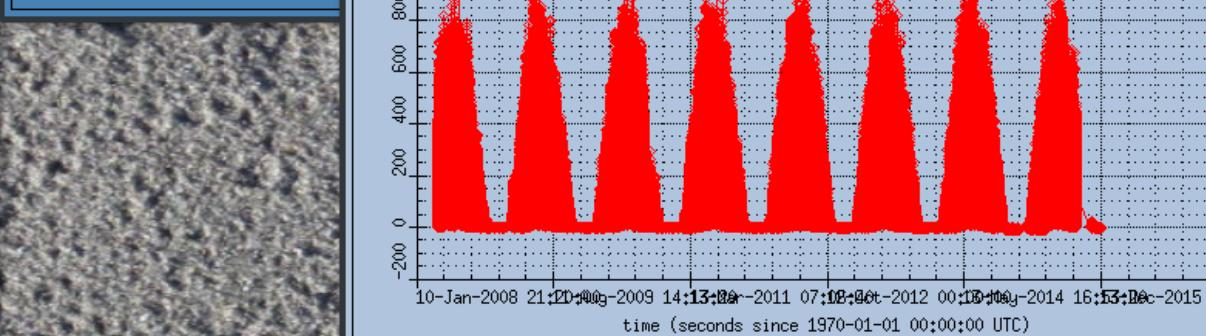
1

Var: stationid

dissenstepp

Din: Name: Mi:

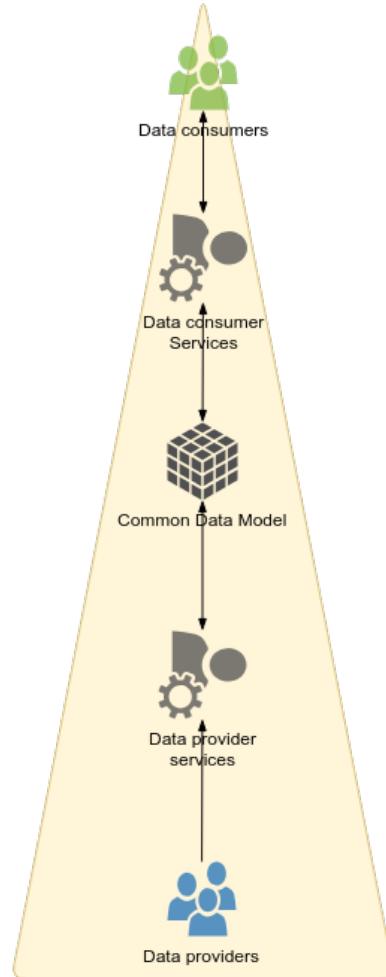
strlen25



shortwave irradiation at the surface from Downwelling surface radiative fluxes at Bear Island

X Axis: time Use Log: X Y X Range Y Range

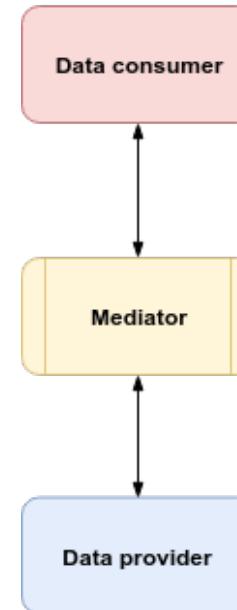
Må strukturere data på tvers av systemer



- Data ligger i siloer
- Må få siloer til å snakke sammen
- Har ikke råd til å bygge alle systemer på nytt
 - Kostnadseffektiv basisteknologi
 - Robust
 - Egnet for
 - automatisering
 - håndtering av ulike typer data
 - skalering ifht volum

Bygge broer

- Gjenbruk av løsninger
 - Nasjonalt og internasjonalt
 - Forvaltning og forskning
 - På tvers av disipliner
- Unngå duplisering av produkter og tilgangsløsninger
 - Overføringer av informasjon på forespørsel
- Selvforskjende data
 - En forutsetning for maskinell behandling



Unwilling

- Do not want to change behaviour, existing tools have worked well.
- Want to continue as before.
- Does not see the benefit of standardisation, until explicitly explained/demonstrated or through new

Willing

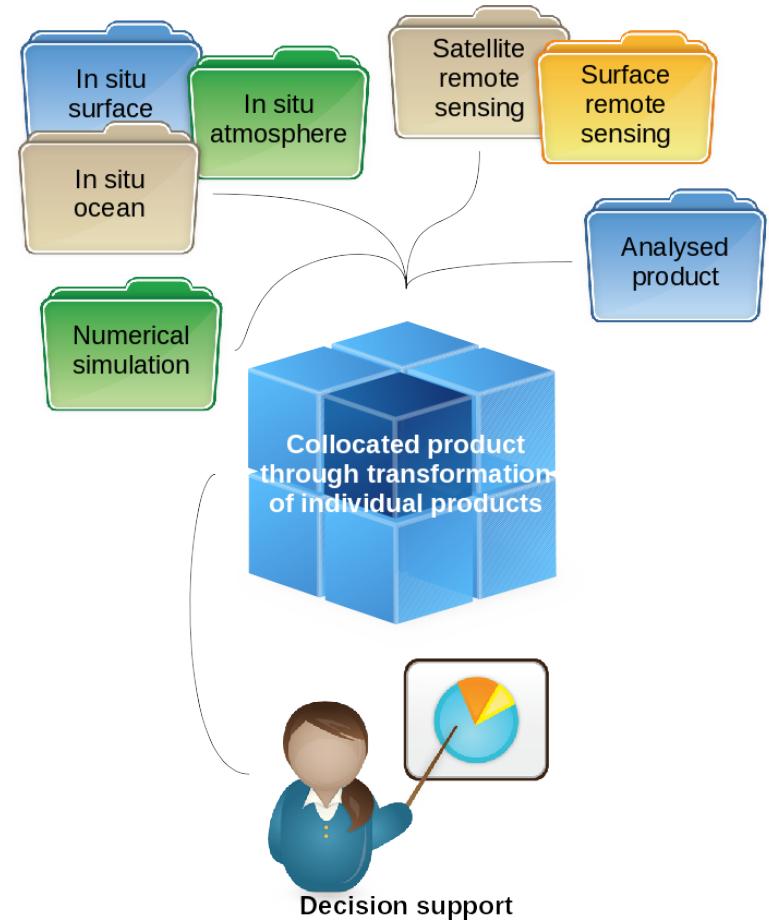
- Wants to translate between provider and consumer.
- Still relies on some sort of standardisation in order to be cost effective.
- Must know dimensions, structures, content, missing values, units, aggregation levels, ...

Unwilling and skeptical to potential users

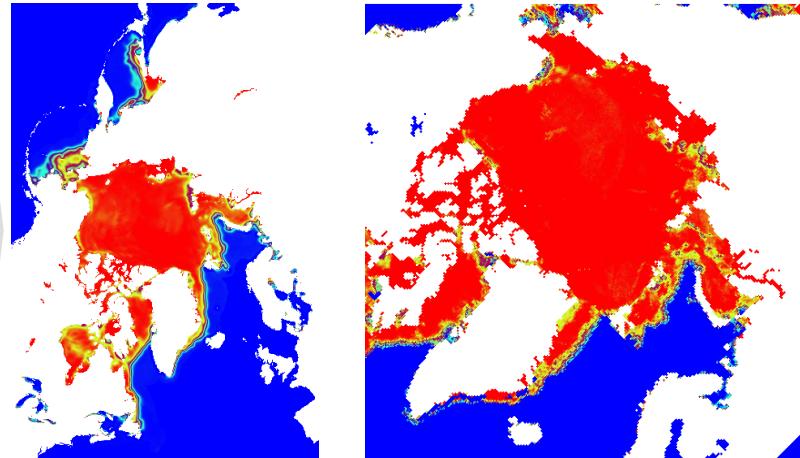
- Do not want to change behaviour, legacy system(s).
- Want to continue as before.
- Understands own requirements (knows the data well).

Strømming av data

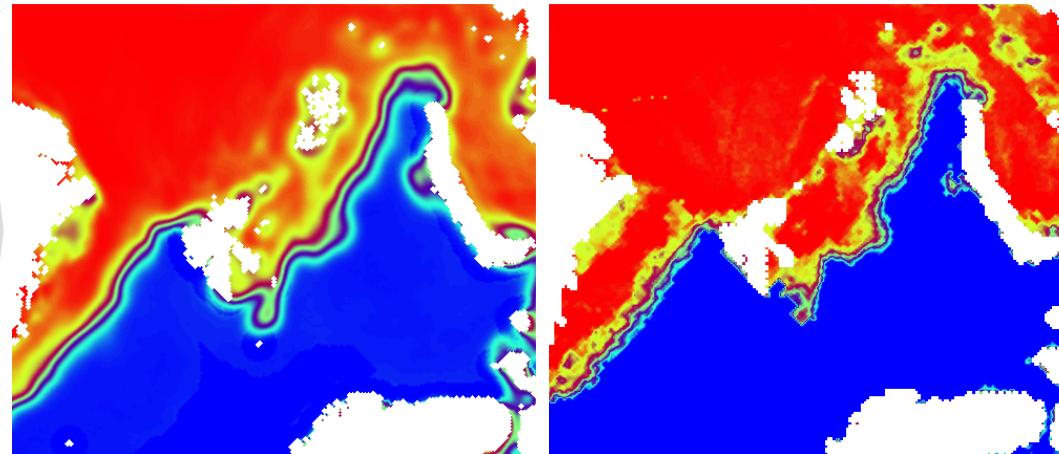
- Open-source Project for a Network Data Access Protocol
 - DAP kobler en enkel datamodell til operatorer
 - Generisk
 - Brukes som en web service (REST)
 - Responderer med dokumenter
 - Innholdsbeskrivelse
 - Dataleveranse
 - Datareduksjon på leverandørsiden
 - Unngår bokholderi
 - Benyttes av
 - Copernicus Marine Environmental Monitoring Service
 - Copernicus Climate Change Service
 - Earth System Grid Federation (IPCC)



Search
results



Transformation
request



Transformation

OGC WPS

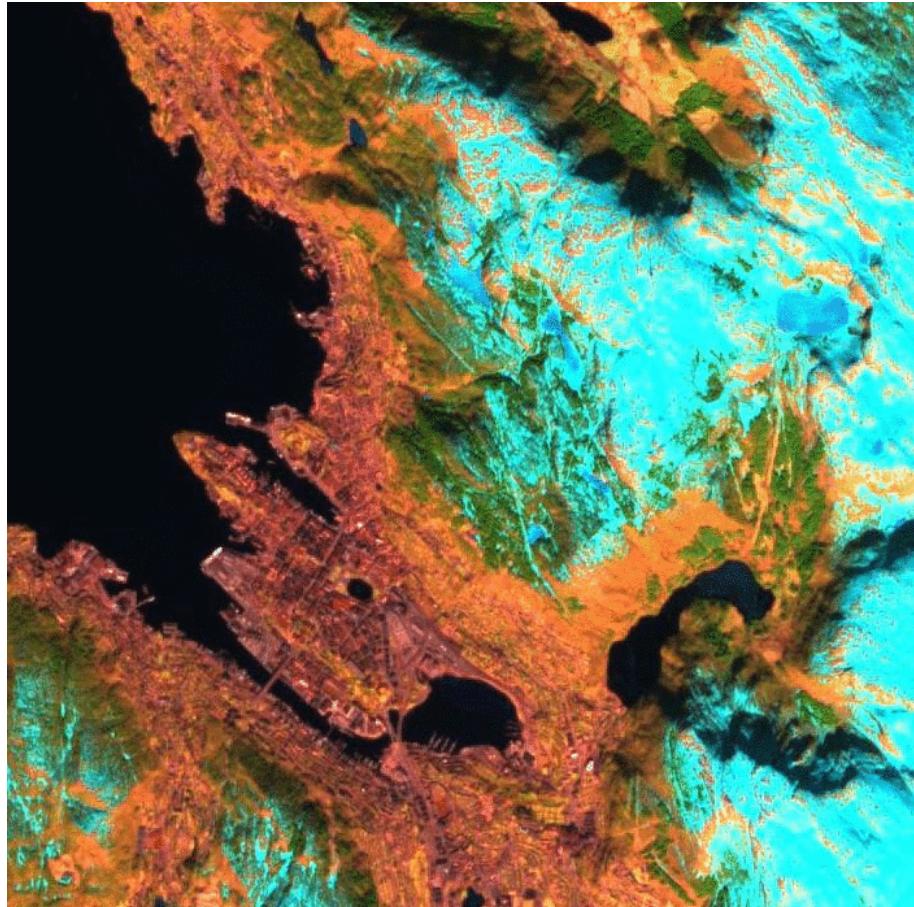
- Subsetting
- Reprojection
- Reformatting
- Relying on OPeNDAP

Example Sentinel-2:

- Extract variables B4, view and zenith angle over predefined AOI
- Reproject from UTM 32N (Tile VKN) to Arctic Polar Stereographic (EPSG:3995)

Result

8.7Mb vs 771Mb file ready to use in favoured projection



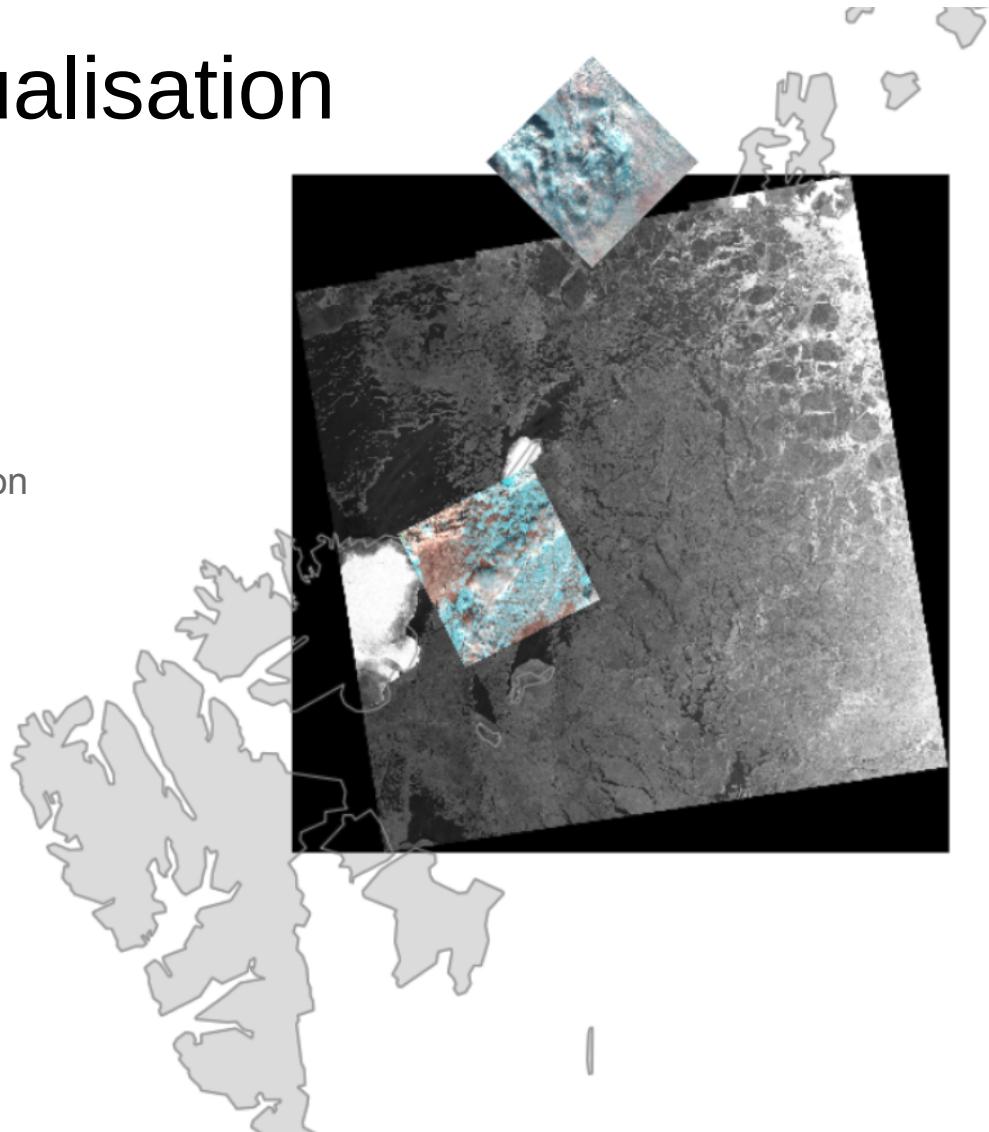
Visualisation

Sentinel-1

- All raw polarizations
- 40x40 m pixel resolution

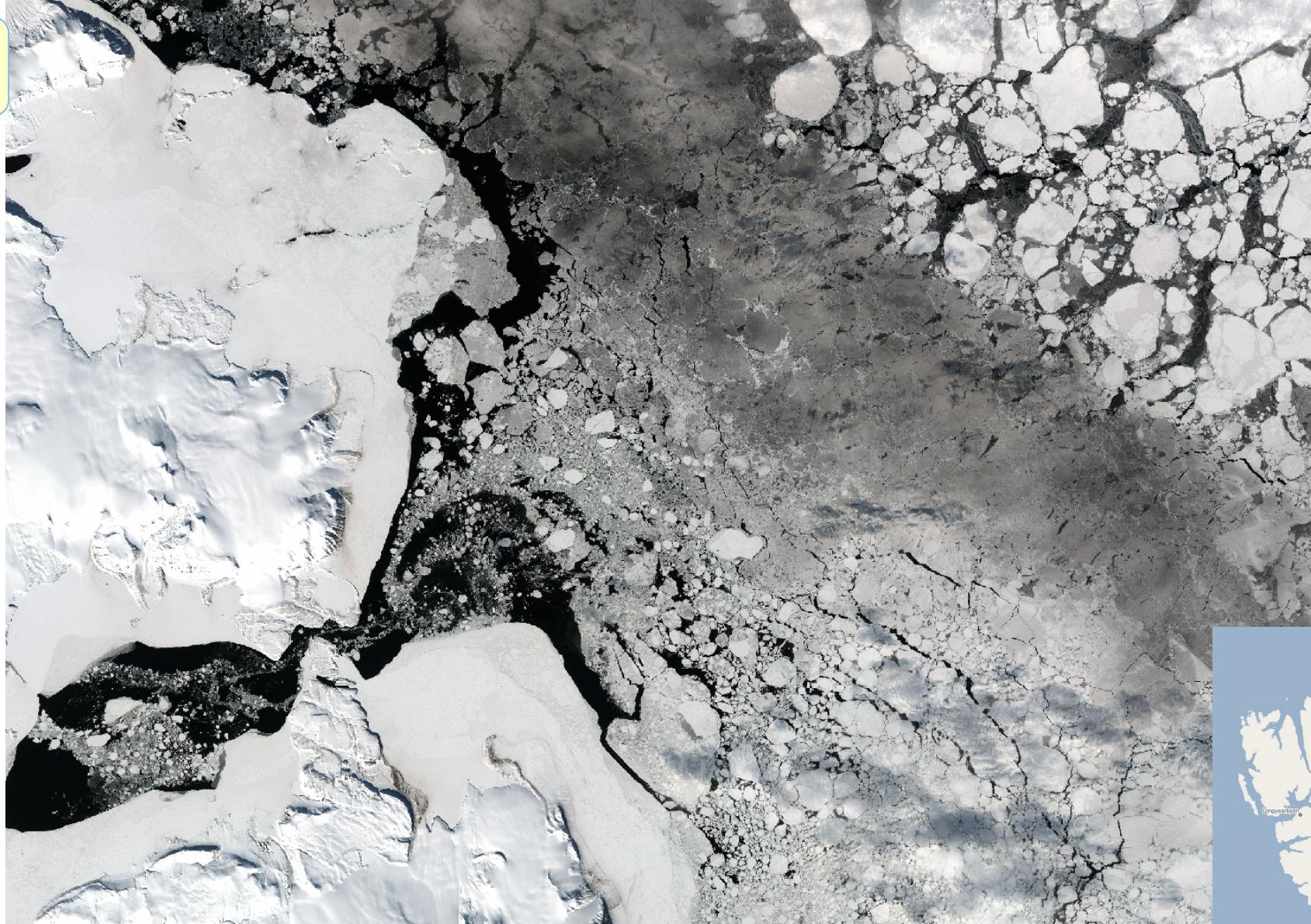
Sentinel-2

- All bands resampled to 10x10 m pixel resolution
- Three RGBs



1.5 km
↔



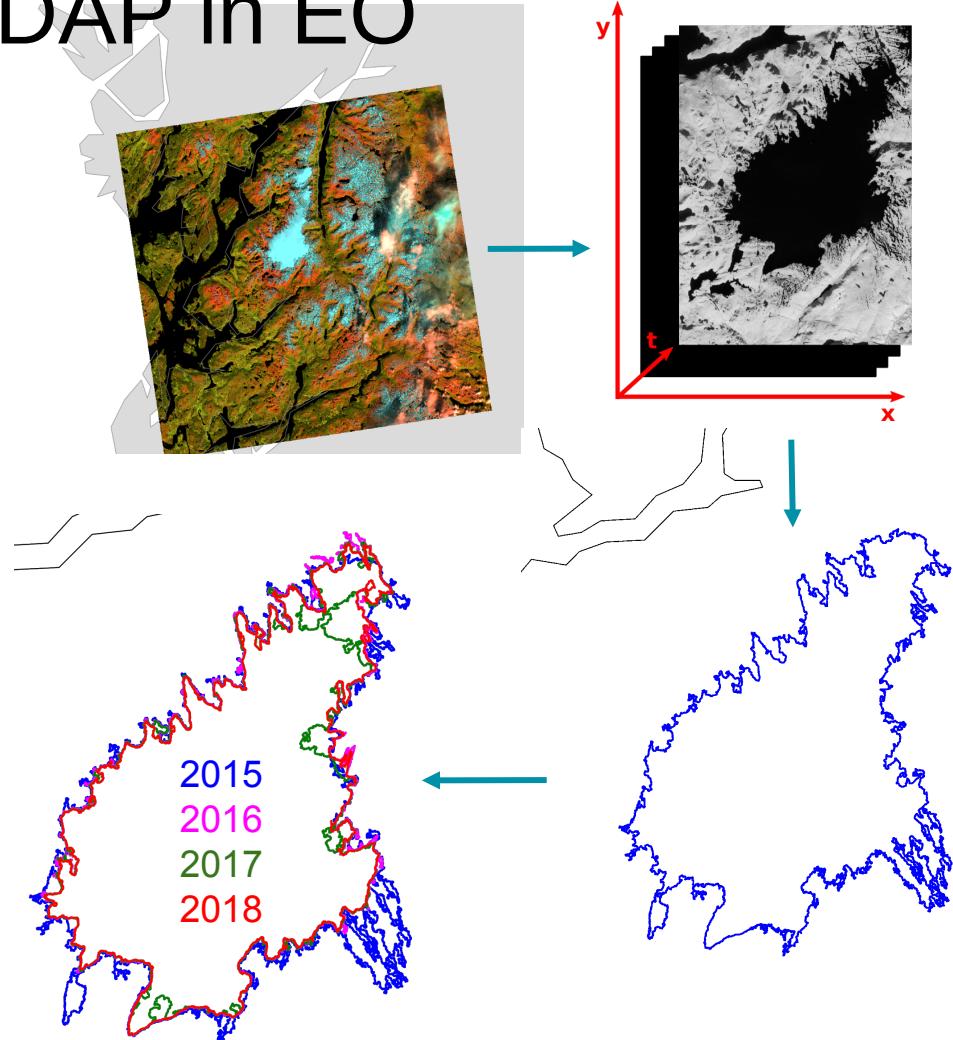


1.5 km
↔



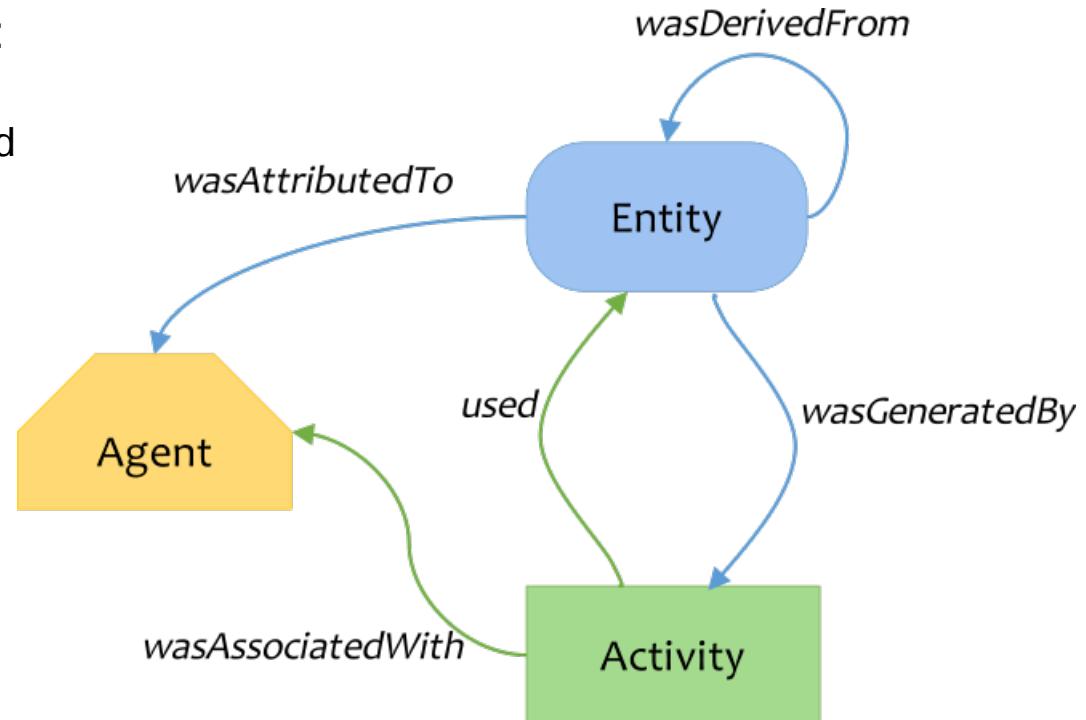
Utilizing OPeNDAP in EO

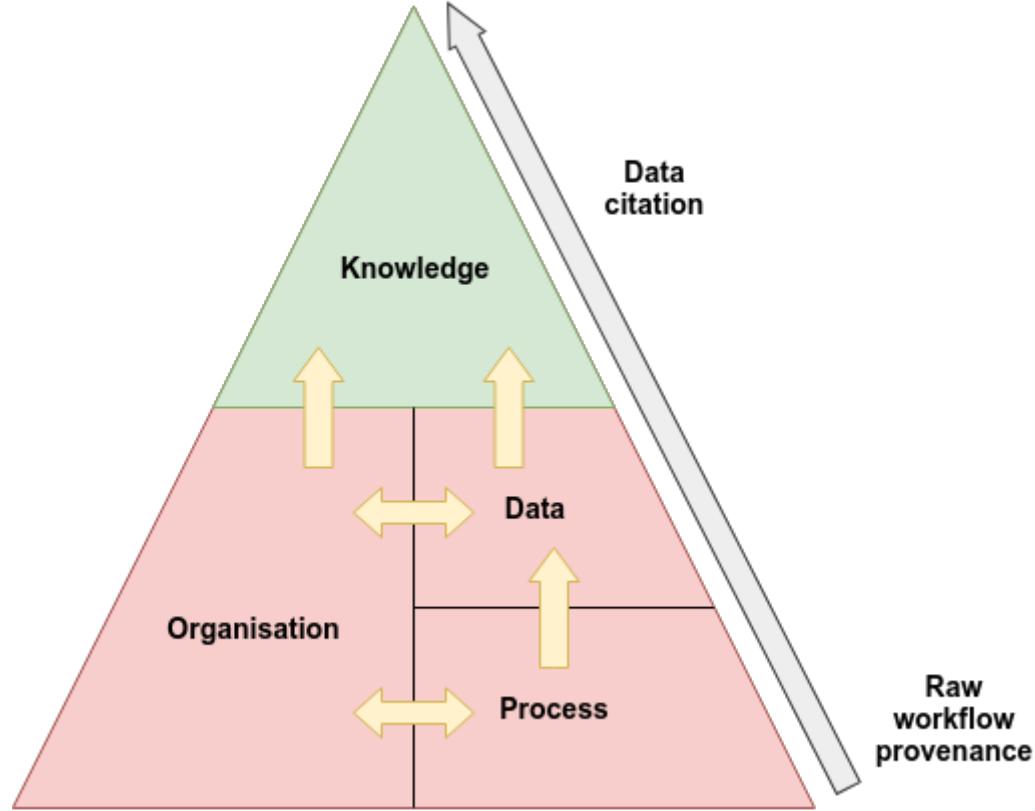
- Task: Decide glacier extent from S2 data in Python
 - Created polygon by means of image slicing.
 - Used 0.2% of all the pixels in the product.
 - Analyse Ready Data / Data Cube
 - Temporal aggregated product covering the same area.
- Available as Jupyter Notebook through gist:
 - <http://nbviewer.jupyter.org>
 - Gist code:
 - d5d29f5b2d691d8ed3c4b3cd65e2009e
- Courtesy of Trygve Halsne



Provenance

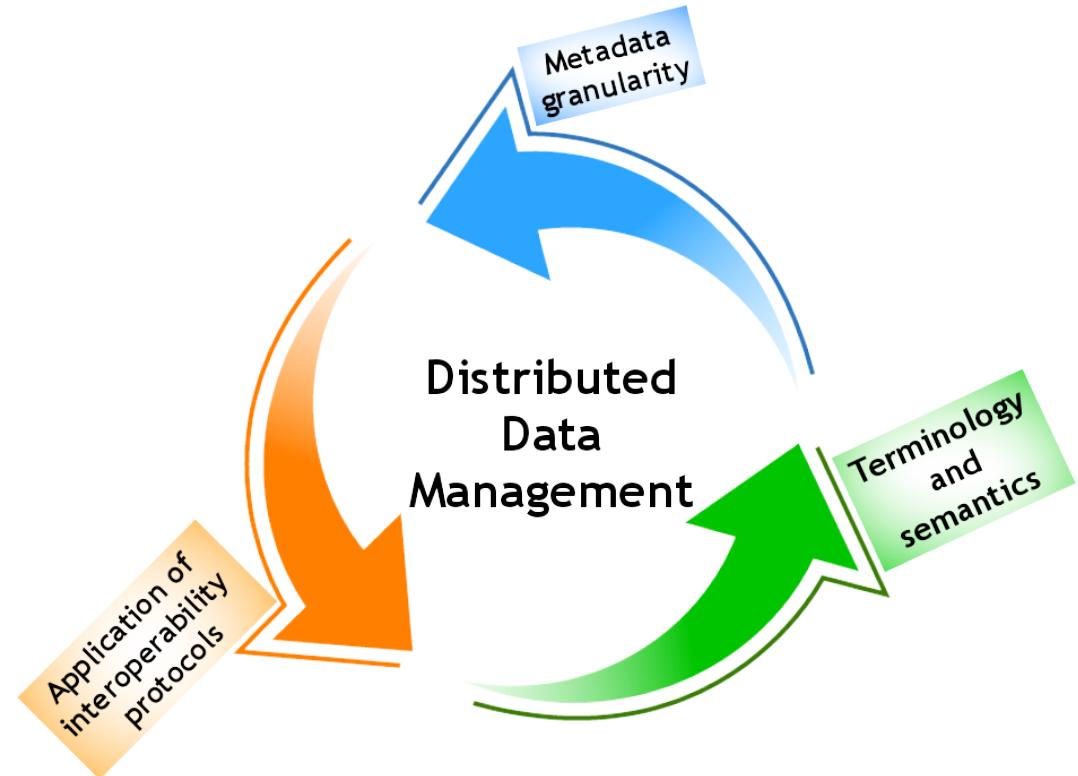
- A provenance framework should support:
 - the core concepts of identifying an object, attributing the object to person or entity, and representing processing steps;
 - accessing provenance-related information expressed in other standards;
 - accessing provenance;
 - the provenance of provenance;
 - reproducibility;
 - versioning;
 - representing procedures;
 - and representing derivation.





Utfordringer med dynamiske data i distribuert dataforvaltning

- Interoperabilitet
 - Søksmetadata
 - Semantisk annotasjon/terminologier
 - Detaljeringsgrad
 - Koding av data
 - Datamodeller/strukturer (-)
 - Semantisk annotasjon/terminologier (-)
- Søksmetadata på samlings- eller datasettnivå?
- Fokus på datasett eller tjenester?
- Semantisk rammeverk og kontrollerte terminologier
 - Selvforklarende data
 - Må også beskrive kvalitet
 - Anvendelse av ontologier og tjenester knyttet til disse



Muligheter

Fokus under store satsinger på statsbudsjettet



2015-08-15

Spørsmål?

