



Array Database Assessment WG

Peter Baumann, Kwo-Sen Kuo
2017-sep-20, Montreal, P10

(Spinoff from Big Data IG)

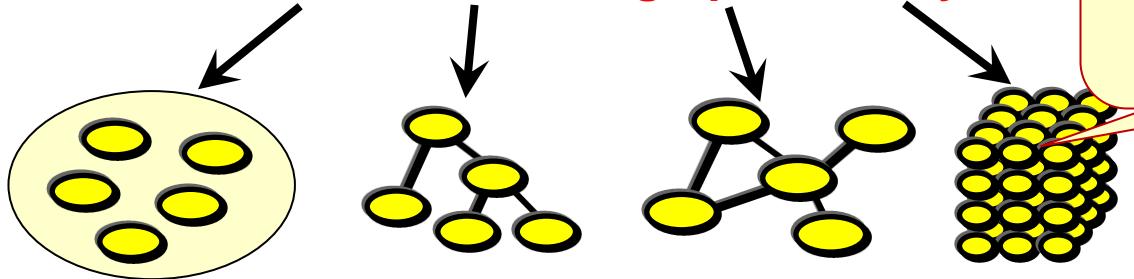
[Dali]

Summary of the Problem

- Multi-dimensional arrays play key role in science & engineering & beyond

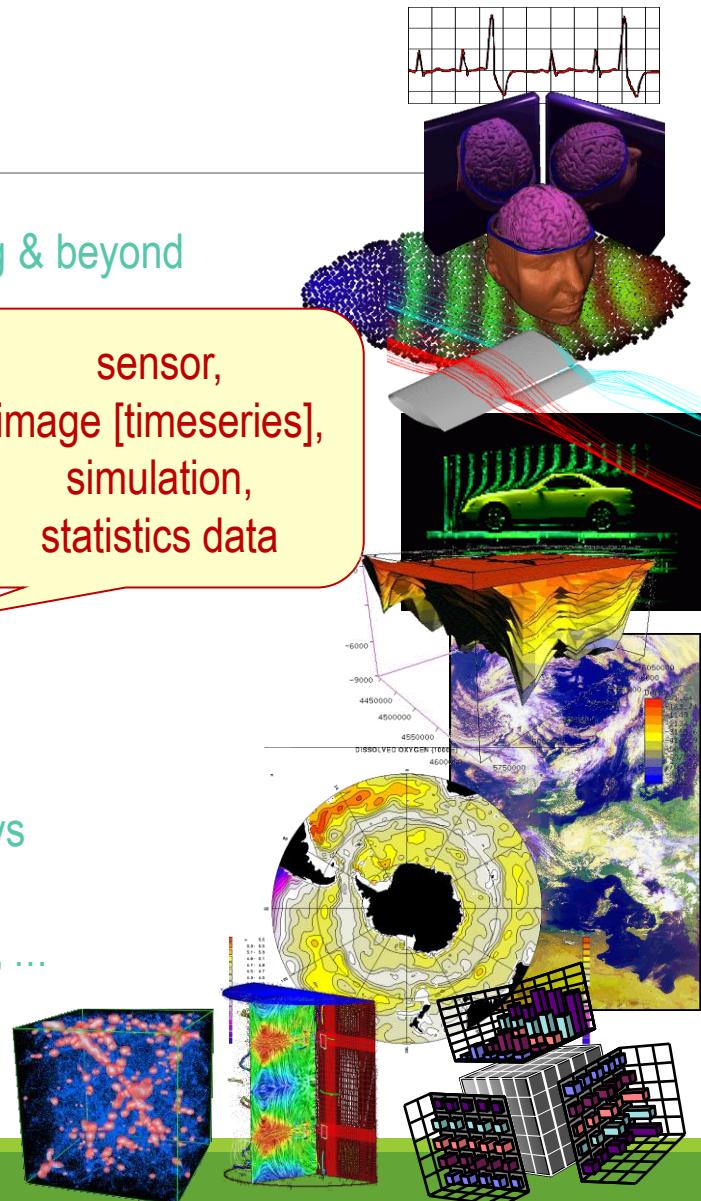
- Increasingly in focus as „datacubes“
- But not supported by SQL nor NoSQL nor NewSQL

sets + hierarchies + graphs + arrays



sensor,
image [timeseries],
simulation,
statistics data

- Array Databases = modeling + query support for massive arrays
 - Server-side evaluation = ship code to data
 - Adaptive partitioning, parallelization, distribution, mixed hardware, ...
- Various tools known – benefits? Maturity? How-tos?



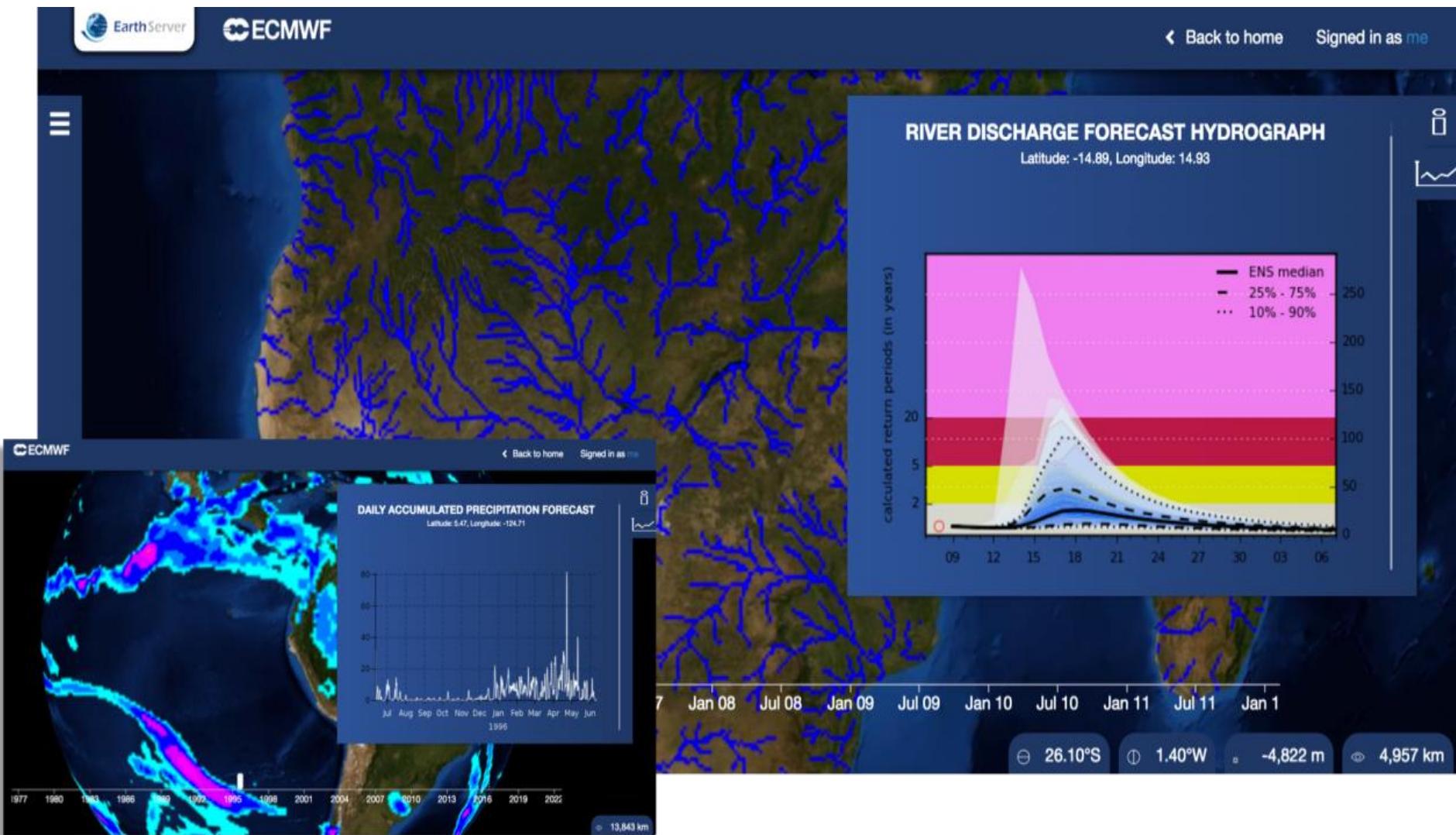
Status of the Deliverable

-
- Report compiled at <https://www.rd-alliance.org/group/array-database-assessment-wg/wiki/array-database-assessment-working-group>
 - Structure:
 - Introduction to Array Databases -- done
 - Array Standards -- done
 - Array Technology -- 17 tools listed, **open-ended**, champion-based
 - Publicly Accessible Array Services -- several, Petascale
 - Array Systems Assessment -- **under work**
 - Summary -- **tbd** at end
 - 36 WG members
 - Petascale services, first benchmarks
 - Visible contribution to datacube stds in ISO, OGC, INSPIRE
 - WG extended until P11

Initial Adopters

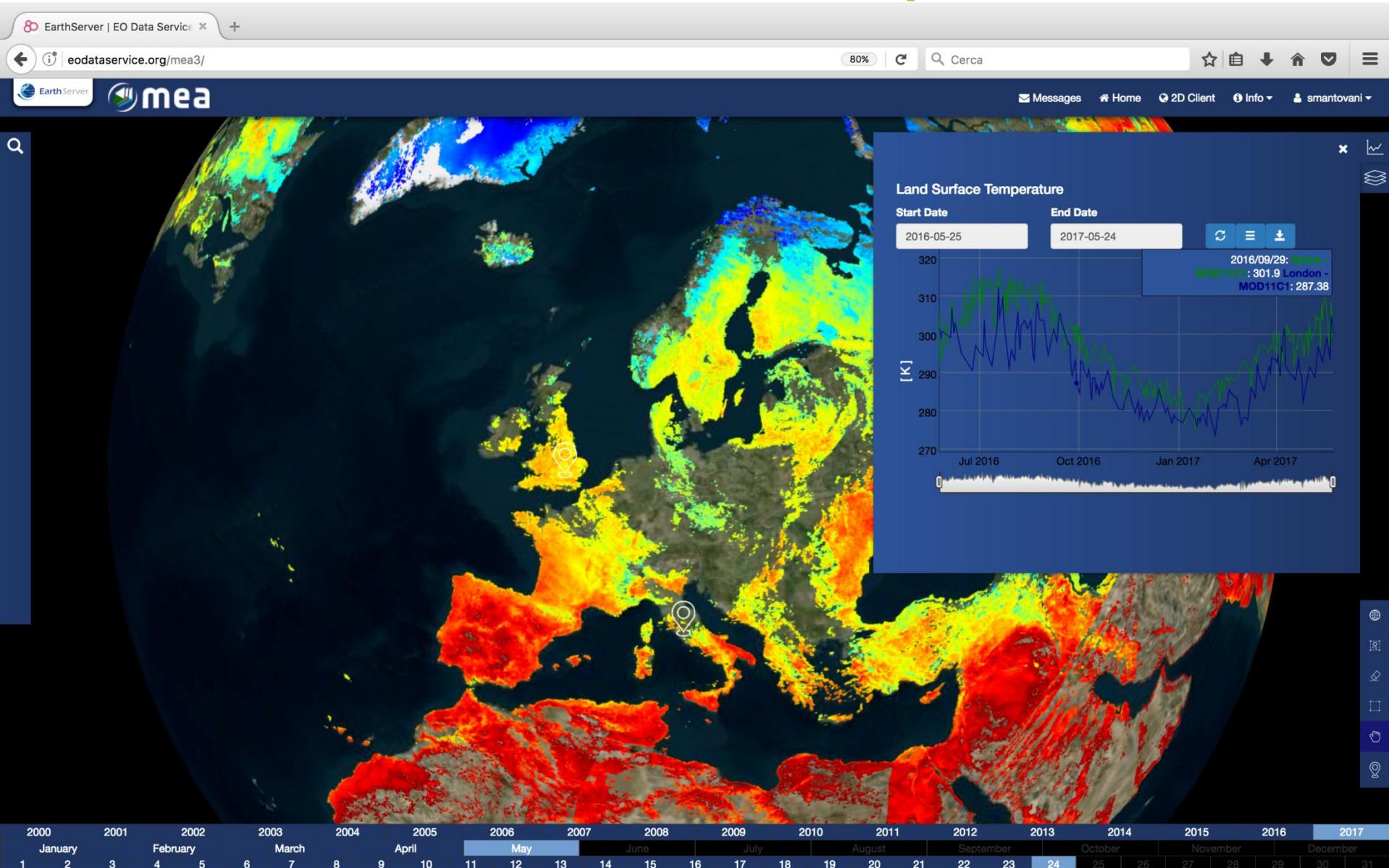
- ECMWF/UK
- PML/UK
- MEEO, IT
- NCI Australia
- EOfarm, GR
- *...and several more*
- ESA, 2017: „The RASDAMAN product is currently the world leading environment in this domain and the standard working horse for OGC standardisation on these innovative data access interfaces.”
- Standards, ex: ISO SQL/MDA (Multi-Dimensional Arrays)

ECMWF: River Discharge



[system used: rasdaman]

MEA: Land Surface Temp, Cloudfree



[system used: rasdaman]

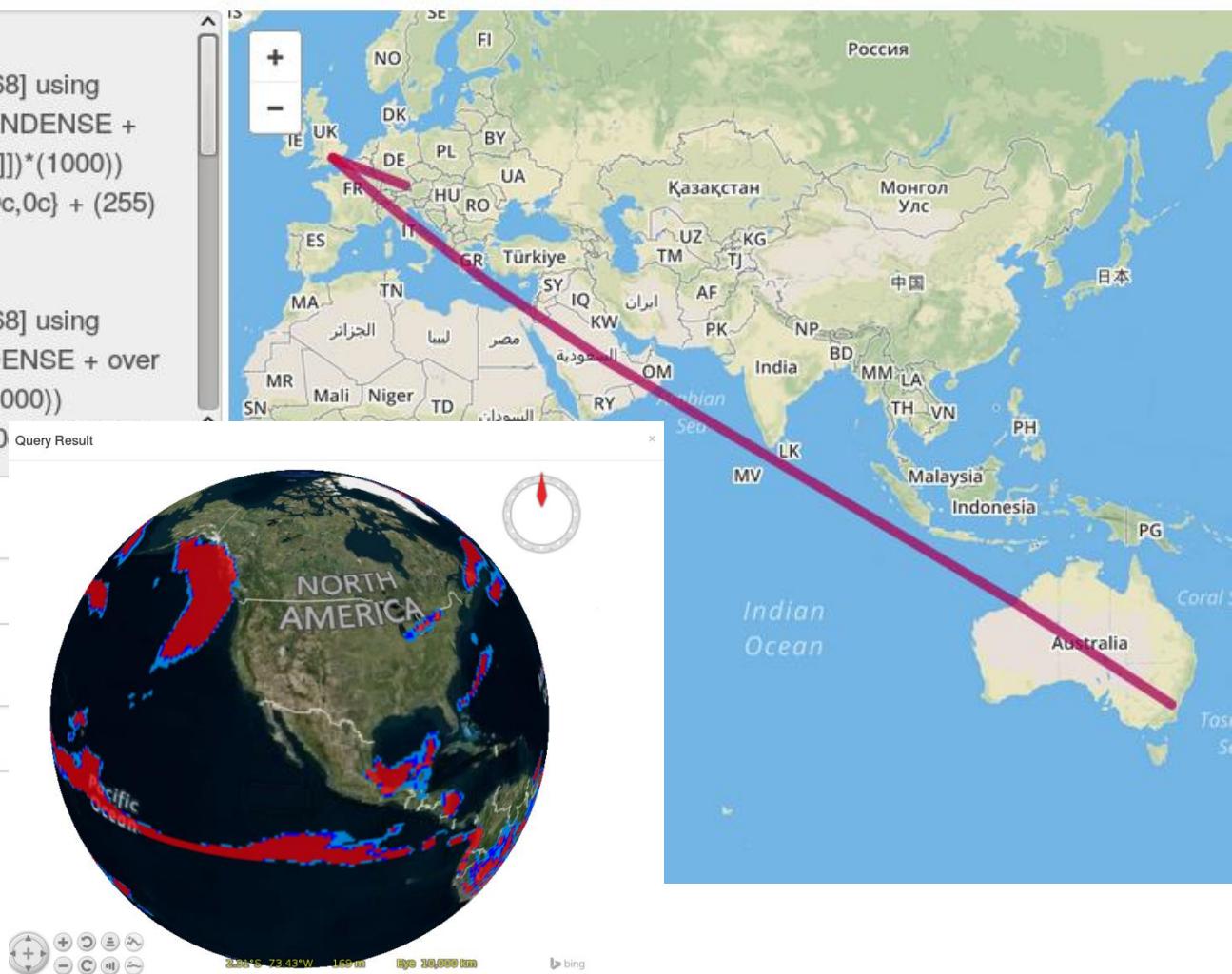
Datacube Federation

SELECT ENCODE(CASE

WHEN (CONDENSE + over i_i in [42364:42368] using
d[0:3600, 0:1800, i_i[0]] / 1423 + 1.47) > ((CONDENSE +
over i_i in [42364:42368] using (c) [*:*, *, i_i[0]])*(1000))
THEN ((255) * {1c,0c,0c,0c} + (255) * {0c,1c,0c,0c} + (255)
* {0c,0c,1c,0c} + (0) * {0c,0c,0c,1c})

WHEN (CONDENSE + over i_i in [42364:42368] using
d[0:3600, 0:1800, i_i[0]] / 1423 + 4) > ((CONDENSE + over
i_j in [42364:42368] using (c) [*:*, *, i_j[0]])*(1000))
THEN ((0) * {1c,0c,0c,0c} + (128) * {0c,1c,0c,0c})

Query Result



Query:

Heavy rainfall risk areas

Server:

ECMWF

Run Query

[system used: rasdaman]

Expected Impact

- Increase **uptake** of Array Databases in Science & Engineering
 - in academic and industrial environments
 - Systems, features, benchmarks, deployments, ...
- Datacube **standards** in OGC, ISO, INSPIRE
 - Providing systematic background
 - Active shaping, eg: ISO SQL/MDA (Multi-Dimensional Arrays)

Expected Impact

Before

- Array Databases **not widely known**, FUD about novel paradigm
- Largely unknown how to use, what benefit „datacubes“ bring
- Software choices unclear
- Lack of datacube interoperability

After

- Report helps establishing broad knowledge and understanding, “datacubes” getting in focus
 - Dissemination work supports spread of datacubes, eg, in CEOS & GEOSS
- Large-scale services demonstrate feasibility
- Annotated listing of software + benchmarks **ease selection**
- Standards getting in place
 - ISO SQL/MDA
 - ISO & OGC „Big Earth Datacube“ stds

Feedback Desired from RDA Community

- Work progressing, but further contributions welcome anytime
- More contributors; participant types (can overlap):
 - Use case providers: data & desirable queries
 - Service providers: storage & compute facilities, hosting datacubes
 - ADBMS software providers: software + support
- Have Earth & Space, want Life sciences
- *Stay tuned for P11 !*

