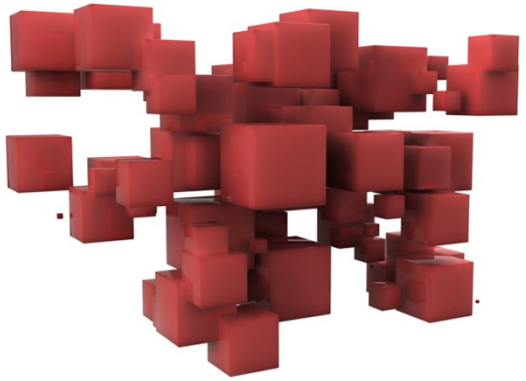
Listen to Community Requirements

Exploring options towards EUDAT next data services

Morris Riedel et al.

Juelich Supercomputing Centre
(on behalf of Peter Wittenburg)

Track 4 – New Services
EUDAT 2nd Conference
28.10. – 29.10.2013, Barcelona



SEVENTH FRAMEWORK PROGRAMME

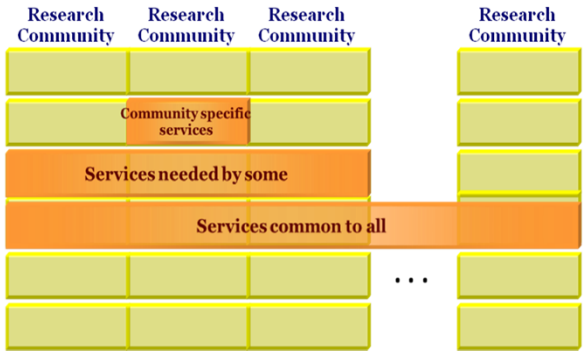
EUDAT

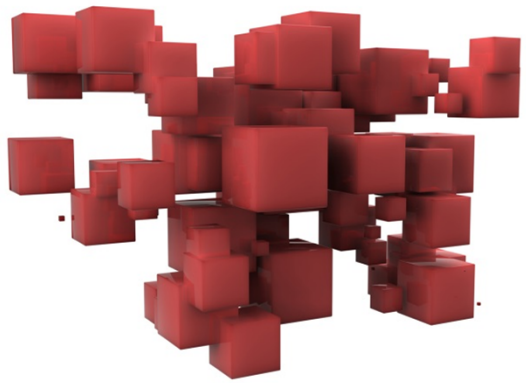
Outline

- What we mean with ‘Listen’
- EUDAT Services
- Surveys and Working Groups
- Porential New Service Areas
- Summary

EUDAT

Listening to Requirements of Communities





Common Services	CLARIN	LW	VPH	ENES	EPOS	INCF	ECRIN	Bio Vel	Dixa	CESSDA	DARIAH	Pan Data	BBMRI	EMSO
Safe Replication	X	o	X	X	X	X			x		x			
Data Staging	o	o	X	X	X									
SimpleStore	X	X	X	X	X	x	x	x	x	x	x		x	
Metadata	X	X	o	X	x	x	x	x	x	x	x	x	x	x

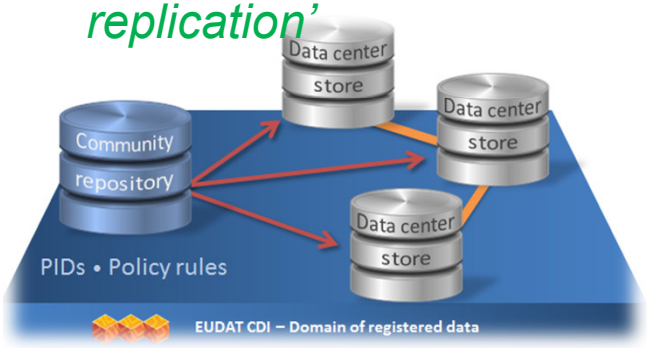
Current and New EUDAT Services

➢ What potential new services are needed?

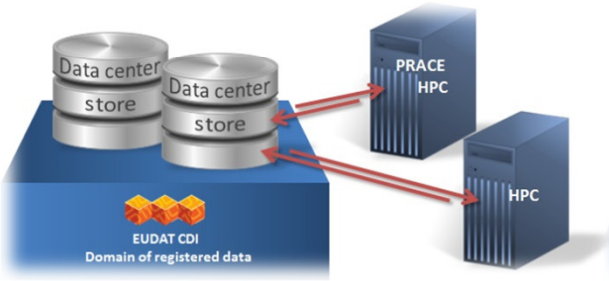
❖ *trusted & simple ad-hoc data sharing with limited number of researchers ('Dropbox')*

✓ *'metadata management'*

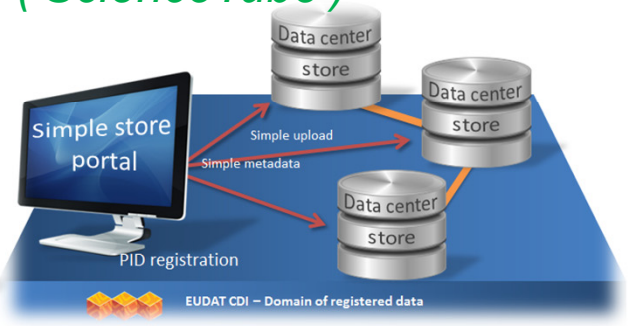
✓ *'trusted replication'*



✓ *'smart transfer to & from computing'*

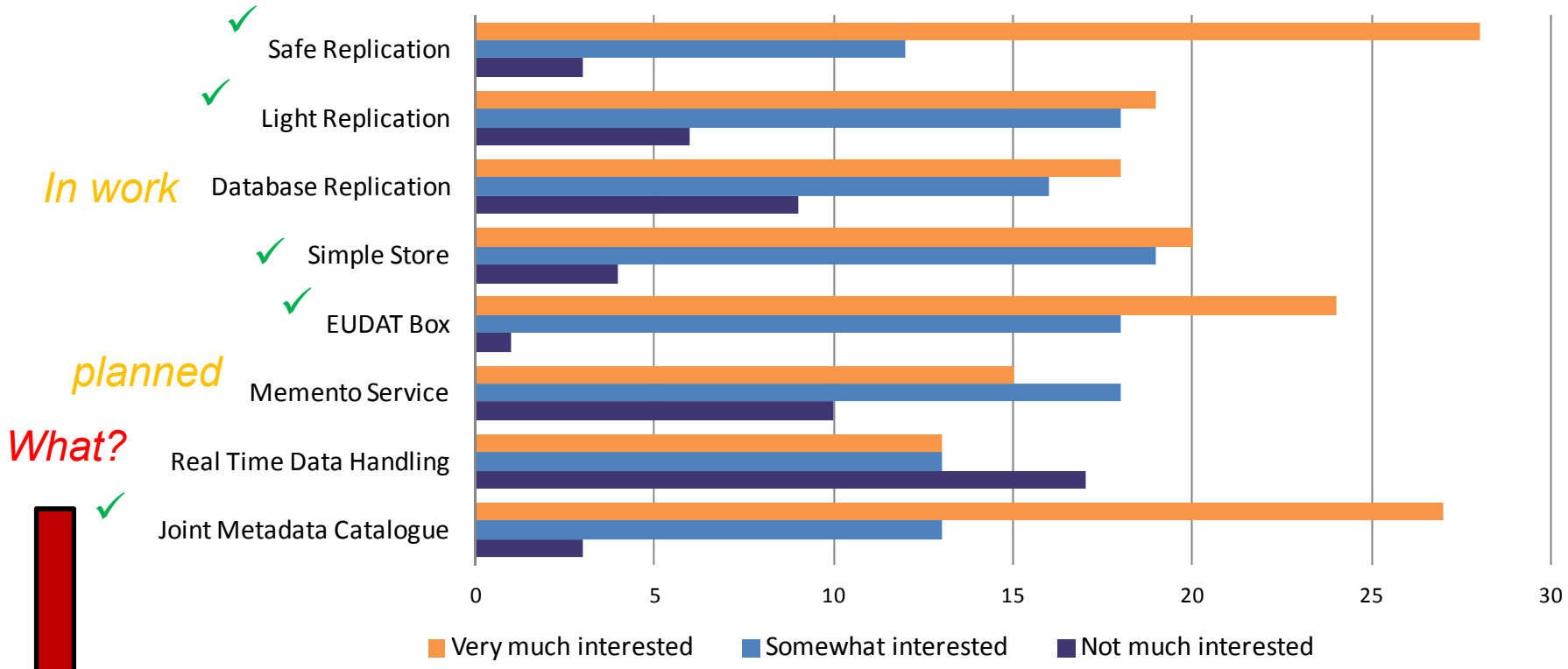


✓ *'organized simple broad sharing' ('ScienceTube')*



EUDAT

Apart from Discussions: Surveys



Requires a deeper understanding of what needs to be done
→ Forming working groups…

EUDAT

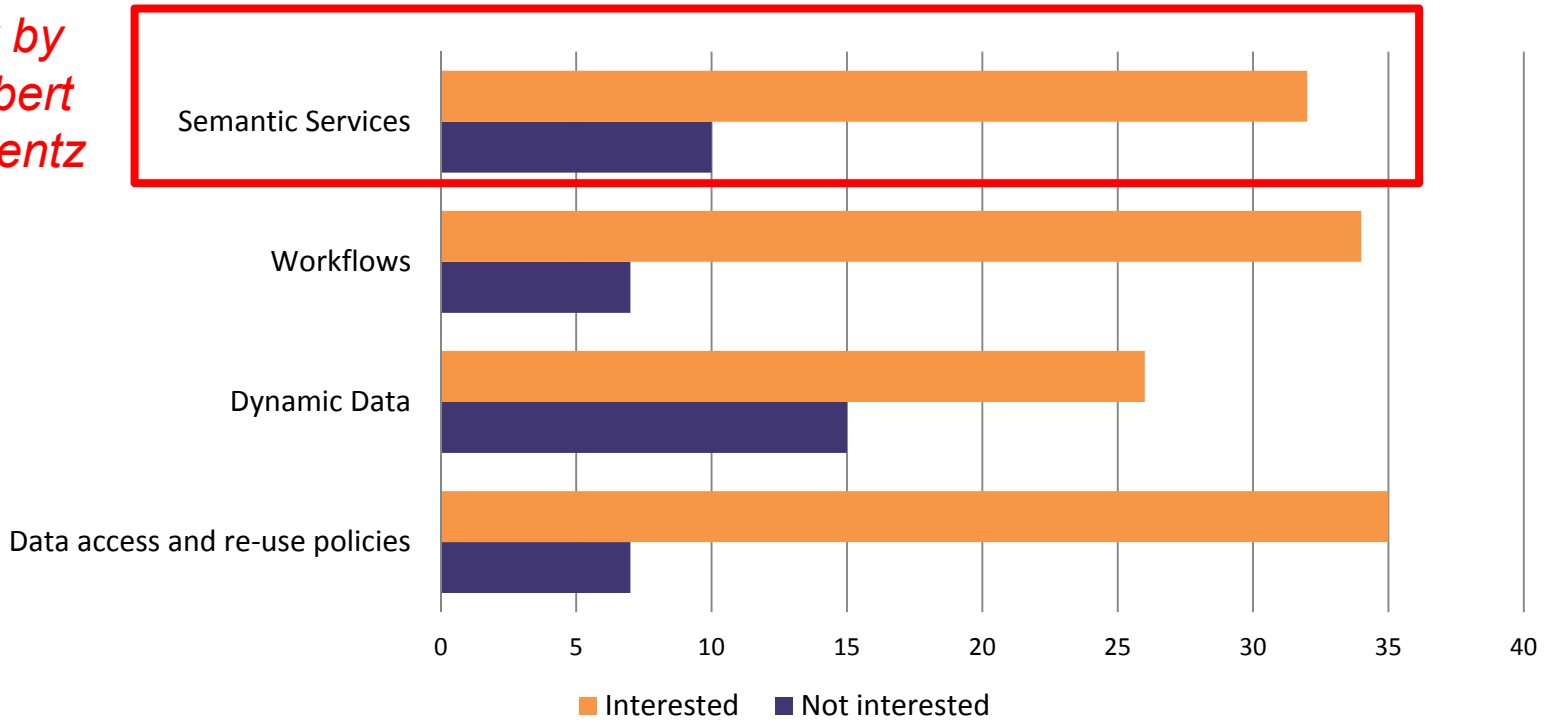
Forming Working Groups

- Not all is *aaS → Bring together…
  - Domain experts in identified interested fields
  - EUDAT-related community representatives
  - EUDAT technologists
- Discuss the challenges and opportunities associated with potential common services across communities
- Reach consensus on the ‘common services’ that will be of most benefit to research communities in their work.
- Series of events
  - **EUDAT Workshop Days, Barcelona, 25.09.– 26.09.2013**
  - Community events, e.g. American Geophysical Union Assembly 2013

EUDAT

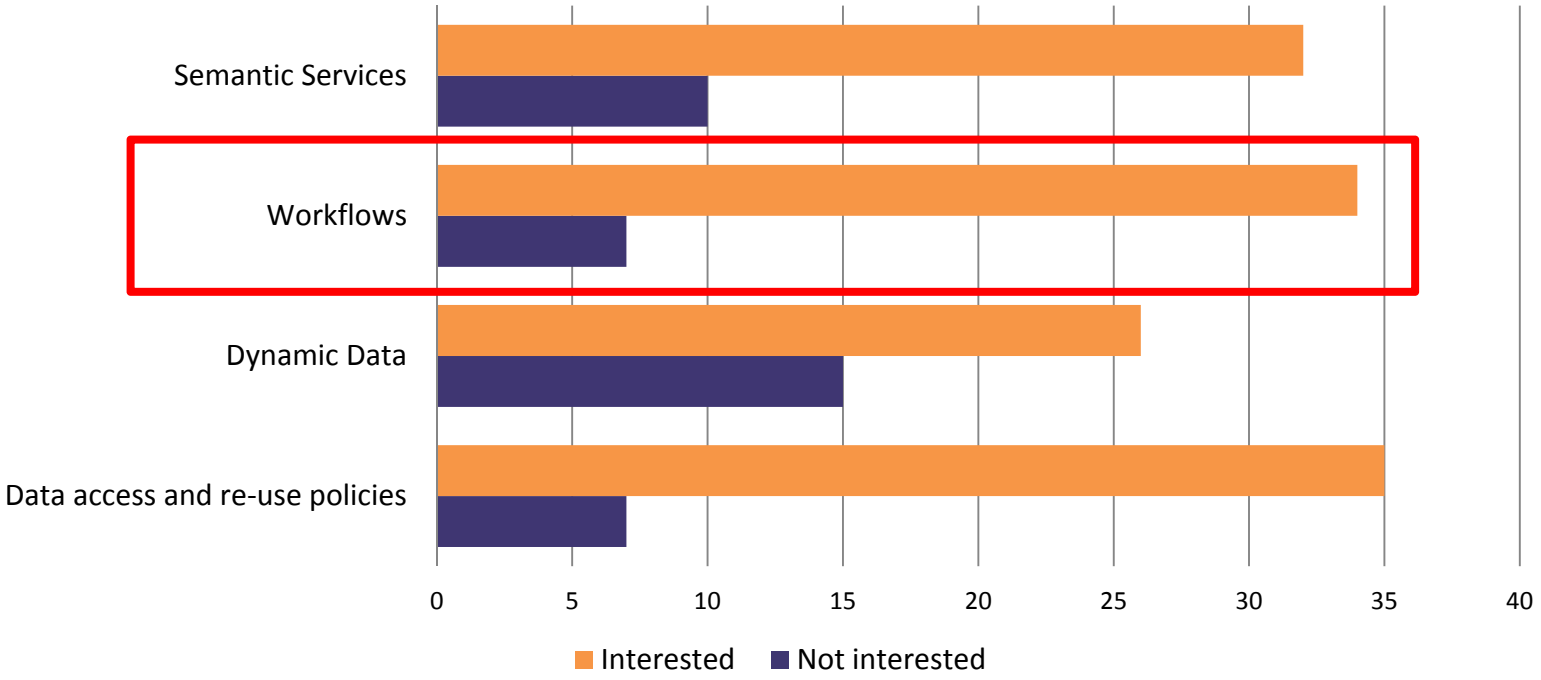
Increase Understanding in Working Groups

Talk by Herbert Schentz



- *Areas could lead to different types of 'new services' → team up with domain experts*

Understanding Workflows



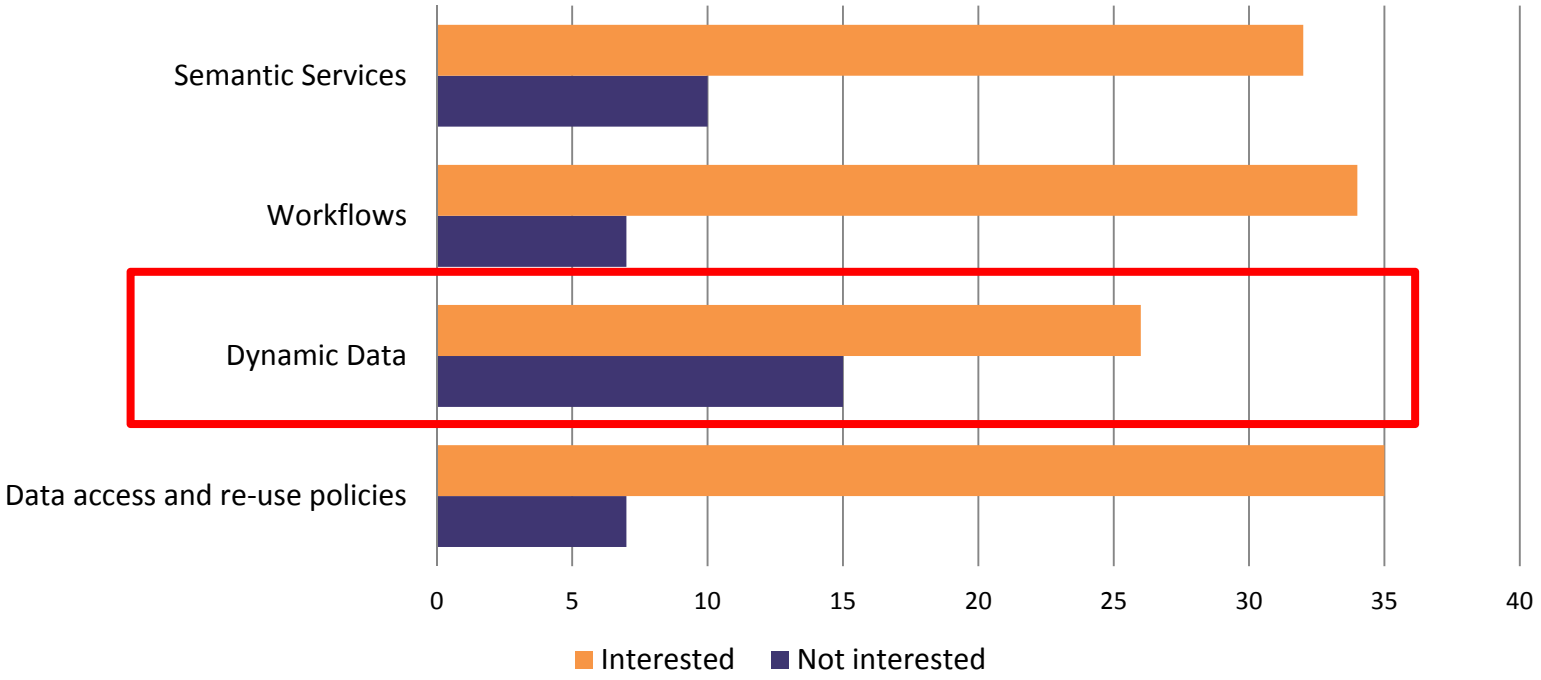
Understanding Workflows

- Understanding the needs of community experts in terms of common services in the area of *'workflow support'*
  - The focus was on common services supporting researchers who need to orchestrate data processing chains (workflow components)
  - E.g. Execute them on computers close to where their data is located.
- ~20 workshop participants have been international experts in the field of 'scientific workflows' (also from DataONE)
  - *'common workflow services elements' related to workflow provenance, APIs for re-use of EUDAT services in workflows, statistics/data mining workflow elements*

➢ *If interested, ask us for the report or join the group!*

EUDAT

Understanding Dynamic Data



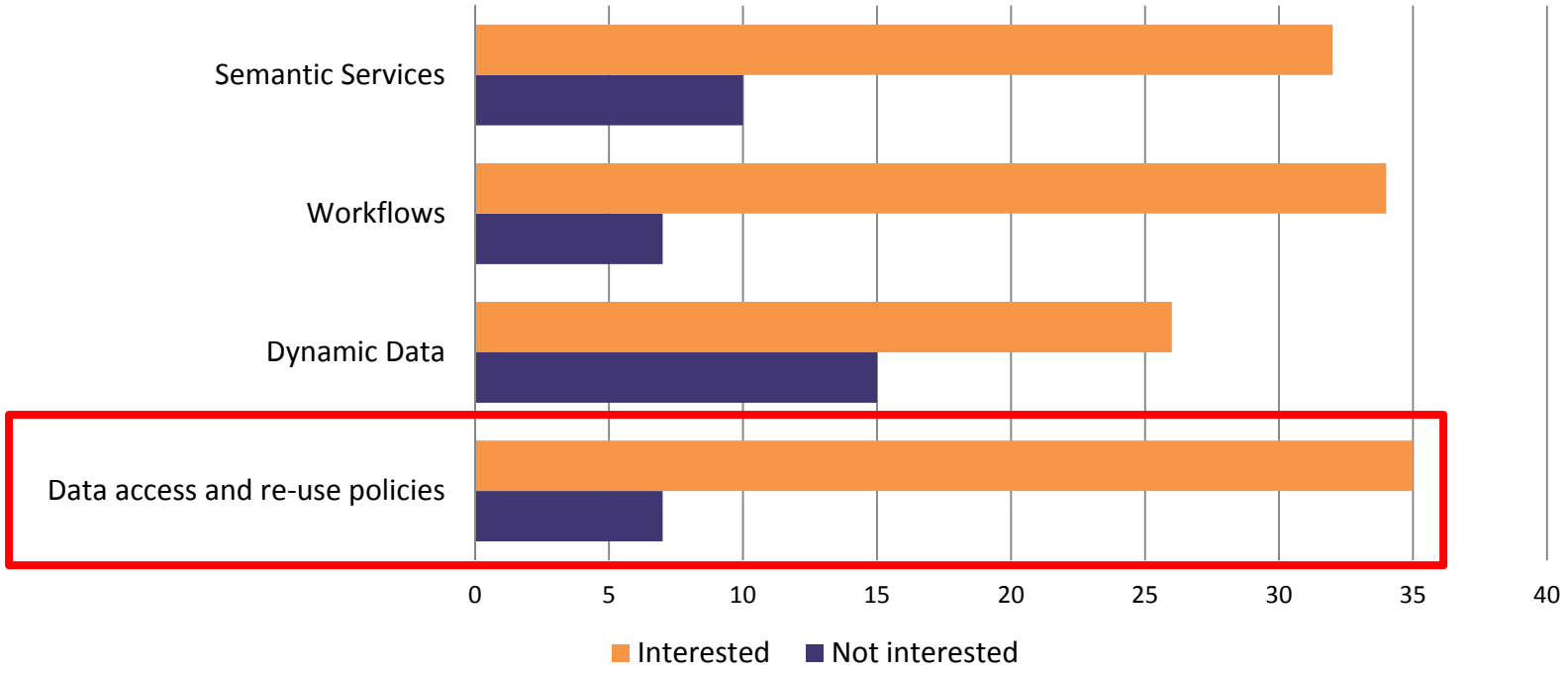
Understanding Dynamic Data

- Understanding common services for Dynamic Data…
  - … addressing real-time data produced by remotely connected field sensors
  - … addressing real-time data from massive crowdsourcing data generated via mobile devices
- Understanding how to overcome challenges
  - Data streams that may be temporarily incomplete
  - Data that may consequently fill up over time (automatically or after manual intervention)
  - Data can be filled up at random moments (special ‘data peek event’ → e.g. crowdsourcing)

➢ *If interested, ask us for the report or join the group!*

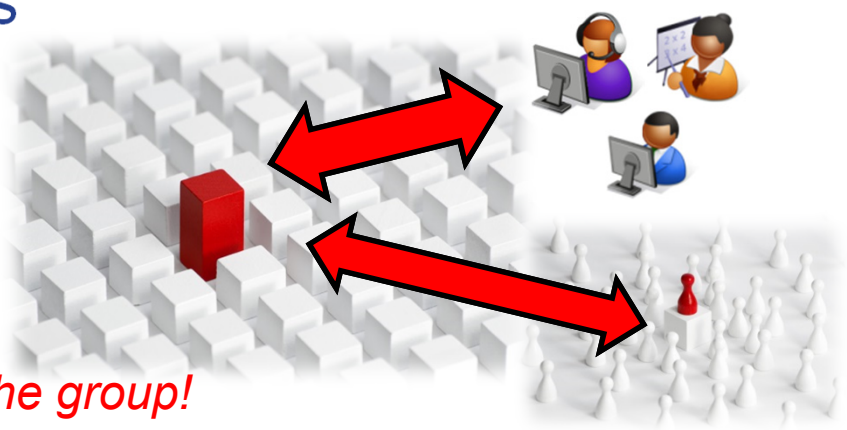
EUDAT

Understanding Data Access & Policies



Understanding Data Access & Policies

- Data access and reuse policies must be made clear
  - … to potential users of data infrastructures
  - … the “data providers” – who need to know how their data will be made accessible and who will be able to access it
  - … the “data customers” – who need to know about access or reuse restrictions before starting working with any of this data
- Discuss common approaches
  - Specify principles regarding open data access and reuse by maintaining data origins and data generator visibility



➢ *If interested, ask us for the report or join the group!*

EUDAT

Summary

- EUDAT services are based on community requirements
  - Already in production or well understood to move towards it
- New EUDAT services are based on community requirements
  - Some of them are well understood and already prototypes are started (e.g. replication of MediaWiki ‘database’)
  - Others require more elaborative discussions with domain experts
- Working groups together with domain experts
  - Four areas identified where ‘some type of services’ might emerge
  - Started Working Groups: Semantic services, Workflow Support, Dynamic Data, Data Access and Re-use Policies

➢ *If interested, ask us and get involved shaping EUDATs future!*

EUDAT

services
measurement
within
fusion
new
funding
Policy-based
concepts
device
Analysis
forms
cross-disciplinary
resources
Centre
climate
Computer
expertise
year
like
Computational
Test
E.g
Computing
disciplines
results
Approaches
construction
Fusion
Enable
different
Cross-Disciplinary
EU
methods
High
big
nodes
PB
community
Environment
Modelling
Key
scientific
important
Science
centers
storage
basis
Infrastructure
files
etc
DKRZ
Use
model
SMAQ
international
engineering
References
formats
simulations
ENES
technologies
brain
increasing
via
EUDAT
physical
IPCC
Big
Project
Data
processing
Common
TB
Methods
computing
Using
European
MapReduce
create
databases
GB
Supercomputing
Work
images
codes
analysis
DLCL
stored
Resources
analyze
Large
Summary
data-intensive
Service
Scientific
levels
often
use
Level
Research
external
JSC
computational
Matters
Concrete
performance
research
e.g
science
HPC
large-scale
Juelich
system
datasets
acquisition
compute
Landscape
System
Climate
modelling
AR5
Hardware
cores
access
hundreds
Nutshell
Services
management
DLCLs
Understanding
structures
Technologies
Earth
distributed
Structure
current
network
Germany
Simulation
directory
project
General
Provide
NASA
Energy
systems
day
Processing
Health