Bridging the gap between Big Earth data users and future (cloud-based) data systems

Towards a better understanding of user requirements of cloud-based data systems

Julia Wagemann[1,2], Stephan Siemen[2], Jörg Bendix[1], Bernhard Seeger[1]
[1]University of Marburg, [2]European Centre for Medium-Range Weather Forecasts
EGU 2020: Sharing Geoscience Online, 7 May 2020

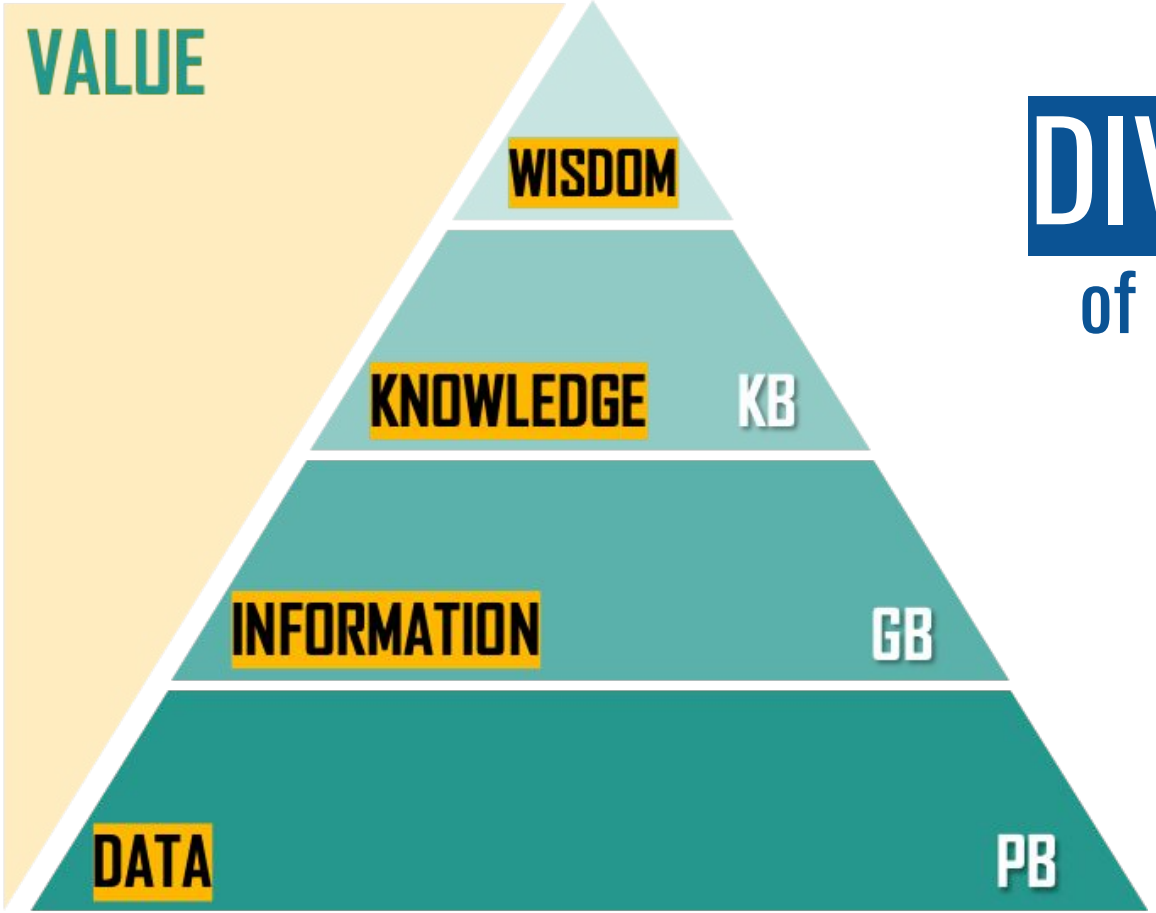
Philipps Universität Marburg

ECMWF

#1 Trend*

DIVERSIFICATION

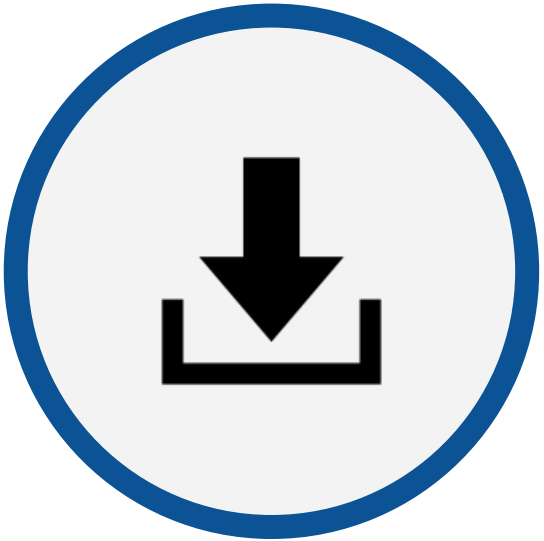
of users and their demands



- End user
- Decision-maker
- Non EO expert
- (EO) expert user
- Intermediate user
- Policy maker
- Geospatial data user

*PwC for European Commission (2019): Copernicus market report

Reproducibility challenge - DATA ACCESSIBILITY



Access

- different data are accessible via different data access systems
- it is still about downloading data
- community-specific data formats (GRIB, NetCDF, GeoTiff)
- data structure and complexity (analyses vs forecast, multiple dimensions)

WEkEO by COPERNICUS

Google Cloud Platform

Google Earth Engine

amazon web services

Climate Change Service

climate.copernicus.eu

Cloud-based data systems -
not only a technical challenge...

User requirements survey

Design
Promotional channels
Time frame
Number of responses

- Online questionnaire
- Open from 12 Nov 2018 to 30 Jan 2019 and 11 April to 31 May 2019
- Promotion via
  - Twitter (with support of CopernicusEU, CopernicusECMWF and Group on Earth Observations)
  - Geospatial mailing lists, such as CODATA, OGC
  - Geospatial communities, such as EGU/AGU ESSI
  - Copernicus C3S and CAMS newsletter
  - Article at Geoawesomeness
  - Medium blog article
  - LinkedIn
  - Individual contact of subject-matter experts
  - ECMWF commercial customers
- 231 respondents

Six topical sections

1	**Personal information**
2	**Information about work**
3	**Current data use**
4	**Data handling**
5	**Data challenges**
6	**Future data services**

2

- Work sector (industry, research, etc.)
- Differentiation between data user / data provider

3

- What kind of data is currently used and would like to be used in the future
- Data formats
- What data applications are of interest?

4

- How is data analysed?
- How are large volumes of data accessed?
- Satisfaction of current data access system
- How is data processed?

5

- Data volume, complex data formats, too many data platforms, data discovery, etc.

6

- Motivation to migrate processing tasks to the cloud?
- Legal policy of a cloud service
- Use of cloud services
- Security aspects of cloud services
- Willingness to pay for cloud services

How is data currently processed and analysed?

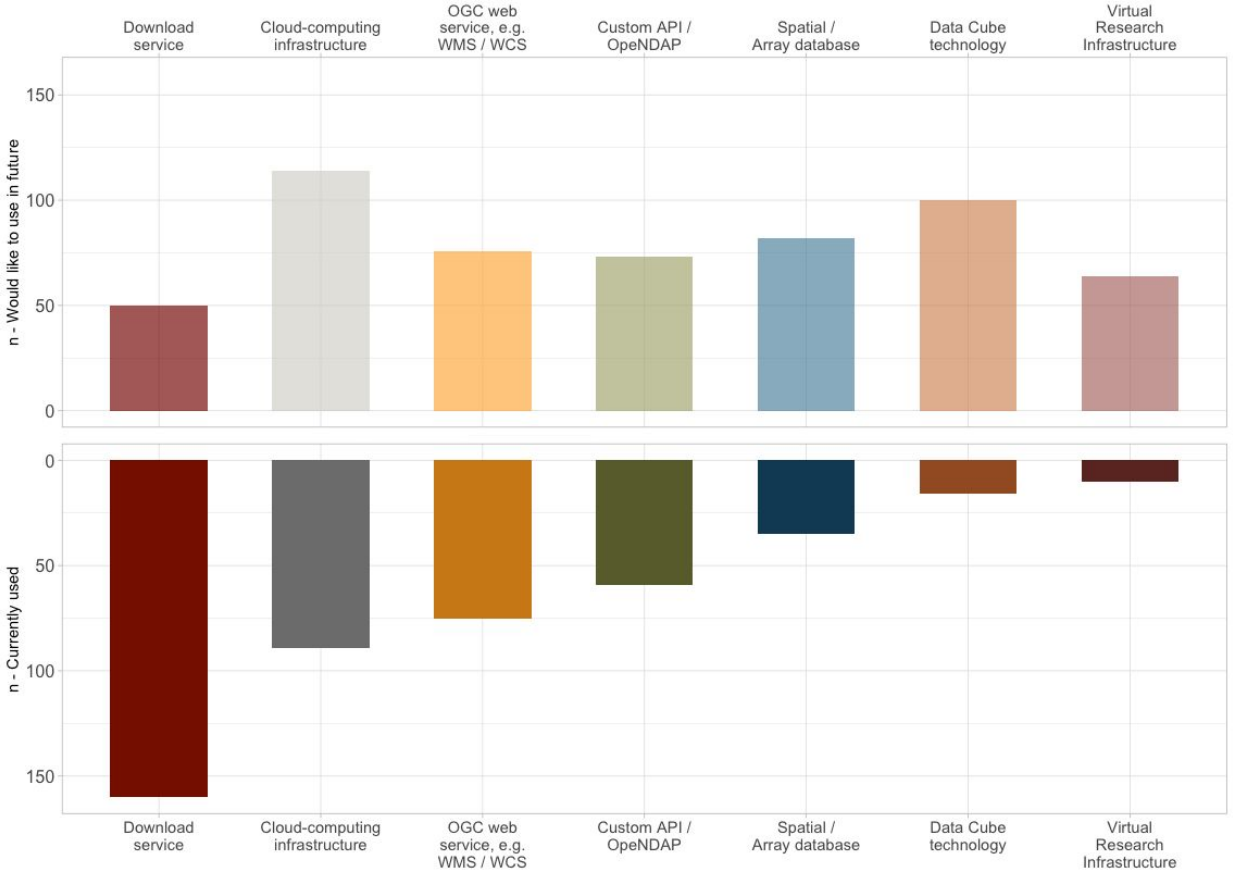
	Always		Sometimes		Never	
Way of data processing and analysis	**n**	**% (n=231)**	**n**	**% (n=231)**	**n**	**% (n=231)**
Code-based processing routines on a local machine	125	54.1	74	32.0	17	7.4
Geospatial software on a local machine, e.g. QGIS	55	23.8	110	47.6	48	20.8
Code-based processing routines accessing cloud services	20	8.7	74	32.0	117	50.6
Code editor in the cloud, e.g. CDS Toolbox or Google Earth Engine editor	17	7.4	86	37.2	108	46.8

> 86% process sometimes or always data with a **code-based processing routines on a local machine**

Top 3 programming languages:

- Python (77.1%)
- R (44.2%)
- gdal (35.1%)

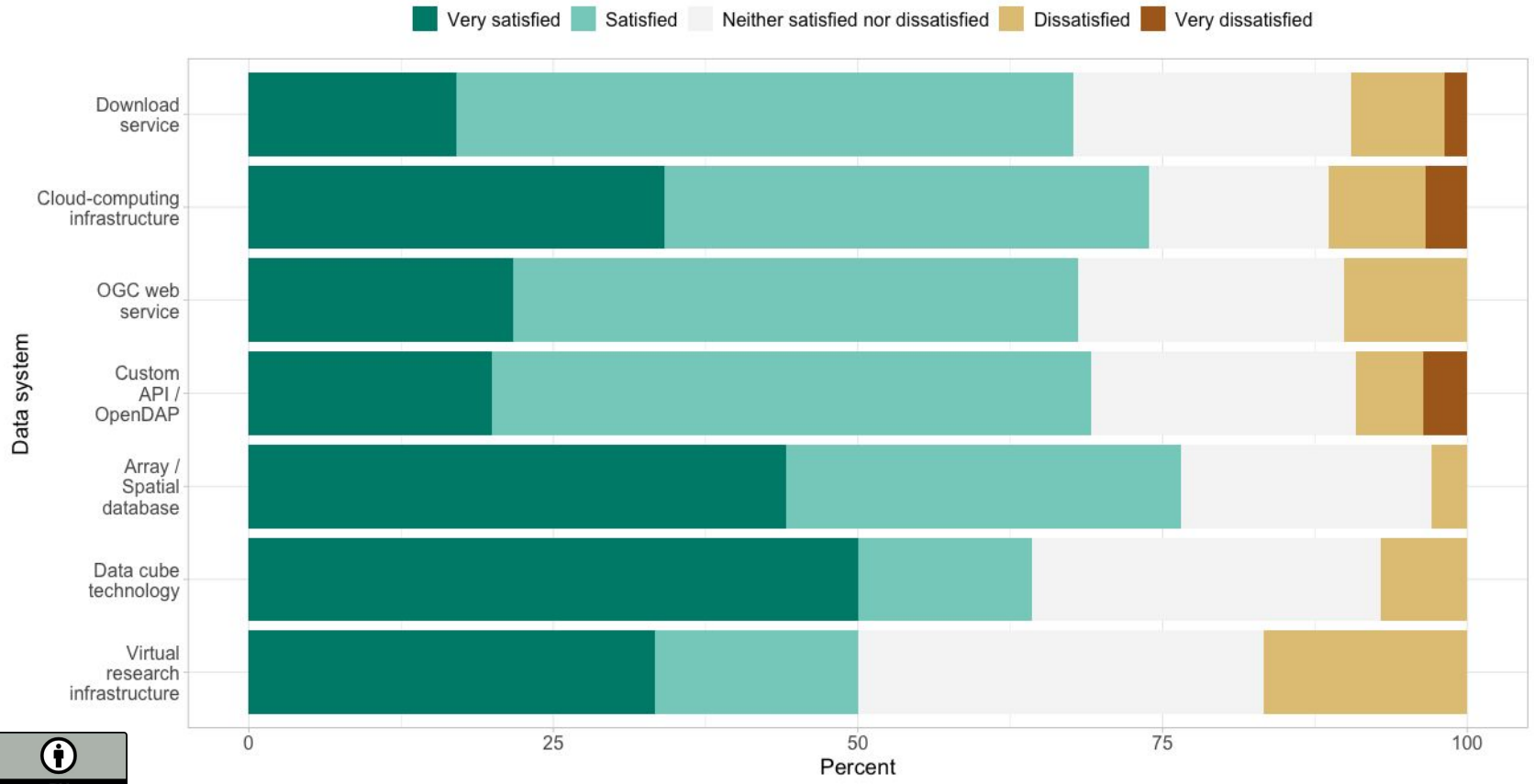
Current and future use of data services



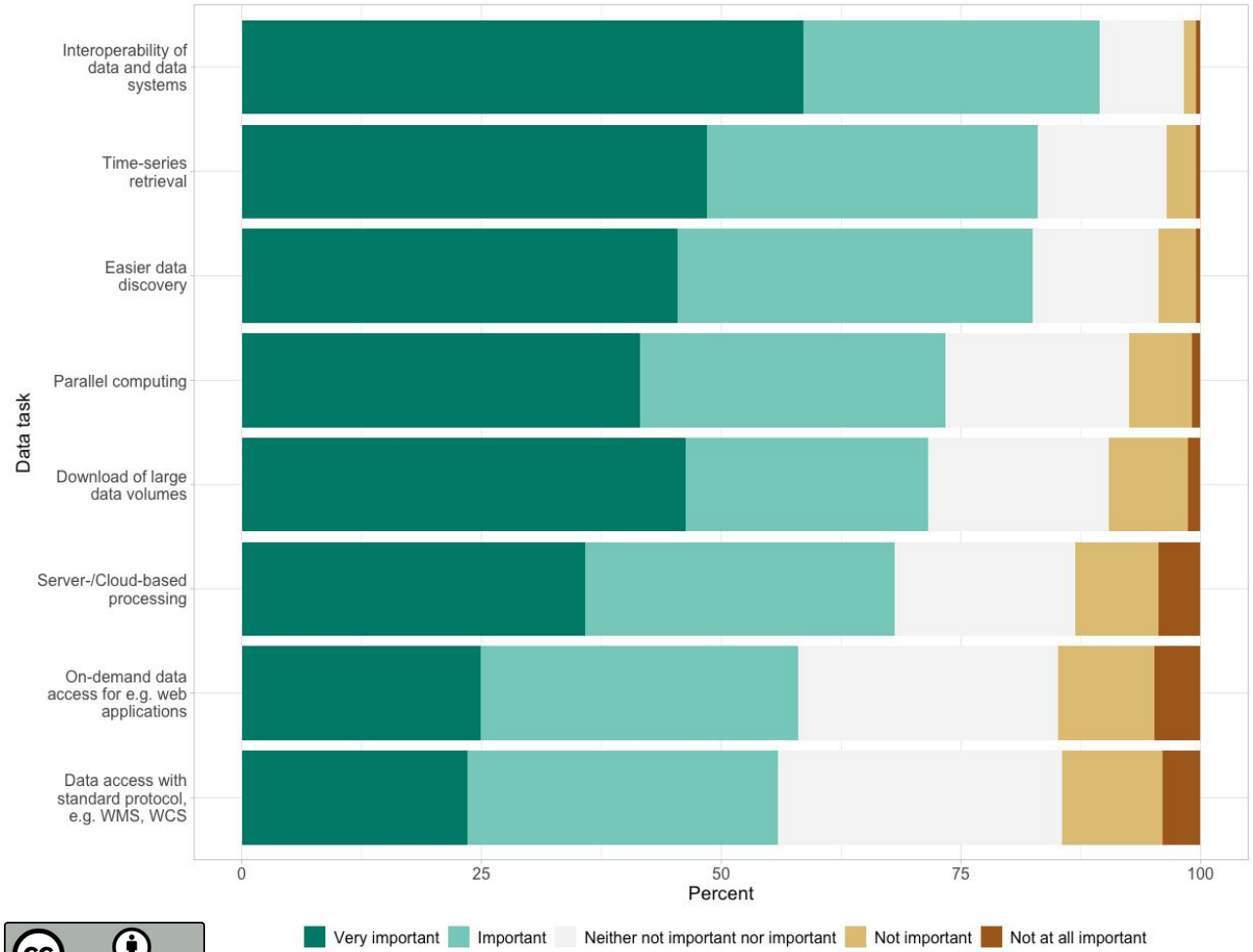
High interest in using **cloud** and **download services** in the future

	Ratio Future use vs. no interest at all
Download service	2.5
Cloud-computing infrastructure	4.2
Others	< 1

Overall high satisfaction level of current data access systems - Around 70% are (very) satisfied with the data access system they use



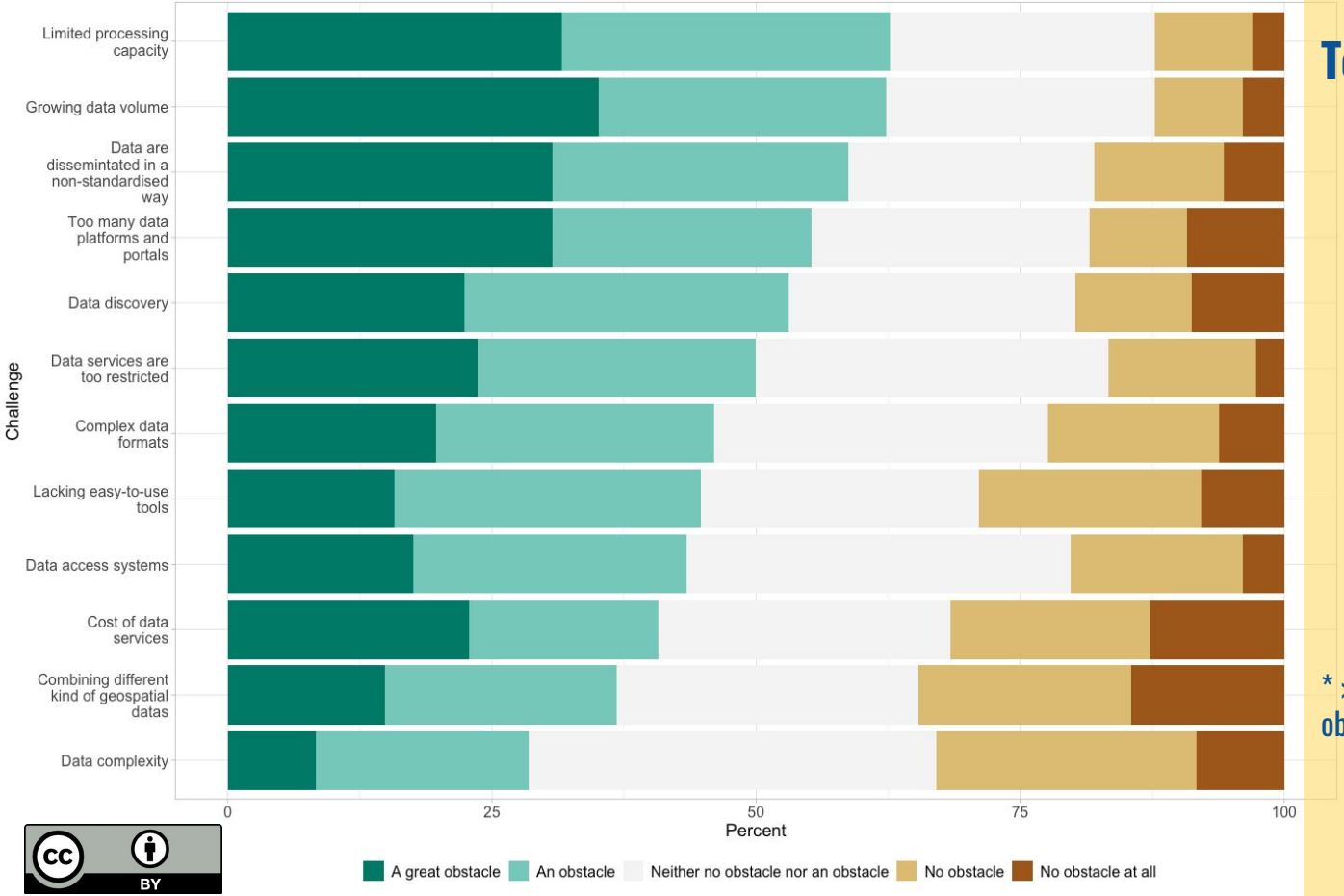
Importance of data tasks



Combination of different data sources and **interoperability of data and data systems** are considered as (very) important

Users **do not draw a direct link between interoperability of data and data systems** and **standardised data access**

Data challenges

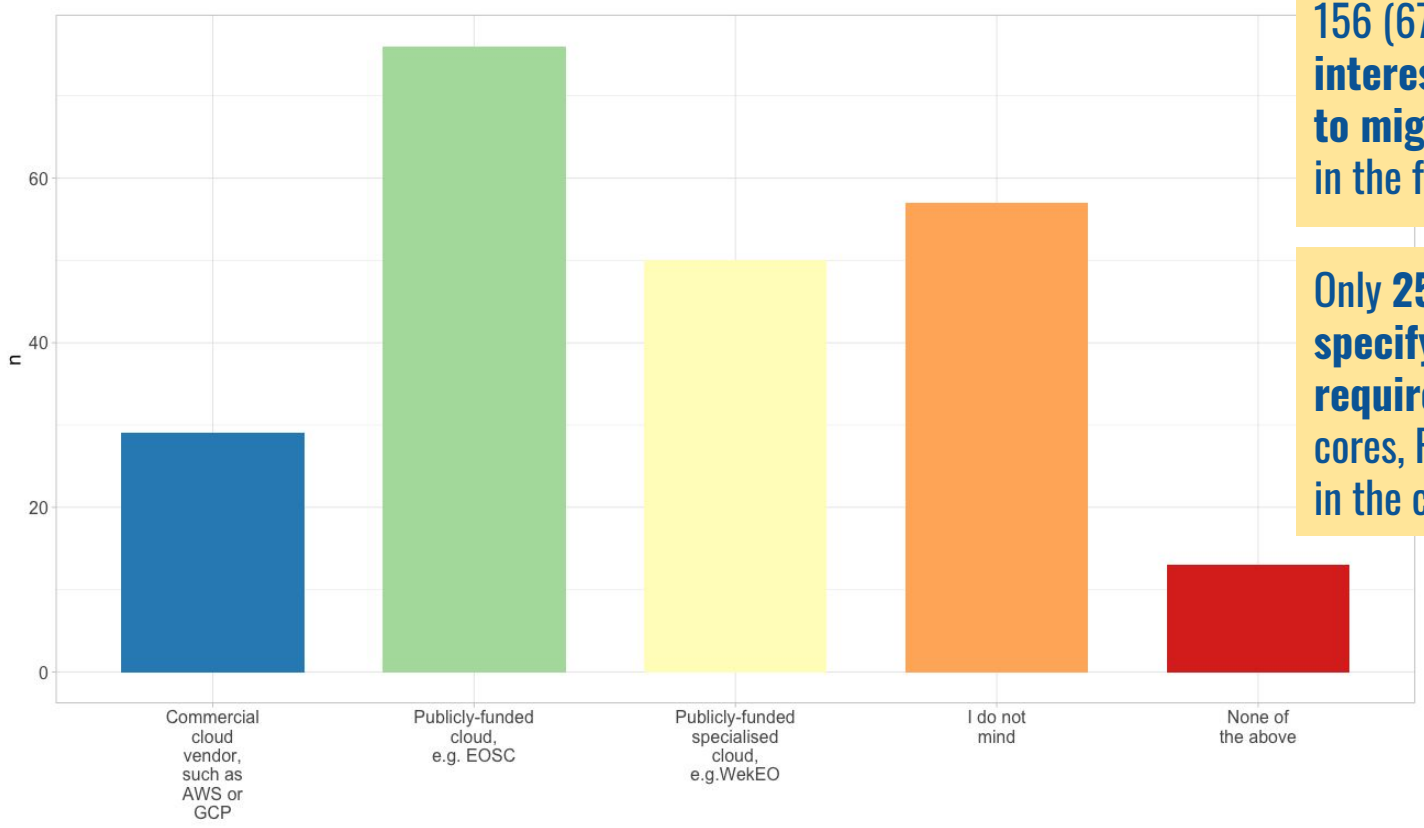


Top 5 data challenges*:

- Limited processing capacity
- Growing data volume
- Data are disseminated in a non-standardised way
- Too many data platforms and portals
- Data discovery

* > 50% of respondents rated it as an obstacle or great obstacle

Preferred legal policy of cloud services



156 (67.5%) are either **interested or very interested to migrate to cloud services** in the future

Only **25% would be able to specify technical requirements** (number of cores, RAM, etc.) for their tasks in the cloud

Conclusions

Despite the high interest in using cloud-based services, **many users face technical hurdles** in using them.

Users are not familiar with working with cloud-based systems. **A shift will require a change in mindset and time.**

Combining different types of data is one of the most important tasks users do. Current cloud solutions do not facilitate this need and **face the problem of data interoperability**

Thank you!

Further information:
Results of the entire study will be submitted shortly in form of two research articles.

Julia Wagemann

PhD candidate at University of Marburg
Visiting Scientist at ECMWF

@JuliaWagemann