ECMWF Workshop: Weather and Climate in the cloud | 8-10 February 2021

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

Towards a better understanding of user requirements

Julia Wagemann[1,2], Stephan Siemen[2], Bernhard Seeger[3], Jörg Bendix[1]

[1] Laboratory for Climatology and Remote Sensing, Philipps University Marburg
[2] ECMWF
[3] Department of Mathematics and Computer Science, Philipps University Marburg

Twitter: @JuliaWagemann

ECMWF Strategy 2021-2030 - A users' perspective

Cloud computing

Artificial Intelligence
Machine Learning

Open Data

Paradigm shift

New user
requirements

Diversification
of users

The need to better specify *'users'* of Big Earth Data

Big Earth Data systems are developed for 'users', but users are diverse

Term 'user' is broadly applied, but users differ in their domain as well as data and skills literacy

No clear definition of Big Earth Data value chain and stakeholders involved

The need to categorize (cloud-based) systems - An 'attempt'

CDS Toolbox

openEO

European Open Science Cloud

Amazon Web Services

Data cubes

Google Earth Engine

Community cloud

Cloud-native

Copernicus Data and Information Access Service (DIAS)

Analytics Platform

Climate Data Store

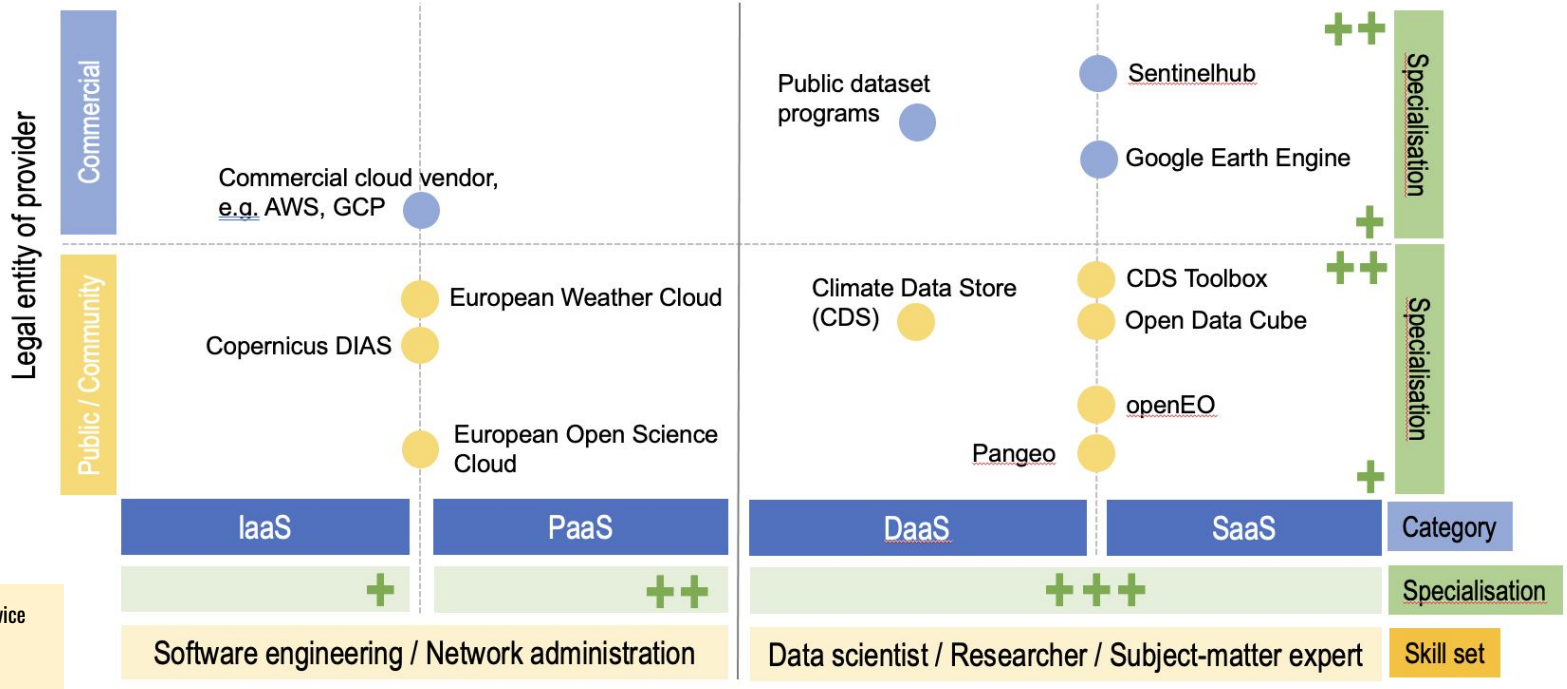
Pangeo

Google Cloud Platform

European Weather Cloud

... and many more

The need to categorize (cloud-based) systems - An 'attempt'



IaaS - Infrastructure-as-a-Service
PaaS - Platform-as-a-Service,
DaaS - Data-as-a-Software,
SaaS - Software-as-a-Service

Note: Graphic does not aim to present a full picture of the landscape of cloud systems for Big Earth Data, but rather provides a categorisation framework

Survey: User requirements of Big Earth Data

When:
- Nov 2018 - Jan 2019
- Apr - May 2019

Six categories
- 32 questions

1) Personal information

2) Work information

- 231 respondents
- majority from Europe and USA / Canada
- 70% between 30-50 years
- around half indicated to work at University, followed by Government and Established Company

3) Data use

4) Data handling

5) Data challenges

Analysis of the current state

Wagemann et al. (2021): Users of Open Big Earth Data - An analysis of the current state. *(under review)*

6) Future data services

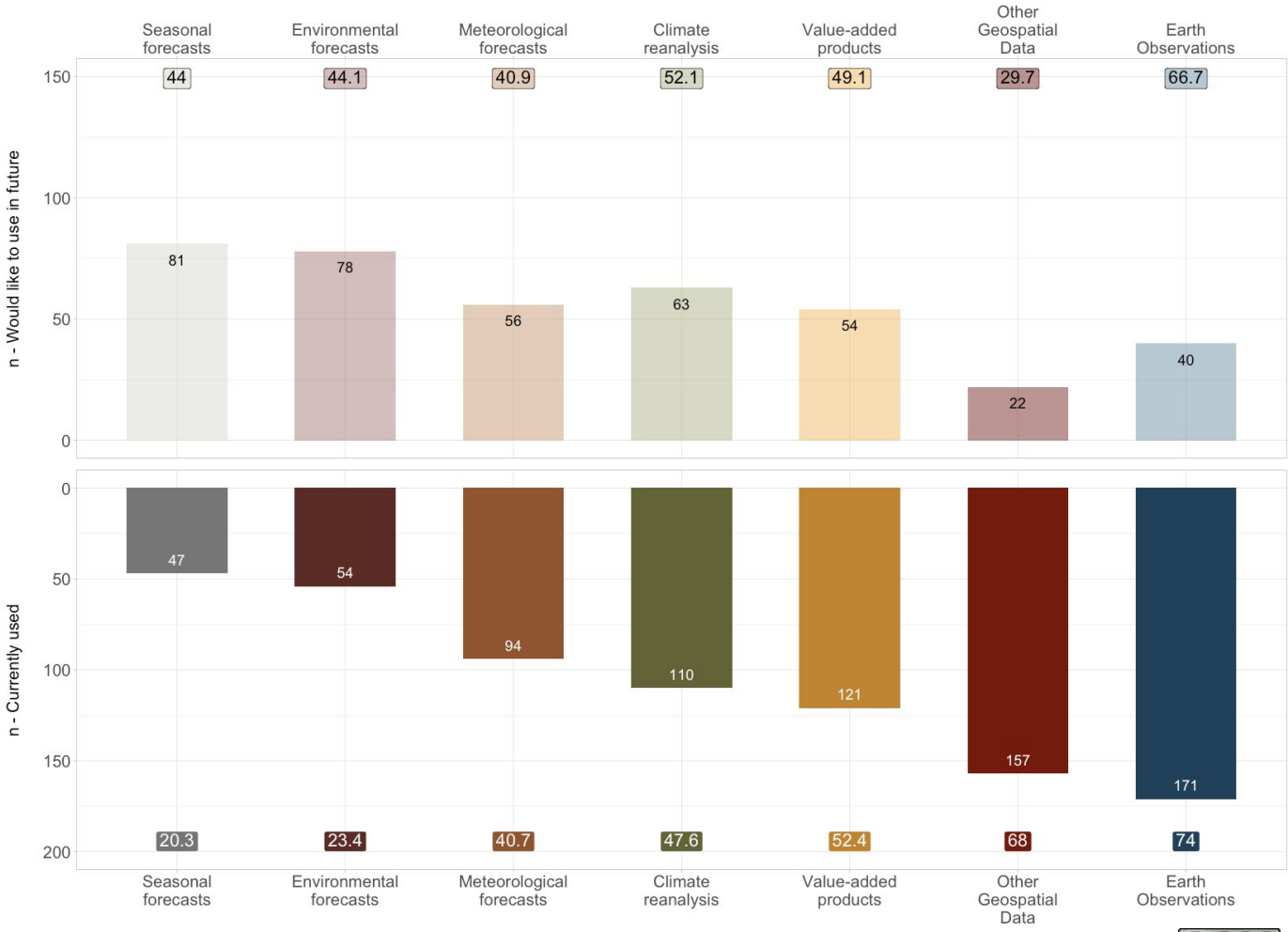
Future requirements

Wagemann et al. (2021): A user perspective on future cloud-based services for Big Earth Data (in preparation)

Forecast data currently used least, but interest for future use

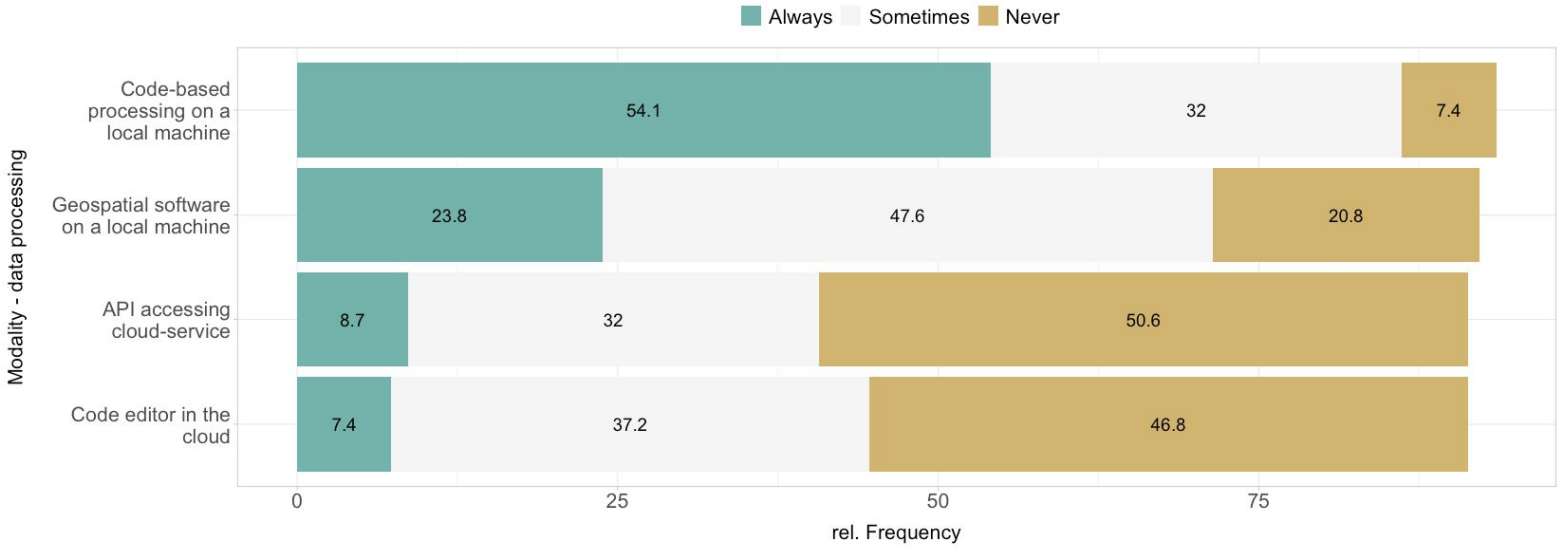
Continued interest in Earth Observation and climate reanalysis

Current and Future Use



Code-based processing on a local machine is prevailing modality

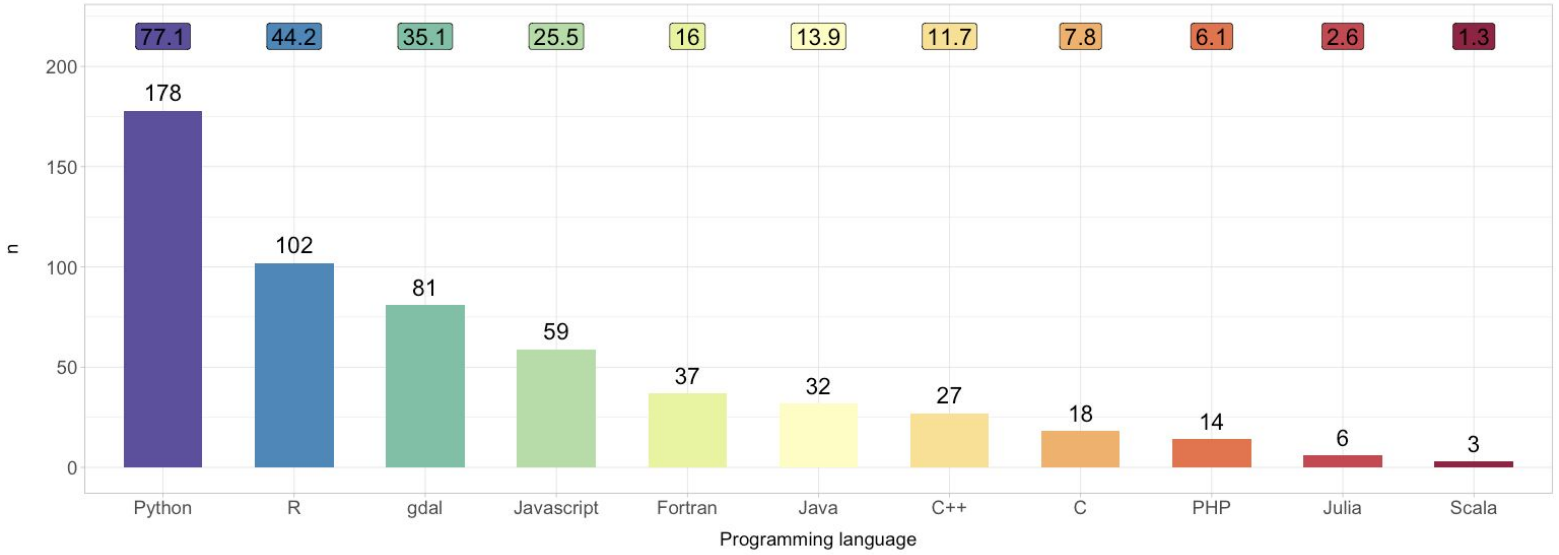
2 out of 3 use additionally desktop-based software



Data handling modality

Python and R - most used programming languages

Python is preference for meteorological and climate data twice as much as R



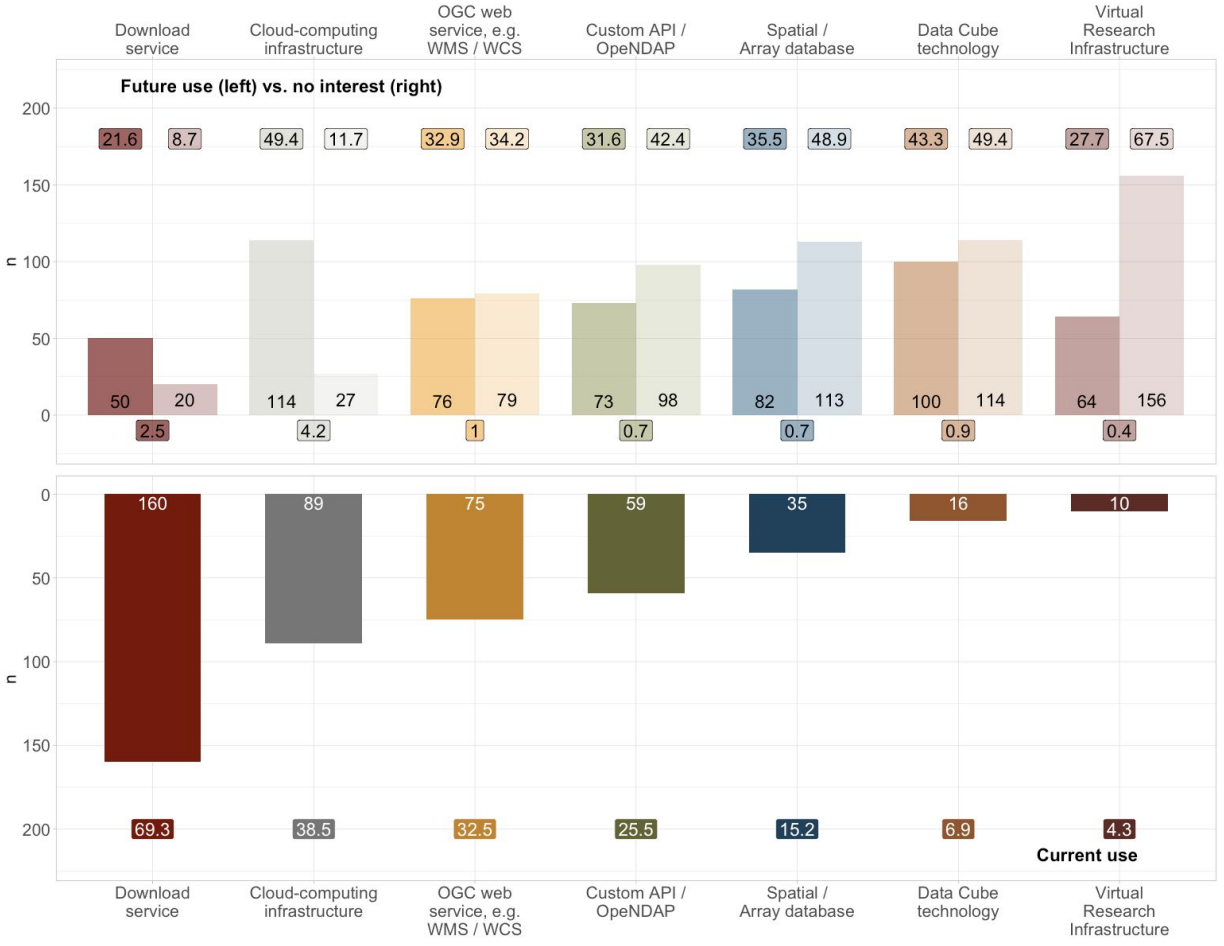
Data handling modality

Download service is prevailing mode of data access

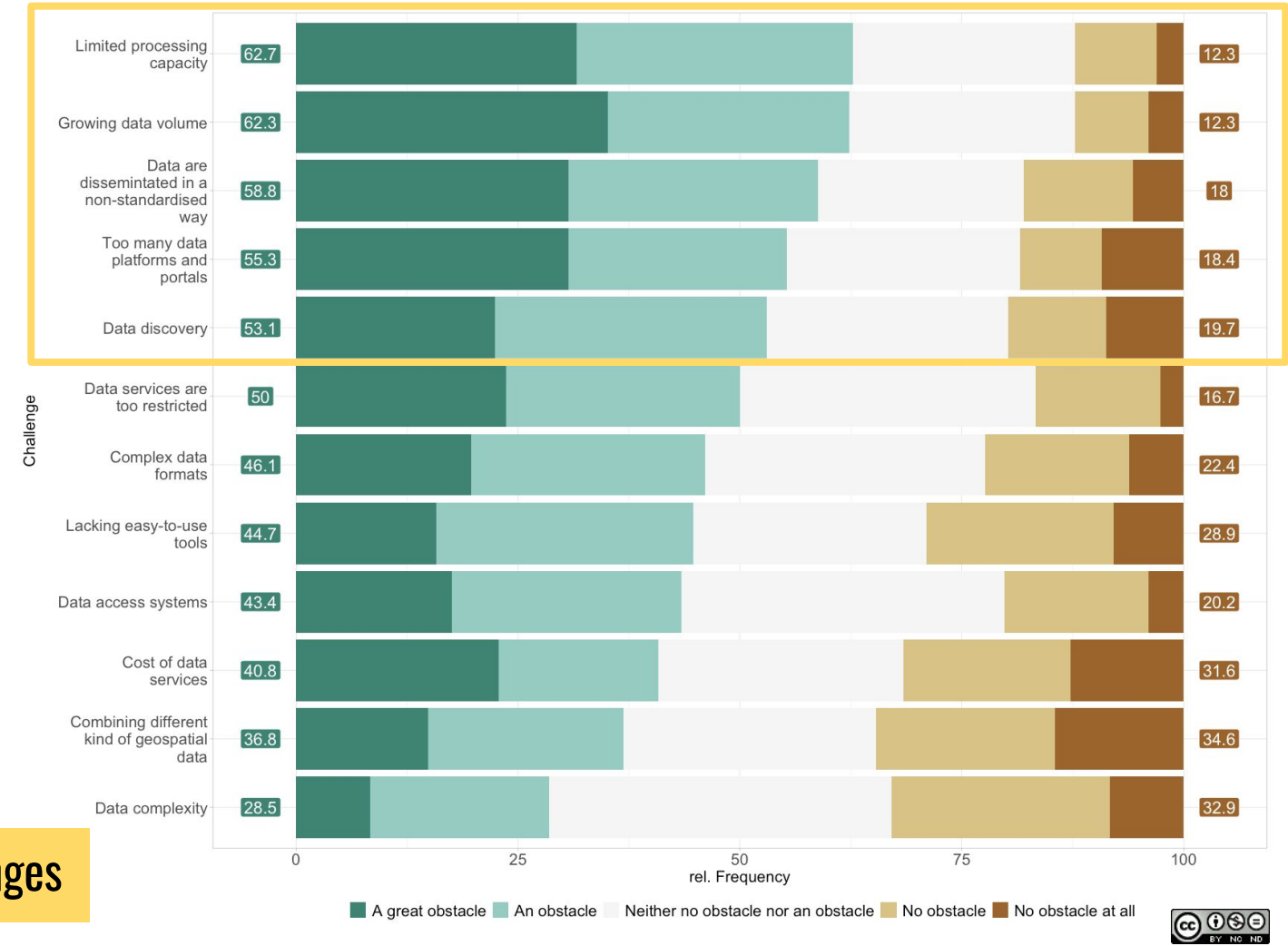
Ratio between ‘future use’ and ‘no interest’ of importance

Overall high satisfaction rate - more than 60% are either satisfied or very satisfied

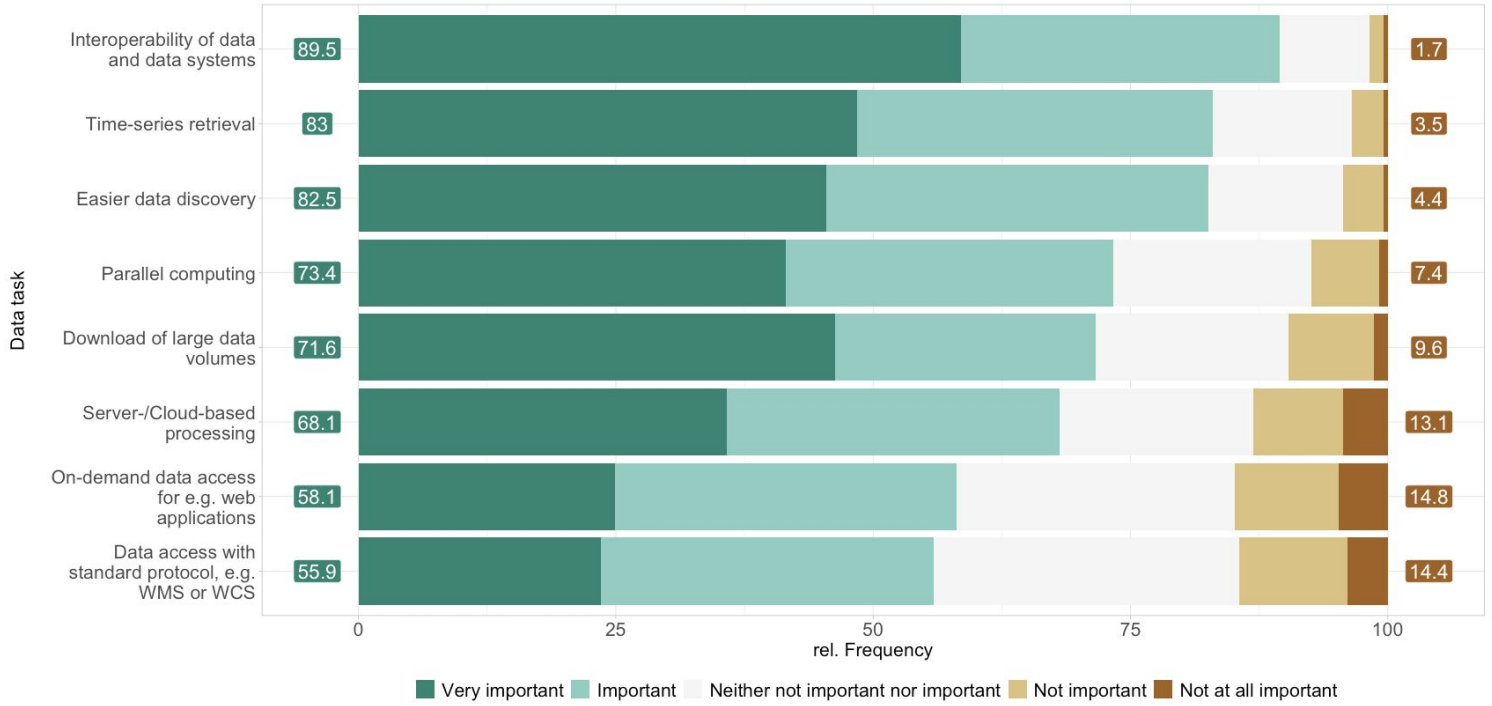
Data access - Current and Future



Top 5 challenges are related to ‘finding’, ‘accessing’ and ‘interoperating’ Big Earth Data



Big Earth Data challenges



Importance of data analytics aspects

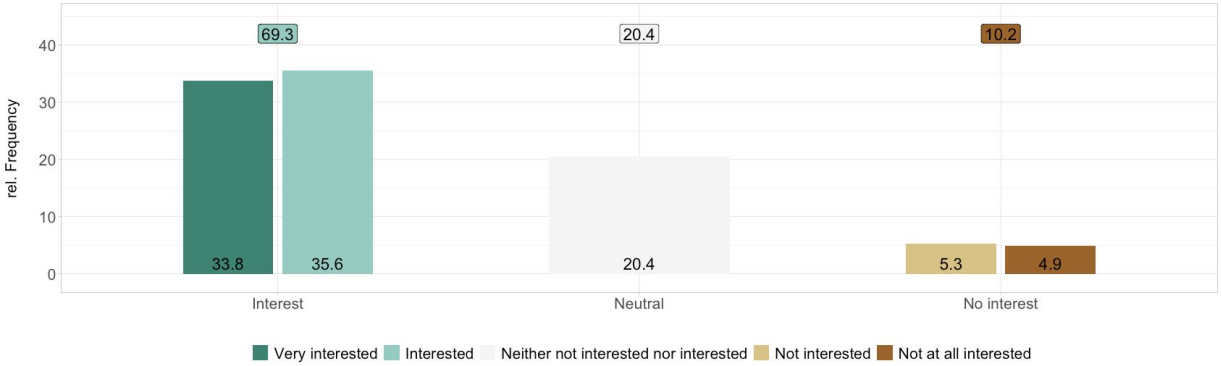
Interoperability of data vs. data access with standard protocols, e.g. WMS / WCS

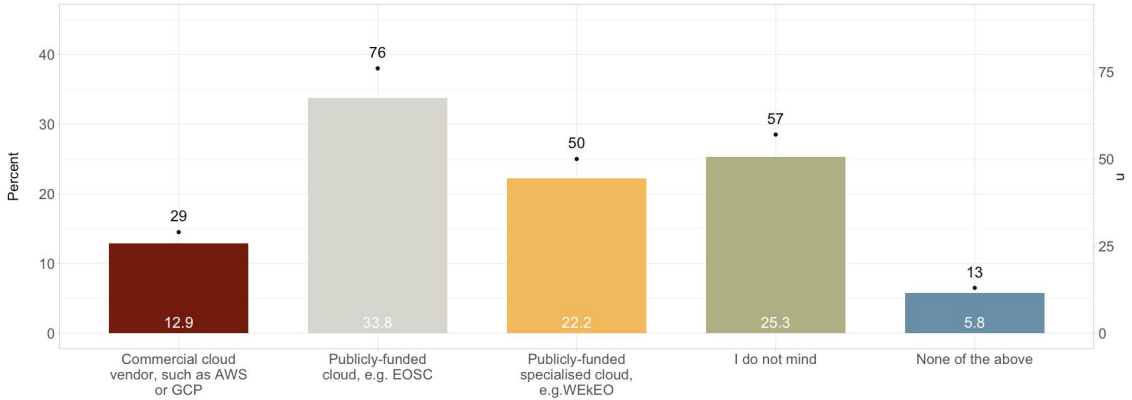
70% consider *'download of large data volumes'* as (very) important

Users perspective on future (cloud-based) services

Almost 70% indicate to be interested or very interested to migrate to cloud services

1 out of 4 are able to specify their technical requirements for storage and processing



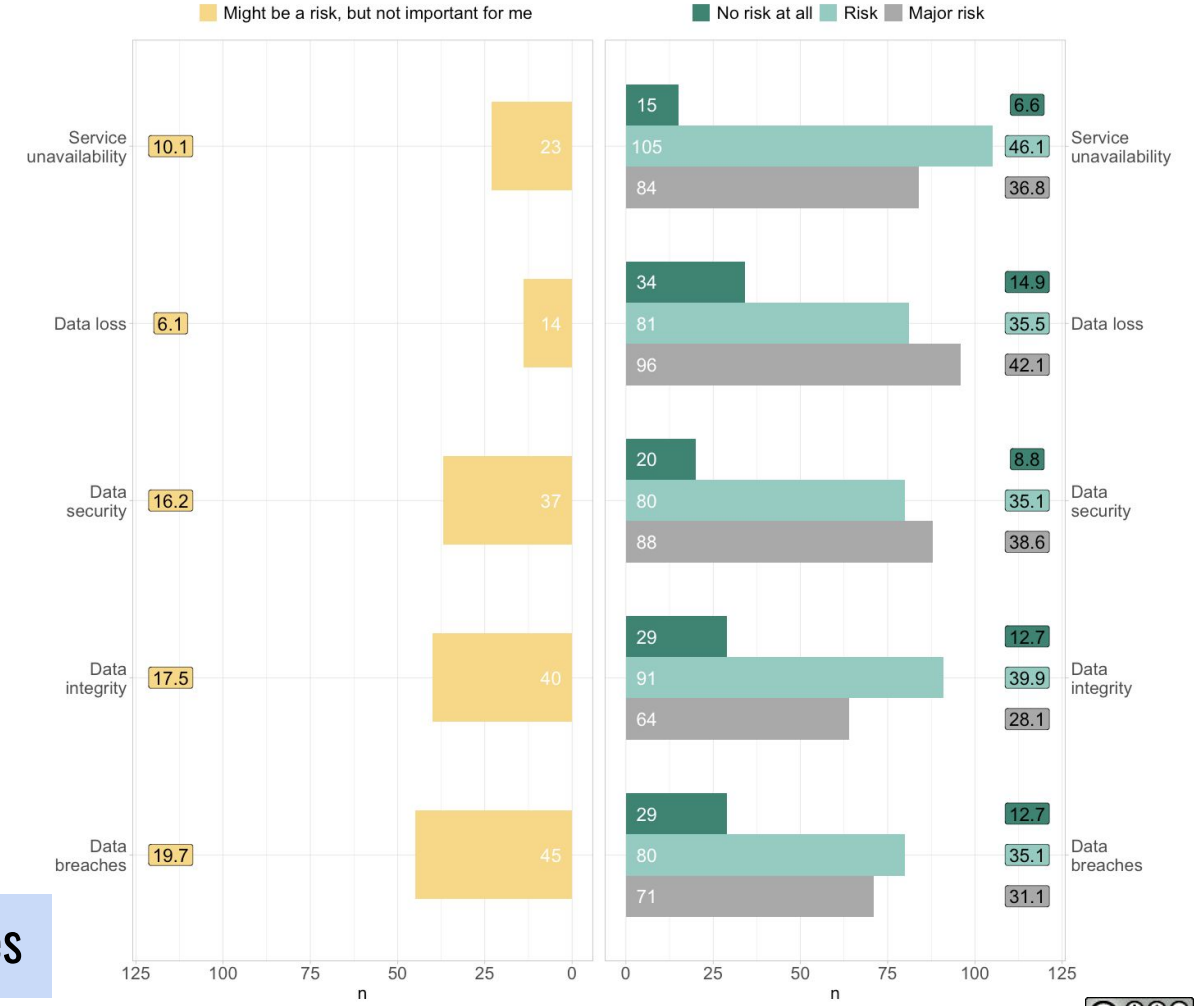


More than half prefer publicly funded cloud services (general or domain-specific clouds)

1 out of 4 ‘do not mind’ the legal policy

2 out of 3 rate all security aspects as risk or major risk

Other risks mentioned: vendor lock-in or migration to a different cloud provider



Security aspects of cloud services

Example data workflows

Run ML or forecast models

Analysis of long time-series information

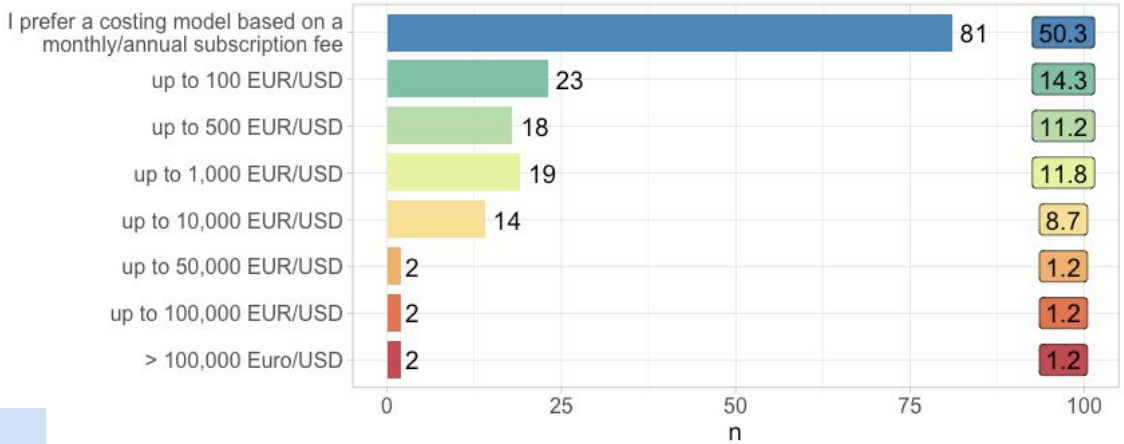
Shortening the processing time

Generating gridded (Level 3) climate products

Downscaling

50% make their willingness dependent on the cost of processing

Nearly 30% indicated to not be willing to pay for processing



Willingness to pay for cloud services

ECMWF Strategy 2021-2030 - A users' perspective

Cloud computing

Artificial Intelligence
Machine Learning

Open Data

Paradigm shift

New user
requirements

Diversification
of users

Summary: Current State - Are Big Earth Data FAIR?

Findable

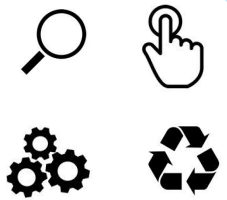
'Data discovery' and 'too many data platforms and portals' among top 5 challenges

75% rate 'easier data discovery' as (very) important

Accessible

Downloading data is prevailing mode of data access

'Limited processing capacity' and 'growing data volume' top 2 challenges



Interoperable

Importance to 'combine different data sources'

'Non-standardised dissemination of data' among top 3 challenges

Reusable

Reusability is limited when the first three principles are already challenging

Summary: Future requirements - How to bridge the gap?

General interest to use cloud services

Shortage in skills

Scepticism in cloud security and emerging costs

Building up **TRUST** through strengthening capacities

Data providers

Prioritise interoperability

Coordinated efforts to better define users and their needs

Follow community standards

Data users

Prepare (and be open) for change

Be literate in more than one programming language

Data trainers

Train the new generation of Big Earth Data users how we expect them to work in the future

ECMWF Philipps Universität Marburg

THANK YOU!