



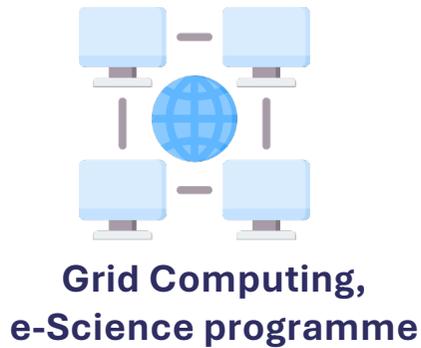
Governance for a distributed infrastructure to support climate science

Philip Kershaw, Head CEDA, RAL Space, STFC
NFCNetwork+ Meeting, 6 Mar 2025



Big Data driving a federated approach

Early 2000s



Communities
of best practice

Global Organization for Earth
System Science Portals



Research federations

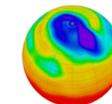


RAL Space



National Centre for
Atmospheric Science

NATURAL ENVIRONMENT RESEARCH COUNCIL



Centre for Environmental
Data Analysis

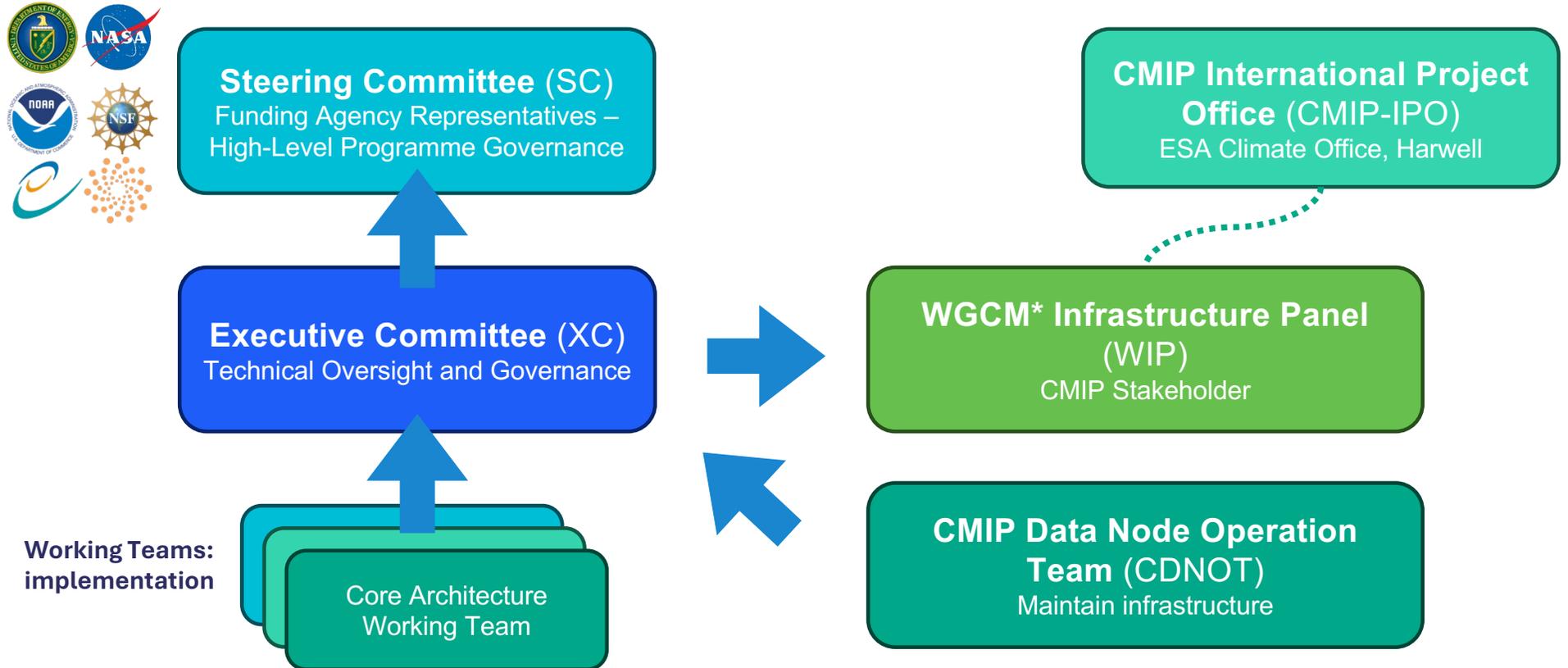
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

ESGF Objectives

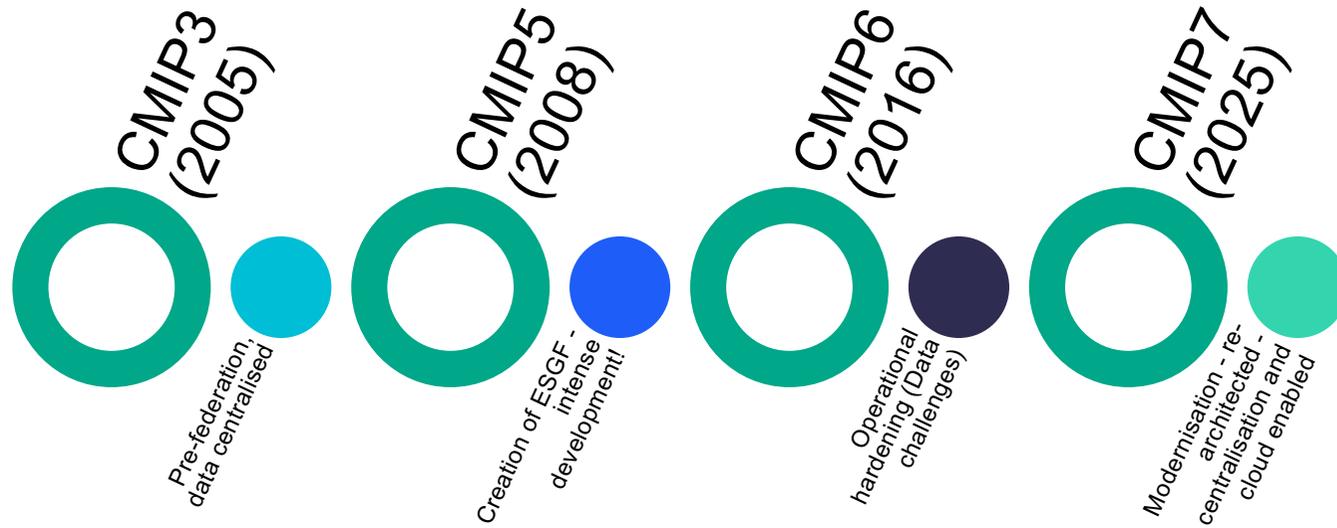
To provide access to climate data primarily *MIP (Model Intercomparison Project) activities. ESGF may support access to other Earth science data but the core community to support is *MIP.

Address the challenge of hosting and access to large volume climate datasets through the provision of a distributed system of collaborating data provider organisations linked together in a federation.

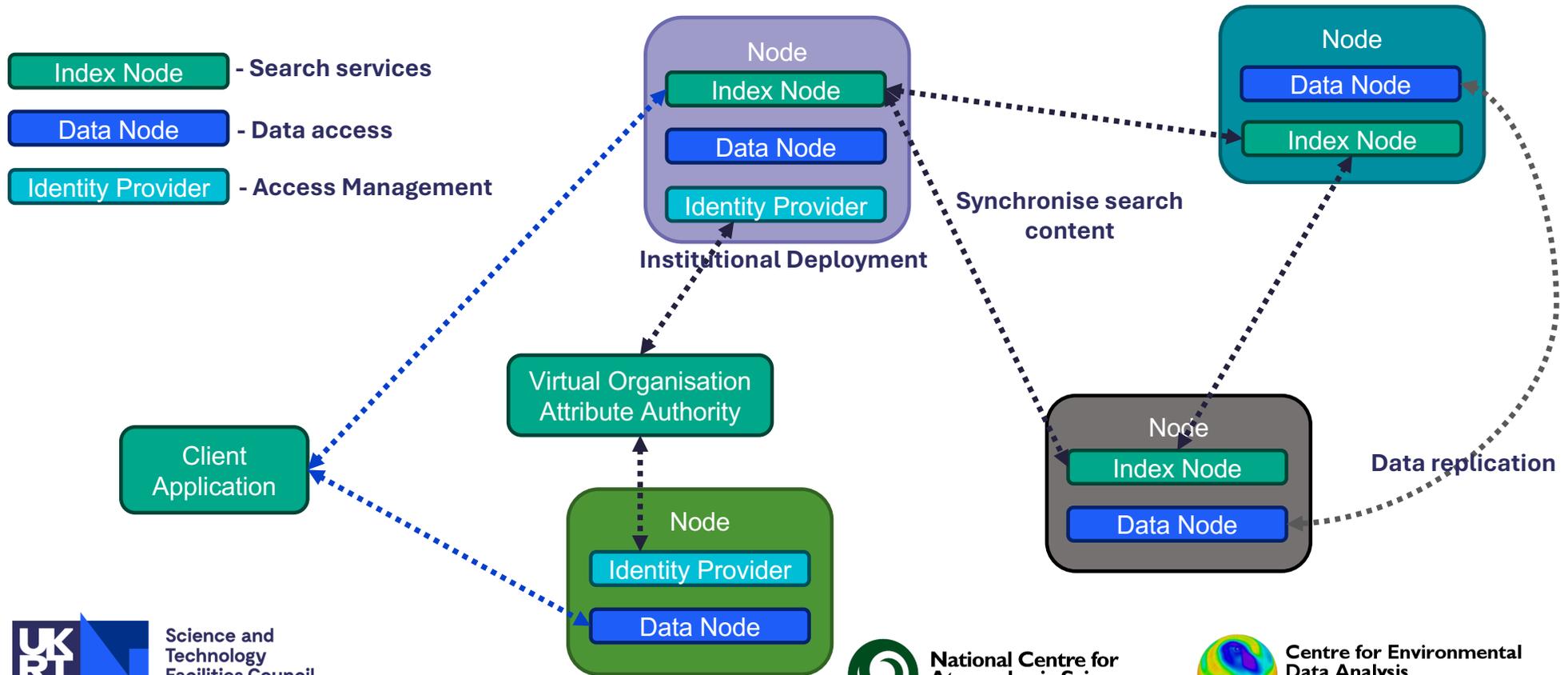
Governance Structure



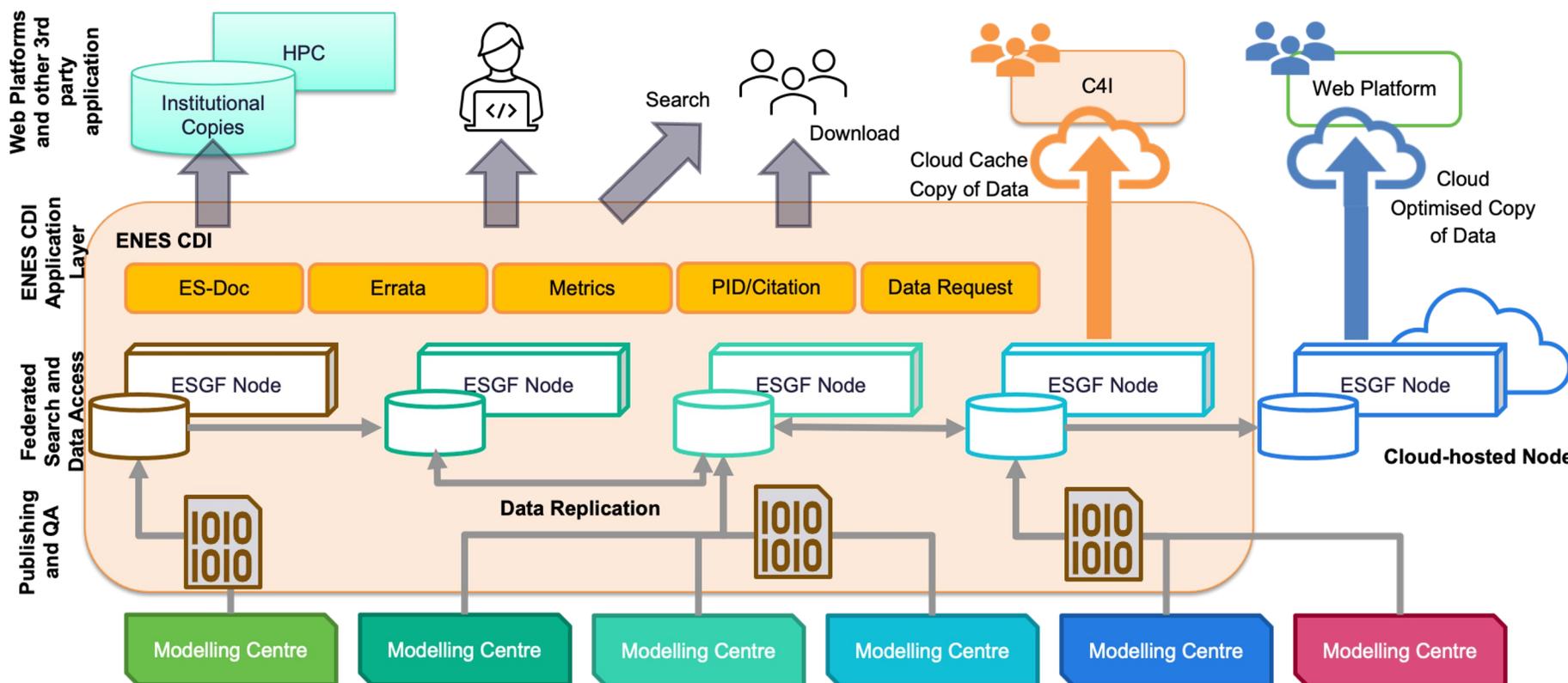
ESGF and CMIP history in one slide



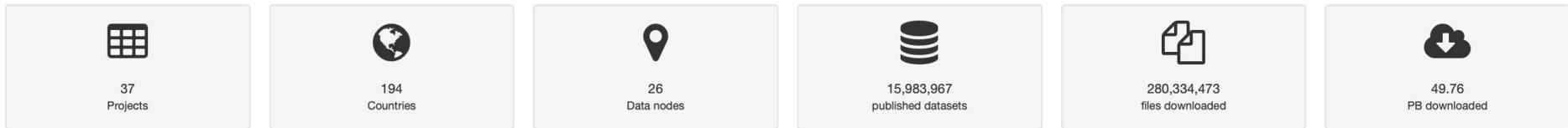
Original ESGF 1.0 High-Level Architecture Blueprint



ESGF as part of an eco-system



Successes

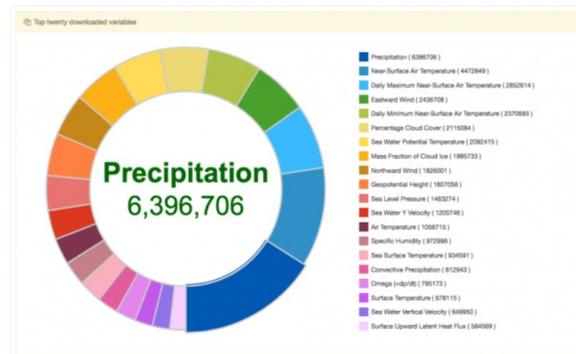


ESGF Federation



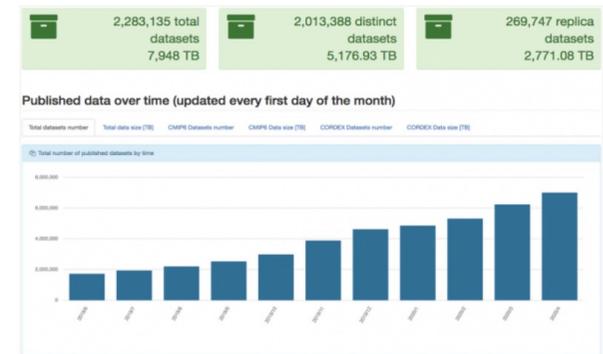
This user interface provides a set of data usage and publication metrics across the Earth System Grid Federation. Statistics refer to the period January 2018 to present.

Data usage



Cross-project and project-specific sections, with a rich set of charts and tables, provide different views about the data downloaded across the ESGF federation.

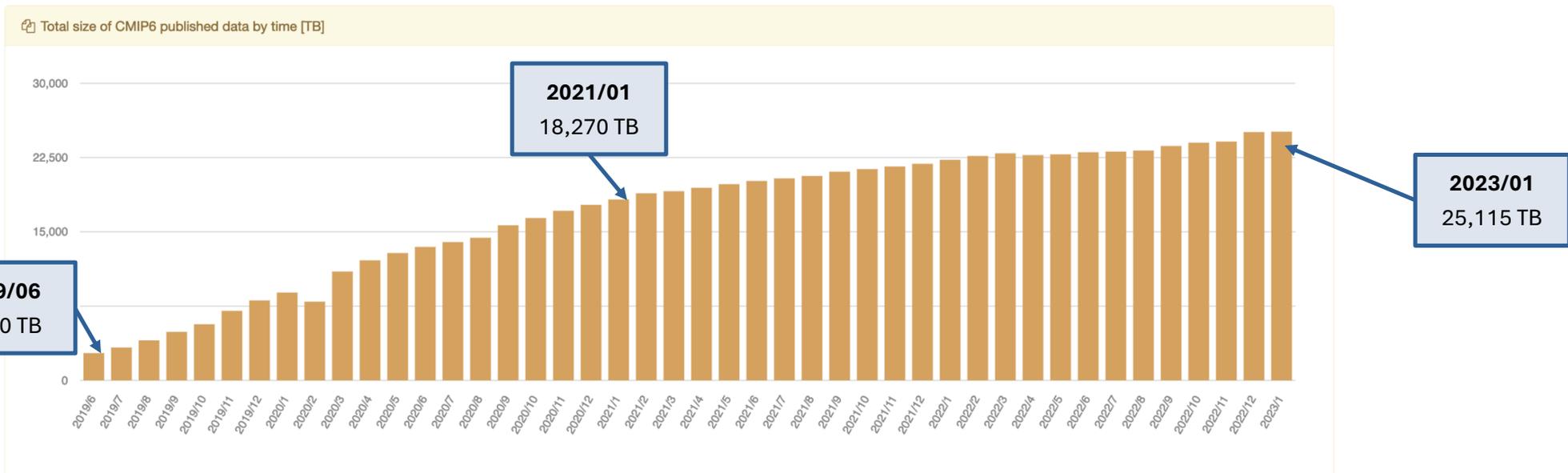
Data publication



A view of the total amount of data published and available through the ESGF infrastructure gives users an in-depth view about the ESGF data archive.

Successes – CMIP6 in Figures

 <p>CMIP6</p>	<p>13,606,408 total datasets</p> <p>25,212.14 TB</p>	 <p>CMIP6</p>	<p>6,759,621 distinct datasets</p> <p>14,295.37 TB</p>	 <p>CMIP6</p>	<p>6,846,787 replica datasets</p> <p>10,916.77 TB</p>	<ul style="list-style-type: none"> - 44 institutions - 120 models
---	--	---	--	---	---	---



Issues

Keeping the software contract – local fixes to federation-wide problems

Data publishing

Complex solutions missing the user need

Esp. IAM solution

Understanding the needs of an operational living system

Clarity about what each partners can provide

Investing in ongoing development

Managing disagreements

Funding seesaw between US and European funding lines

Lessons Learnt

A clear common goal and deadline galvanised the partners towards creating a solution – a SMART* target

Set up clear governance and responsibilities from the outset

Firm, enforced standards with data worked for interoperability

Agree service levels for partners so that there are clear expectations

Development and Operational needs to live together in harmony 😊

Implement simple and clear governance for technical changes to avoid incompatibilities

Be user-focussed

* Specific, Measurable, Achievable, Realistic, Timebound

Summary

15 years of federation is an achievement to be proud of!

It has continued on the basis of goodwill without legal agreements or formalisms that might be expected for such an undertaking

There is a strong sense of a common purpose and desire to collaborate towards a common goal

