

International Data Commons

The future of data exploitation
in ~~30~~ 10 years' time

ICRI 2018, Vienna



Today's scenario

- Today's cross-RI and cross-border access is difficult due to **different funding models, access and provisioning policies**
 - Data and service provisioning to international user communities possible only when supported by sound business models or existing collaboration agreements. Today only a few structured int. research groups have achieved this.
- **Needs of large investments for the creation, processing, preservation, access and reuse of research data** → will the funding match the anticipated needs of future data-intensive science?
 - Opportunities for economies of scale and aggregation of demand can arise with joint provisioning of infrastructure common components
- Major **separation** between **data preservation and data exploitation** infrastructures in many disciplines
 - Ris and e-Infrastructures should collaborate to support the entire research workflow of an experiment

The International Data Commons

A federation of research data, computing, applications and other open science resources, responding to the problem of **scalable access to research data** through a **new data provisioning service approach** that is complementary to the traditional data download model.

The Data Commons should....

Allow to **discover, access and analyze** major research datasets and information for third-party exploitation

Provide **access to the data & data products close to processing facilities while avoiding duplication** of local data storage & compute infrastructures across research performing organizations in Europe

The Data Commons should...

- Offer a **hybrid distributed compute** platform (HTC, HPC, cloud) and integrated **rich portfolio of scientific application tools** supporting self-service provisioning
- Offer tools for scalable data movement across data preservation infrastructures and **distributed interconnected network of “data hubs”**
 - Provide integrated capabilities for **publishing and sharing scientific outputs** from experiments to support open science
- Support **federated authentication and authorization** for use of existing personal credentials and easy to use access channels

The Data Commons require...

- **Coordinated development, operations, procurement, management and provisioning** across RIs and e-Infrastructures sharing expertise and technologies of common interest
 - The **federation of existing/future publicly funded digital infrastructures** leveraging **national investments** to achieve economies of scale across Europe
- **Changes in the policies** of the participating infrastructures to support both national and international user groups → **Infrastructures open by default**
- **International cooperation** among digital infrastructures in the world, to scale beyond Europe

The International Data Commons

A **partnership**

- Of Research Communities, Research Infrastructures for **data and application provisioning**
- Of e-Infrastructures and other providers of European and international relevance for **compute/networking/data management, PID and AAI** infrastructure provisioning (network of national Data Hubs)

The **European Open Science Cloud** initiative of the EC will be a concrete step forward towards this vision.