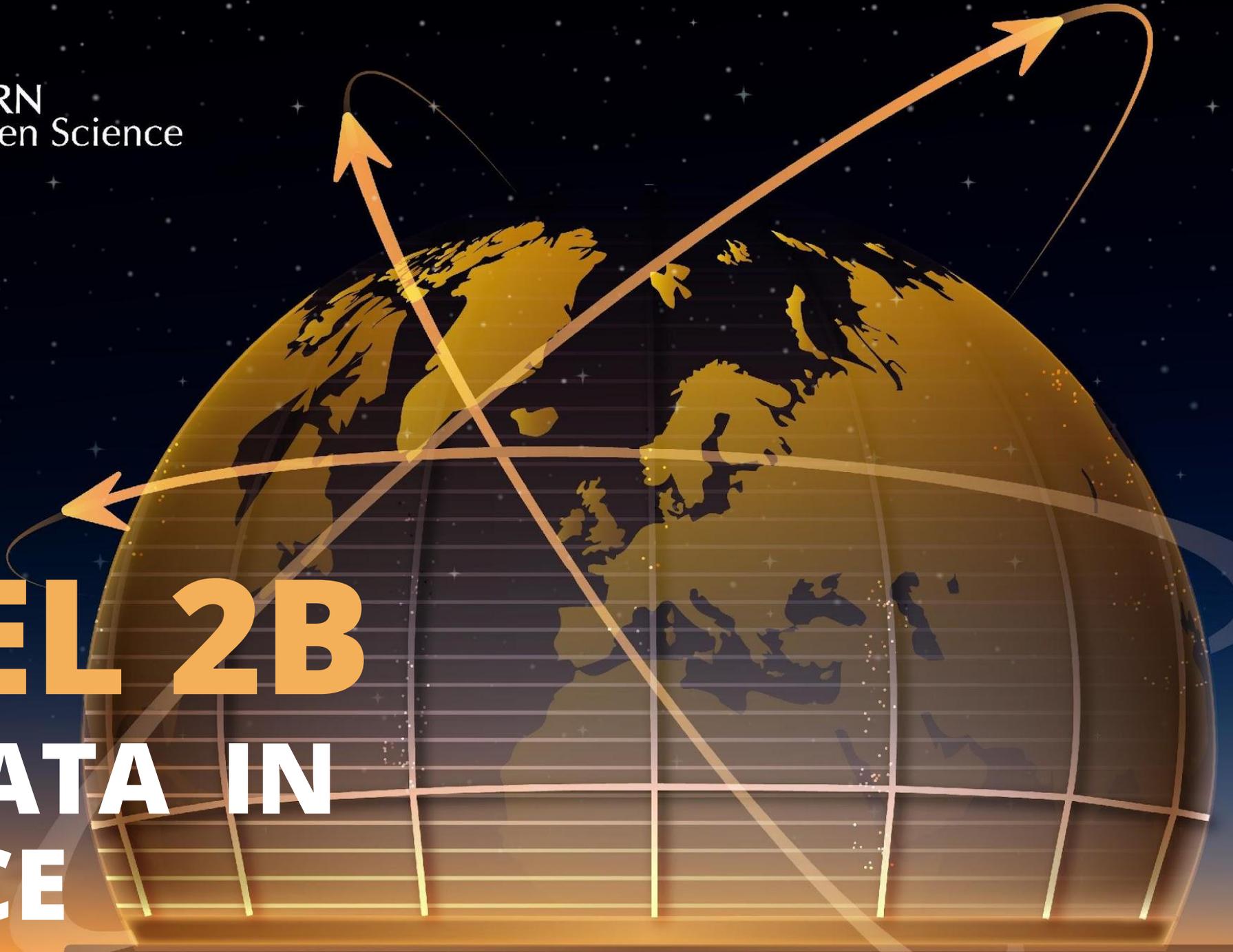




---

# PANEL 2B

## OPEN DATA IN PRACTICE





# PANEL 2B

## OPEN DATA IN PRACTICE



### PANEL CHAIRS

---

#### **Daniel Spichtinger**

Open Science Expert, self-employed &  
Ludwig Boltzmann Gesellschaft

### PANEL SPEAKERS

---

#### **Chelle Gentemann**

NASA, Transform to Open Science Lead

#### **Sarah Jones**

EOSC Association Board Director & GÉANT

#### **Kati Lassila-Perini**

CMS@CERN, Helsinki Institute of Physics

#### **Tibor Šimko**

CERN, Information Technology



# NASA Open Data

---



**Chelle Gentemann**

NASA, Transform to Open Science Lead



# Sharing our NASA Discoveries since 1958

National Aeronautics and  
Space Act of 1958,  
As Amended



The National Aeronautics  
and Space Administration

Section 203(a) of the law that  
created NASA notes...

## FUNCTIONS OF THE ADMINISTRATION

Sec. 203. (a) The Administration, in order to carry out the purpose of this Act, shall—

- (1) plan, direct, and conduct aeronautical and space activities;
- (2) arrange for participation by the scientific community in planning scientific measurements and observations to be made through use of aeronautical and space vehicles, and conduct or arrange for the conduct of such measurements and observations;
- (3) provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof;
- (4) seek and encourage, to the maximum extent possible, the fullest commercial use of space; and
- (5) encourage and provide for Federal Government use of commercially provided space services and hardware, consistent with the requirements of the Federal Government.<sup>10</sup>



# EARTHDATA

OPEN ACCESS FOR OPEN SCIENCE



End User Average  
Distribution Volume

**281.45  
Terabytes/Day**



End User Distribution  
Files Including  
from Cloud

**3 Billion**



Average Archive Growth

**49.15  
Terabytes/Day**



End User Distribution  
Files from Cloud Only

**290.03 Million**



Distinct Users of  
EOSDIS Data &  
Services  
(Google Analytics)

**3.64 Million**



Unique  
Datasets

**15,360**



Website Sessions  
(Google Analytics)

**2.28 Million**



Total Archive  
Volume Including  
in Cloud

**71.64  
Petabytes**



Total Archive  
Volume In Cloud  
Only

**20.2 Petabytes**

## SPD-41a is SMD's updated Scientific Information Policy

SPD-41a is *forward looking* and will apply to all future SMD-funded scientific activities

### Major Policy Updates

- Peer-reviewed publications are made openly available with no embargo period.
- Research data and software are shared at the time of publication or the end of the funding award.
- Mission data are released as soon as possible and unrestricted mission software is developed openly.
- Science workshops and meetings are held openly to enable broad participation.

## Infrastructure: Core Services

**BETA**

# SCIENCE DISCOVERY ENGINE

Empowering open science, the Science Discovery Engine allows you to explore the universe, from the tiniest of cells to the vastness of space, through discovery of NASA's science data, documentation, and code.

Search for...

The SDE is frequently updated. We welcome your [feedback!](#)

<http://science.data.nasa.gov/search>

ads

QUICK FIELD: Author First Author Abstract All Search Terms

← Start New Search mars atmosphere

Your search returned **22,436** results

Sort Date

Show highlights  Show abstracts  Hide Sidebars [Go To Bottom](#)

**AUTHORS**

- Jakosky, B 734
- Forget, F 613
- Montmessin, F547
- Smith, M 527
- Brain, D 473

**COLLECTIONS**

- astronomy 20.7k
- physics 10.1k
- general 500

1 2023EP&S...75...10O 2023/12  
**Cloud trains associated with Martian Mountain Lee Waves on the eastern side of the Phlegra Montes**  
Ogohara, Kazunori; Ro, Maaya

2 2023Icar...40415665K 2023/11  
**A note on the determination of vortex parameters from combined seismic and barometric measurements**  
Kurgansky, M. V.

3 2023P&SS...23405721V 2023/09  
**Effects of stealth CME to ion energisation at ionospheric altitudes of Mars**

### Science Explorer

Extend the primary digital library portal for researchers in astrophysics, planetary science & heliophysics, the Astrophysics Data System (ADS), to support Earth and Biological and Physical Sciences

**Over 90% of the primary planetary science literature is now included.**

# *Open is a misnomer*

---



**Sarah Jones**

EOSC Association Board  
Director & GÉANT





# RDM not Open or FAIR





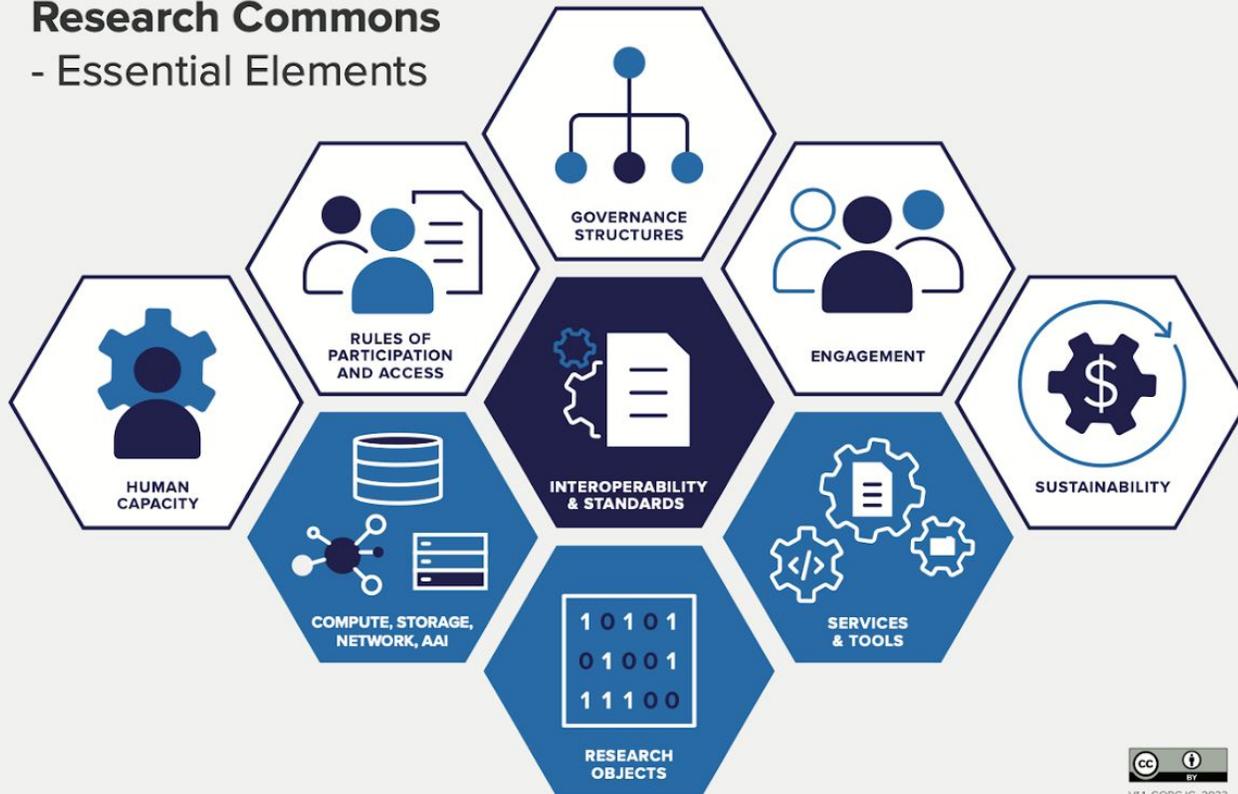
# Build the Commons



RESEARCH DATA ALLIANCE

- There are core elements we need whatever type of Commons we are building
- human and governance aspects (white)
  - technical and content elements (blue)
  - interoperability (dark blue,) “the core glue”

## Global Open Research Commons - Essential Elements





# Commons (or nodes) are everywhere



- *thematic, disciplinary, national, institutional, regional..*



Australian Research Data Commons



Oilthigh  
Ghlaschu





# Focus on interoperability



- *researchers need to be able to connect across Commons*



Oilthigh  
Ghlaschu



eosc



Australian Research Data Commons



# The problem is NOT Open Data it's interoperability

- We need an open API layer, series of protocols, standards, schemas, crosswalks... etc
- Allow data to be transferred (securely), combined, remixed and reused across all different environments researchers work in
- In Europe we call this the **EOSC Interoperability Framework** and we should be working on it globally!



# Open data, in practice

---



**Kati Lassila-Perini**

CMS@CERN, Helsinki Institute of Physics

CMS Data Preservation and Open Access  
group co-coordinator



# Open science - what it takes to make it happen



# Building Open Science Services

---



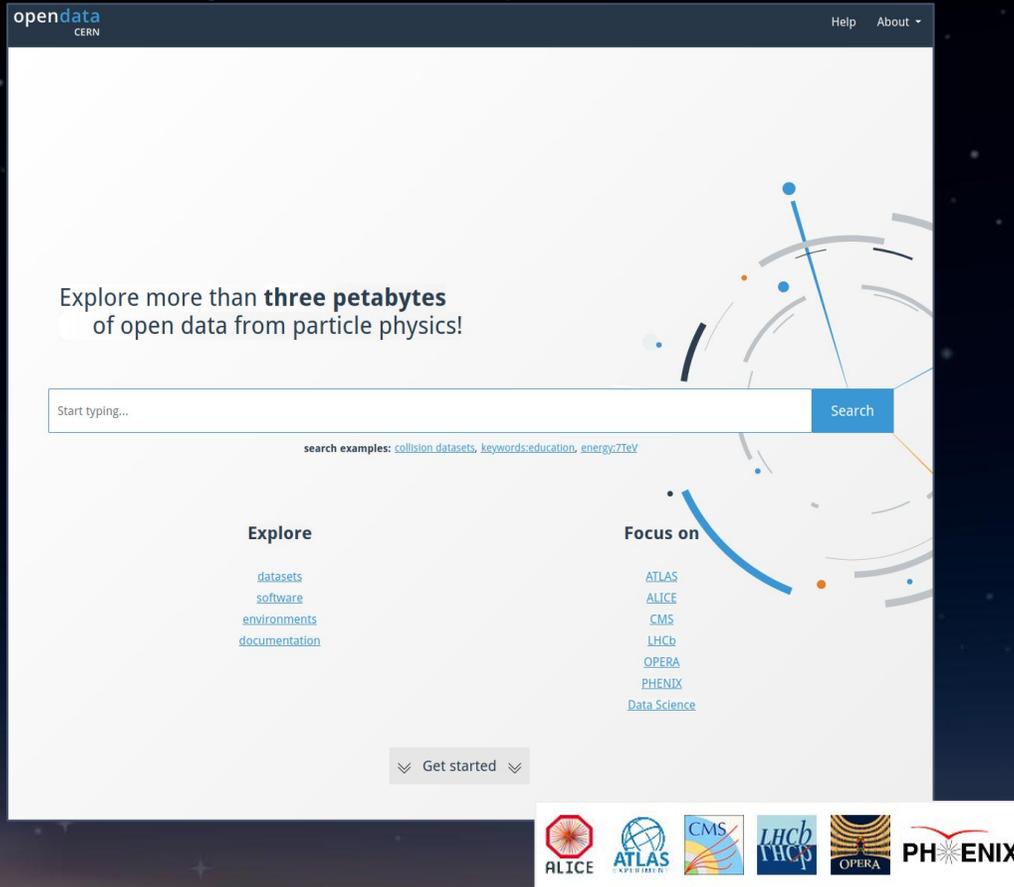
**Tibor Šimko**

CERN, Information Technology

Lead, CERN Open Data

Lead, REANA Reusable Analyses





<https://opendata.cern.ch>

## CERN Open Data

- collision and simulated datasets for research
- derived datasets for education
- configuration files and documentation
- virtual machines and container images
- software tools and analysis example

Over 3 petabytes and 1.5 million files

Launched in November 2014



## Selected observations

- ★ Build "together", not "for"
- ★ Expect unexpected
- ★ Open is not enough
- ★ Data is not enough

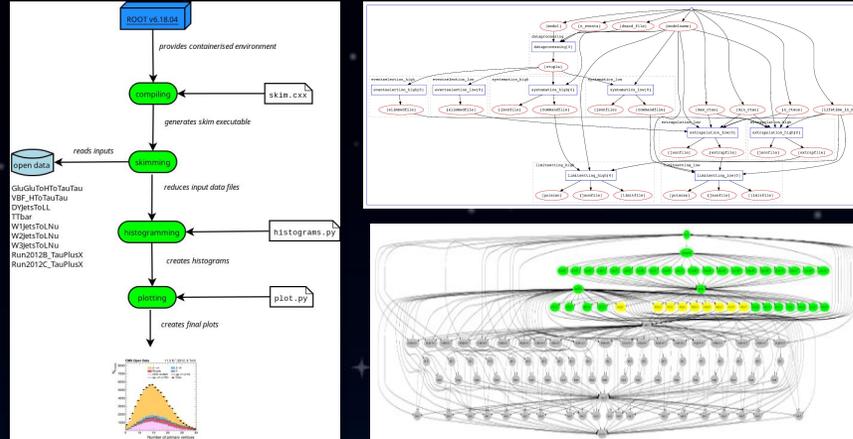


W3JetsToLNu dataset in reduced NanoAOD format for education and outreach

Analysis of Higgs boson decays to two tau leptons using data and simulation of events at the CMS detector from 2012

Description: This analysis uses data and simulation of events at the CMS experiment from 2012 with the goal to study decays of a Higgs boson into two tau leptons in the final state of a muon lepton and a hadronically decayed tau lepton. The analysis follows loosely the setup of the official CMS analysis published in 2014.

Two example results produced by this analysis can be seen below. The plots show the data recorded by the detector compared to the estimation of the contributing processes, which are explained in the following. The analysis has implemented the visualization of 34 such observables.



Workflow is the new Data!

reana

Reproducible research data analysis platform

Flexible	Scalable	Reusable	Free
Run many computational workflow engines.	Support for remote compute clouds.	Containerise once, reuse elsewhere. Cloud-native.	Free Software. MIT licence. Made with ❤️ at CERN.

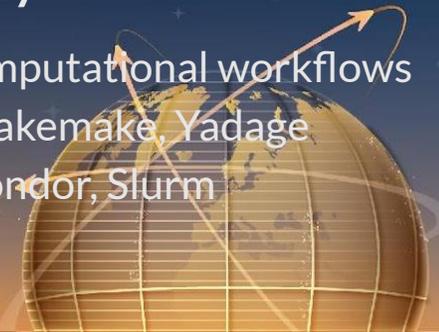
<https://www.reana.io>

Data are accompanied with actionable usage examples

- ★ *Ultimate goal? Facilitate future data reuse*
- ★ *What is the most efficient way to pass the knowledge to future generations?*
- ★ *Data + Code + Environment + Workflow = Reusable Science*

## REANA Reusable Analyses

- run containerised computational workflows
- using CWL, Serial, Snakemake, Yadage
- on Kubernetes, HTCondor, Slurm





**THANK YOU**

