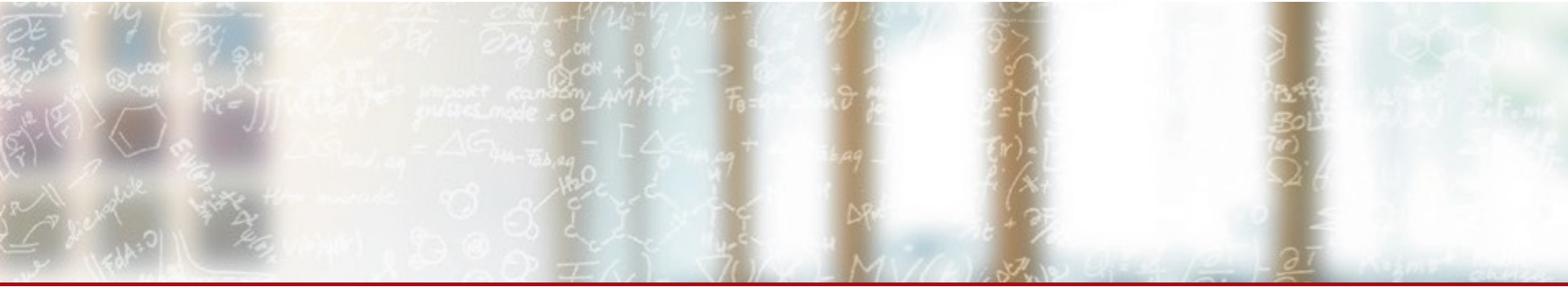
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



The Balance between Openness and Security in the Context of Scientific Research Data

HPC-ch Forum on HPC and Data as a Service

Victor Holanda Rusu, CSCS

October 5th, 2023

You can always take more than nothing.

— Lewis Carroll, Alice's Adventures in Wonderland

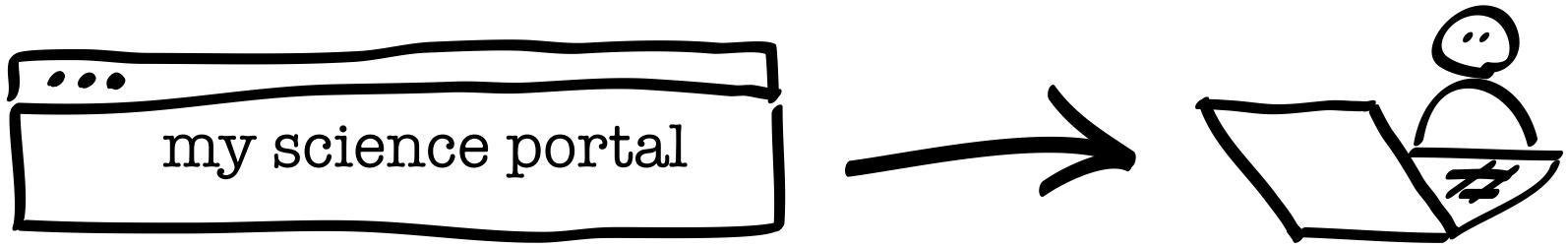
Setting up the Foundations

Defining our terms

What's Data as a Service?

The DaaS concept

- DaaS is a cloud computing model that provides on-demand access to data on a subscription basis.
- Data is hosted and maintained by a third-party service provider and made available to users or applications over the internet.
- Instead of managing and maintaining their own data infrastructure, organizations can rely on DaaS providers to deliver the data they need in a timely and cost-effective manner.



The “Science is a Global Endeavour” Concept

The driving forces

- This is a widely accepted concept that transcends the scientific community.
- As research fellow: no one tells you that, you live it!
- Be open to accept new concepts and ideas is part of Science.
- Scientific labs are full of different nationalities and cultures.
- Almost every country has a research institute; one reads papers from all over the world.
- International conferences and some local ones are packed with “foreigners”.
- This was “true” even before globalization.

The “Science is a Global Endeavour” Concept

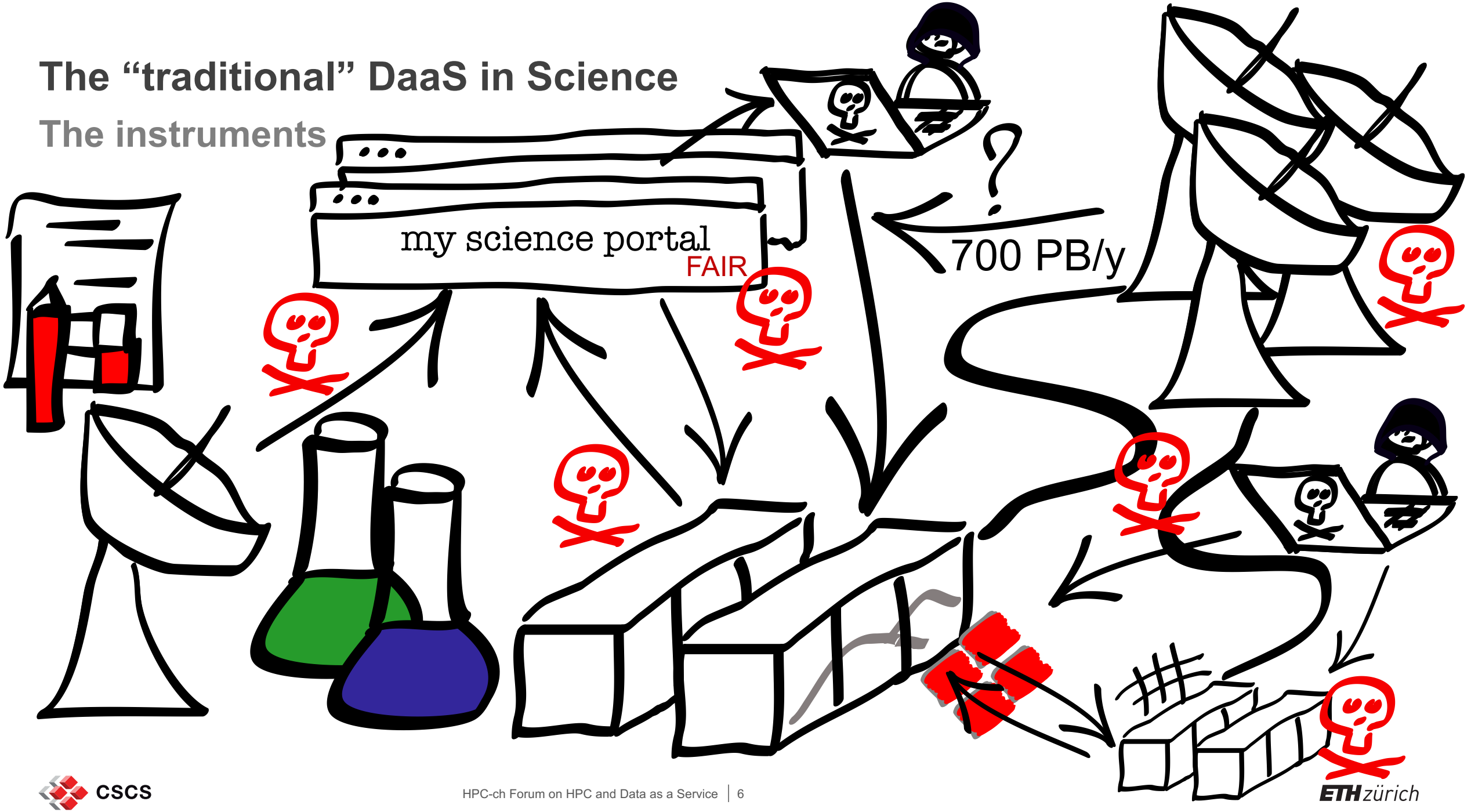
The natural consequences

- International Collaboration
  - Scientific research frequently involves collaboration among researchers and institutions from different countries.
  - Collaborative projects bring together diverse expertise, resources, and perspectives to address complex scientific questions.
- Global Challenges
  - Many of the most pressing scientific challenges facing humanity, such as climate change, pandemics, and space exploration, are global in nature and require international cooperation to address effectively.
- Shared Knowledge
  - Scientific knowledge is a collective endeavour, and the findings of one group of researchers can benefit scientists and society worldwide.
  - Open sharing of research findings and data is essential for the advancement of science.
- Access to Resources
  - Scientists often access international facilities, data, and resources to conduct their research.
  - These resources may include large scientific instruments, telescopes, or international research stations.
- Peer Review and Publication
  - Scientific research is typically subject to peer review by experts from various countries.
  - Journals and conferences in most scientific fields are open to contributions from researchers worldwide.
- Scientific Diplomacy
  - Science plays a role in international diplomacy and fosters cooperation and understanding among nations.
  - Collaborative scientific projects can promote peaceful relations and goodwill.
- Education and Talent Mobility
  - The global nature of science means that scientists, researchers, and students frequently move across borders to study, work, and contribute to research efforts.

Openness is not an option. It is fundamental pillar of what we do

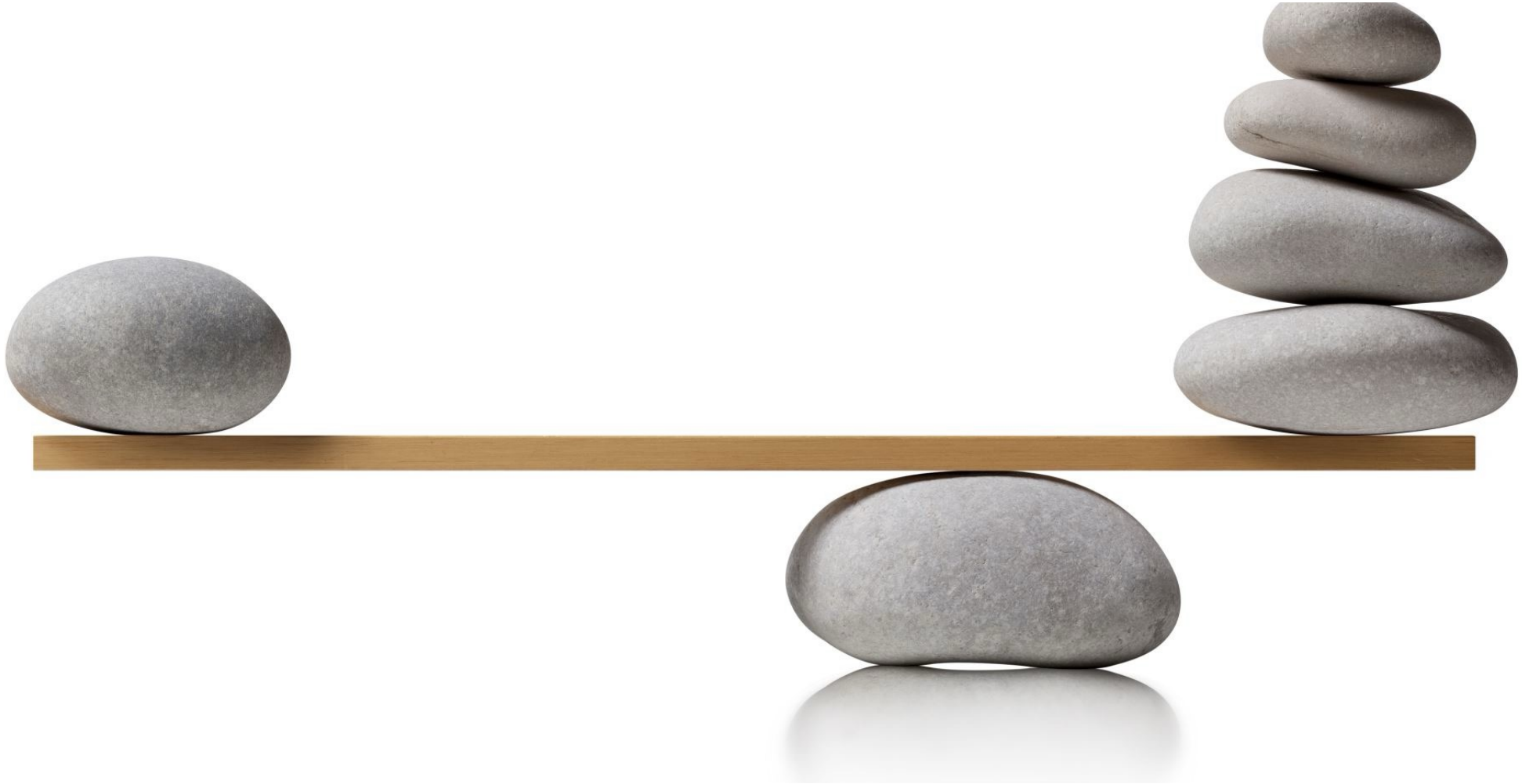
The “traditional” DaaS in Science

The instruments



The balance

The challenges



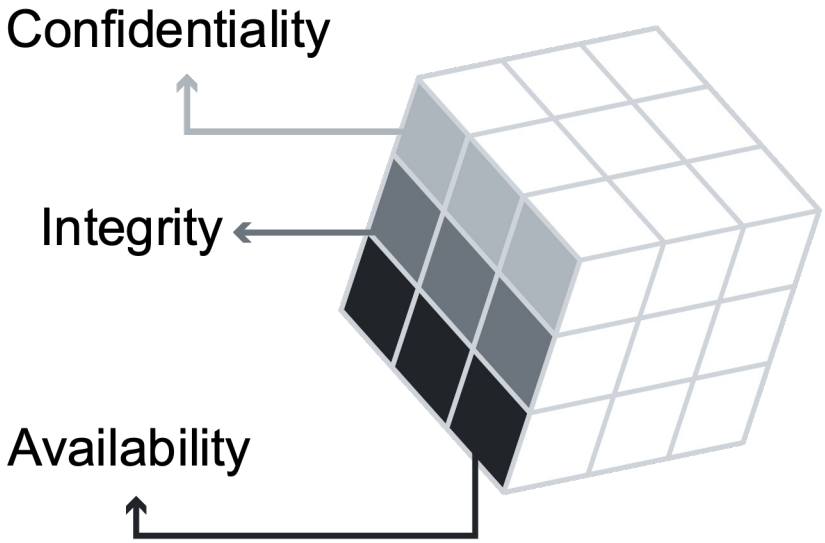
What can we do?

Rely on principles – The McCumber Cube

- It is a model framework created by John McCumber in 1991.
- Aims to help organizations evaluate and establish information security initiatives by considering all of the related factors that impact them.
- This security model has three dimensions:
  - The foundational principles for protecting information systems.
  - The protection of information in each of its possible states.
  - The security measures used to protect data.

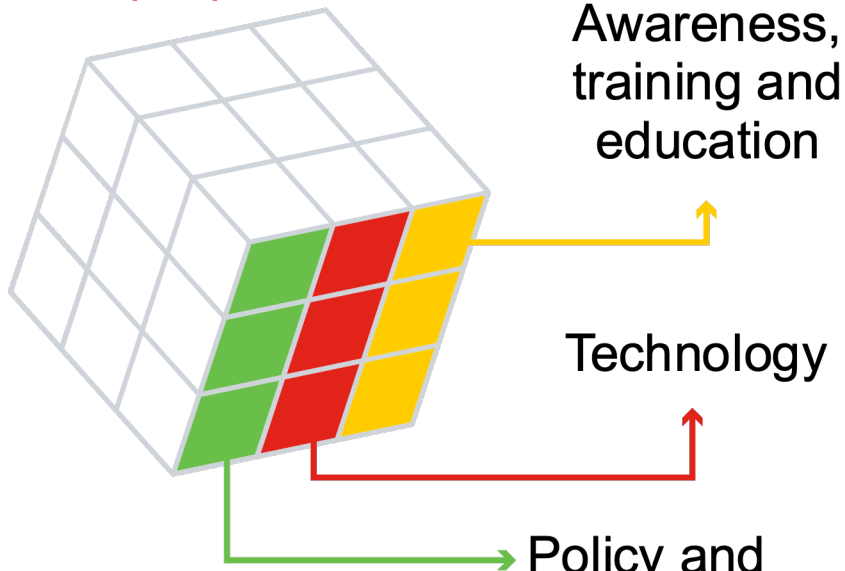
The McCumber Cube

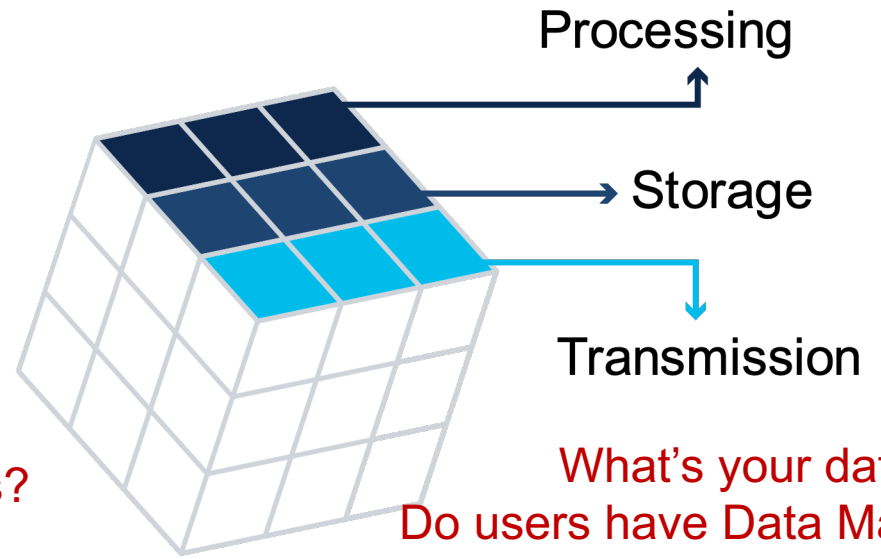
We live in multiple dimensions



What are the priorities?
How do you plan to achieve them?
Do you need strong consistency?
Are you aware of the CAP theorem?
Do you implement Mandatory Access Controls?

This is where people tend to concentrate



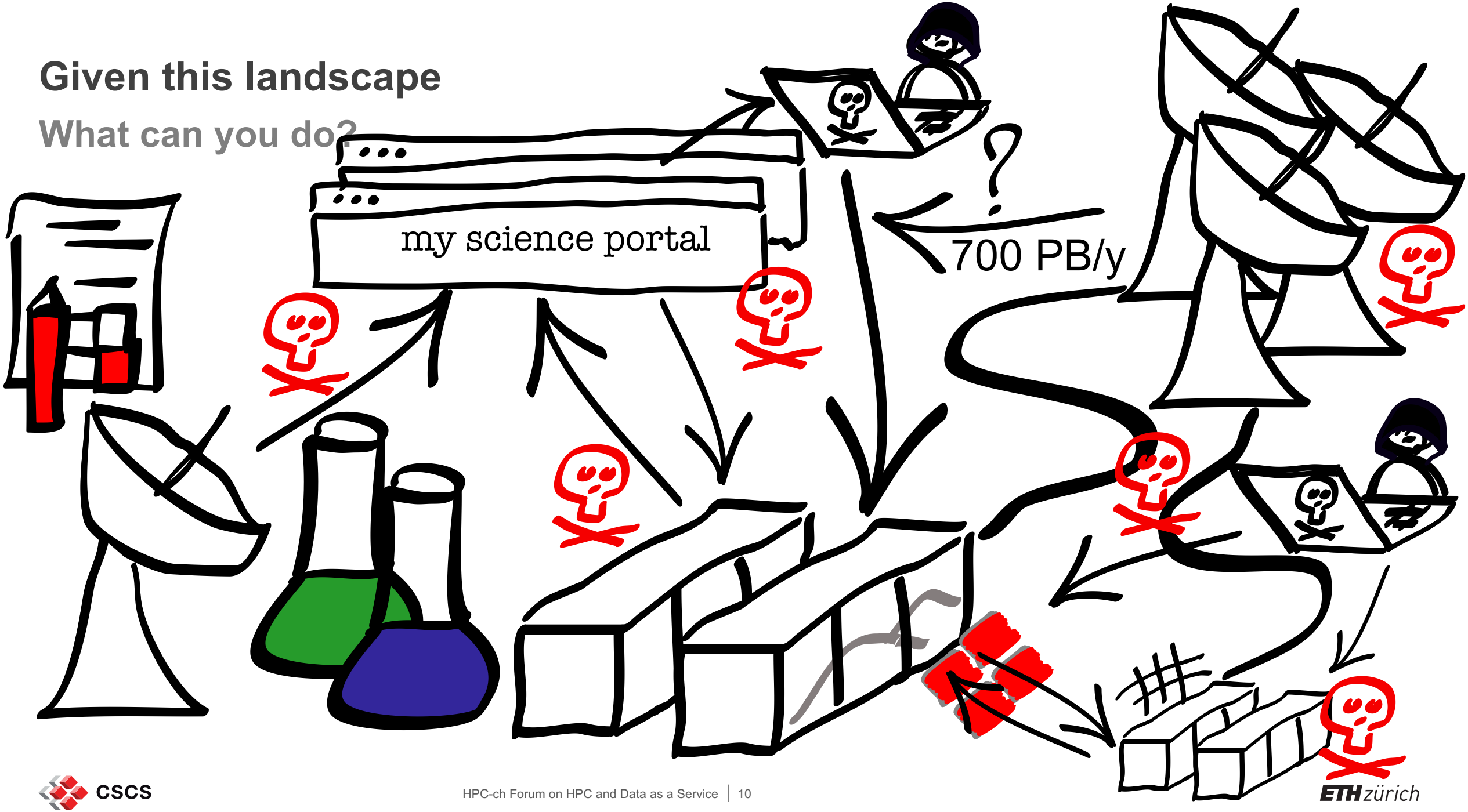


What's your data offering?
Do users have Data Management Plans?

Can you implement a single security approach to all the different use cases?

Given this landscape

What can you do?



What can we do as a HPC centers?

Divide and Conquer approach

- Identify what one can offer
  - For a given system what are the combinations of the CIA x Data Stages x Security Controls?
- Develop security profiles for each business offerings
  - Shared security responsibilities?
  - Business Associate Aggrements?
- Develop and operate multiple HPC platforms, not HPC machines
  - Isolate business and different security profiles
- Provide technologies and procedure to help onboard users
  - Data Management Offerings, help users create the Data Management plans
- Engage the Scientific communities you support
  - Scientists need to understand that competition goes beyond their peers
  - Demand clarity in the security governance, architecture and technologies from consortiums
- Identify acceptable compromises and risks
  - Isolate the threat internally to mitigate impact on other customers

What can the scientists do?

Share with us their needs

- Identify Data Security requirements
  - Do we need to know more about security?
  - Should we develop a HPC-ch cross community SETA program?
  - National awareness program, in the lines of CISA Secure Our World?
- Develop Research Data Management Plans
  - Understand your Research Data lifecyle
- Push your communities to develop secure by default solutions
- Implement Secure Development practices in your software projects
  - Implement regression testing
  - Static Application Securety Testing (SAST)
  - Dynamic Application Securety Testing (DAST)
  - Perform and publish Software Component Analyses (SCA)
  - Sign software releases
  - Provide artifact checksums to implement CRC processes

The Data Management Plan

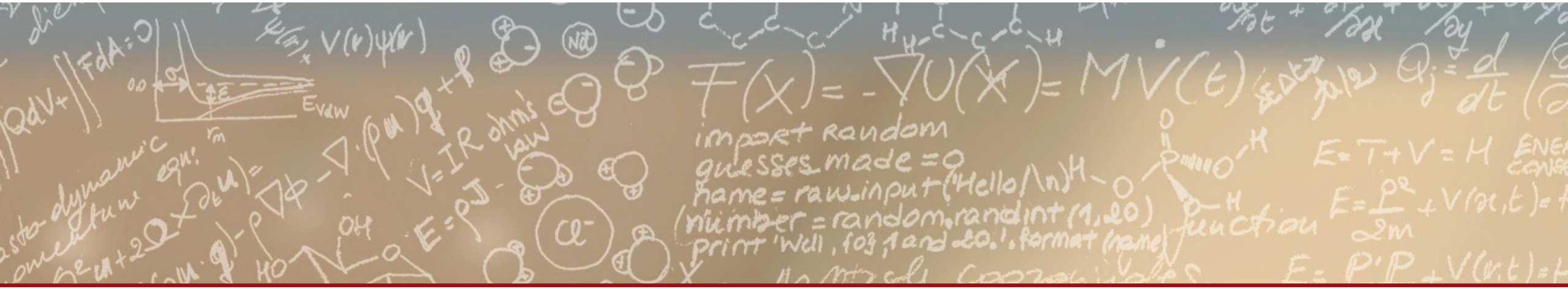
The user side of things

- Users and projects must develop a process of providing the appropriate labelling, storage, and access for data at all stages of a research project
- The plan should be composed of seven stages
  - Data Collection
  - Documentation and Metadata
  - Ethics and Legal Compliance
  - Storage and Backup
  - Selection and Preservation
  - Data Sharing
  - Responsibilities and Resources

The Data Management Plan

The user side of things

- Types of data:
  - What is the source of your data?
  - In what formats are your data?
  - Will your data be fixed, or will it change over time?
  - How much data will your project produce?
- Contextual details (metadata): How will you document and describe your data?
- Storage, backup, and security: How and where will you store and secure your data?
- Provisions for protection/privacy: What privacy and confidentiality issues must you address?
- Policies for re-use: How may other researchers use your data?
- Access and sharing: How will you provide access to your data by other researchers? How will others discover your data?
- Archiving and providing access: What are your plans for preserving the data and providing long-term access?
- Roles and plan oversight: Who will be responsible for aspects of data management throughout the project, and what resources are required for implementation?



Thank you for your attention.

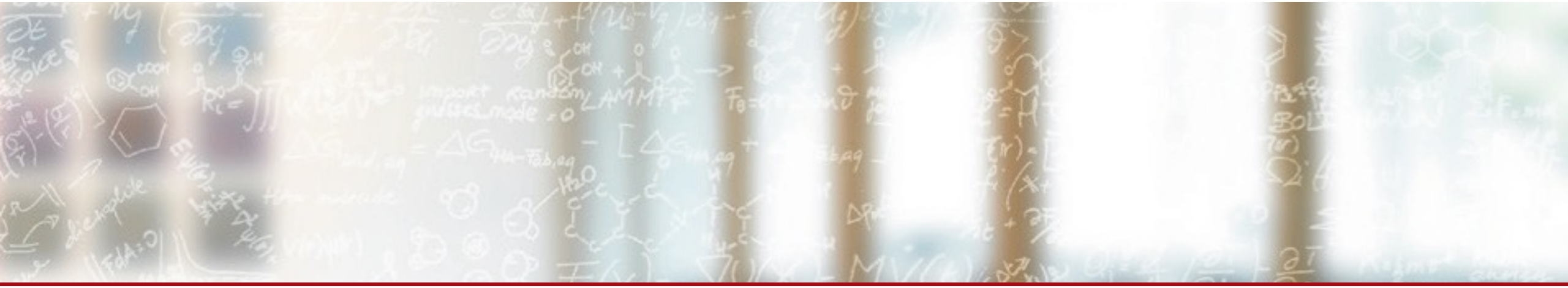
Questions?

“Where should I go?" -Alice.
"That depends on where you want to end up." - The Cheshire Cat.”

— Lewis Carroll, Alice's Adventures in Wonderland

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



The Balance between Openness and Security in the Context of Scientific Research Data

HPC-ch Forum on HPC and Data as a Service
Victor Holanda Rusu, CSCS
October 5th, 2023

"But that's just the trouble with me. I give myself very good advice, but I very seldom follow it."

— Lewis Carroll, Alice's Adventures in Wonderland