

## Identity Challenges in a Big Data world

**Paul Millar**

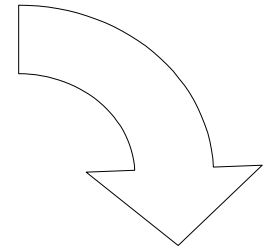
GridKa School 2014,  
Karlsruhe, Germany



Authentication



Authorisation



Can read a file?

Can write a file?

Can delete a file?

Stage file from tape?

Can create directory?

# Who are you?



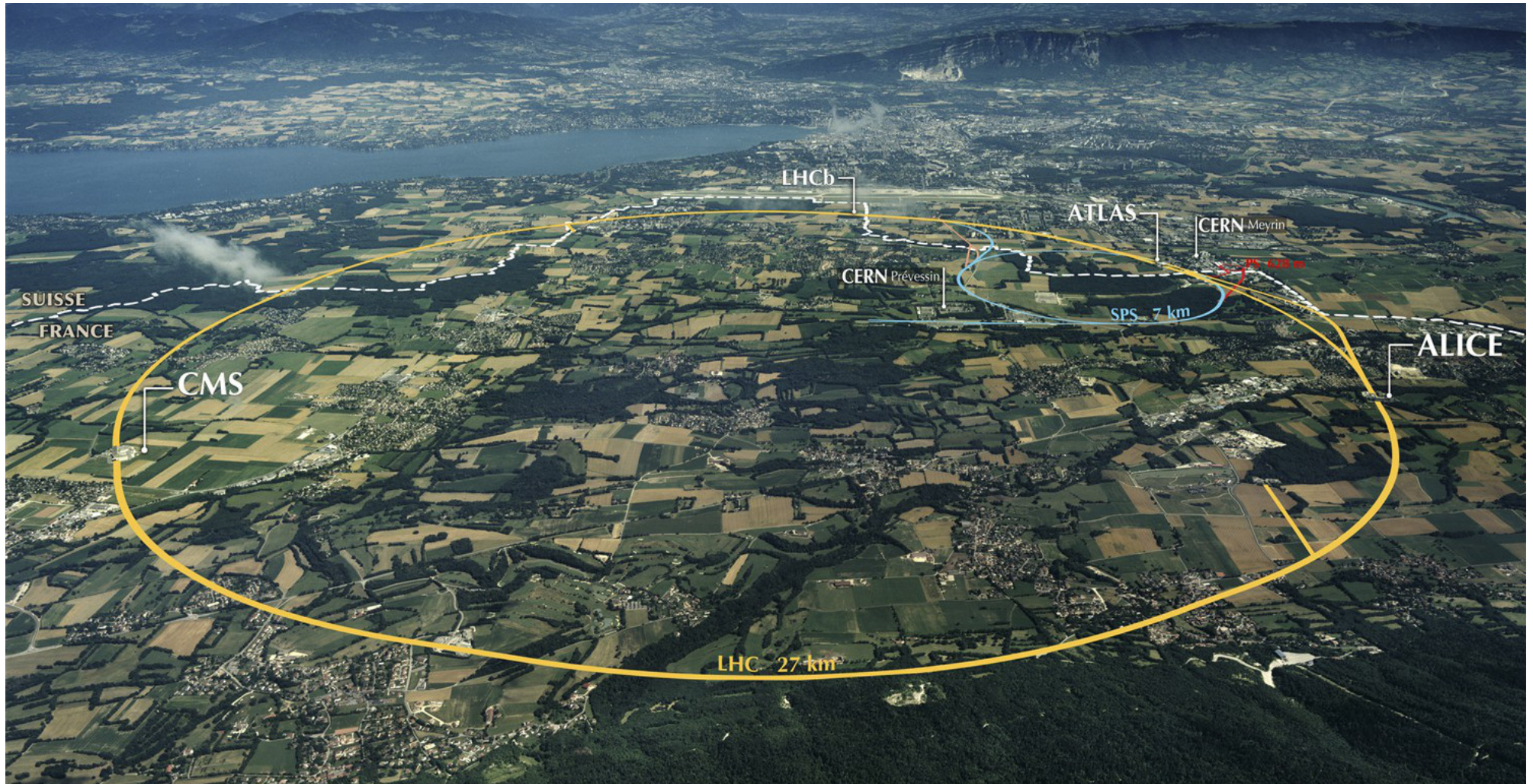
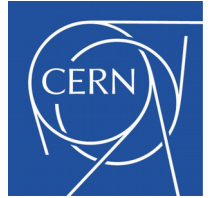


# DESY: photon science





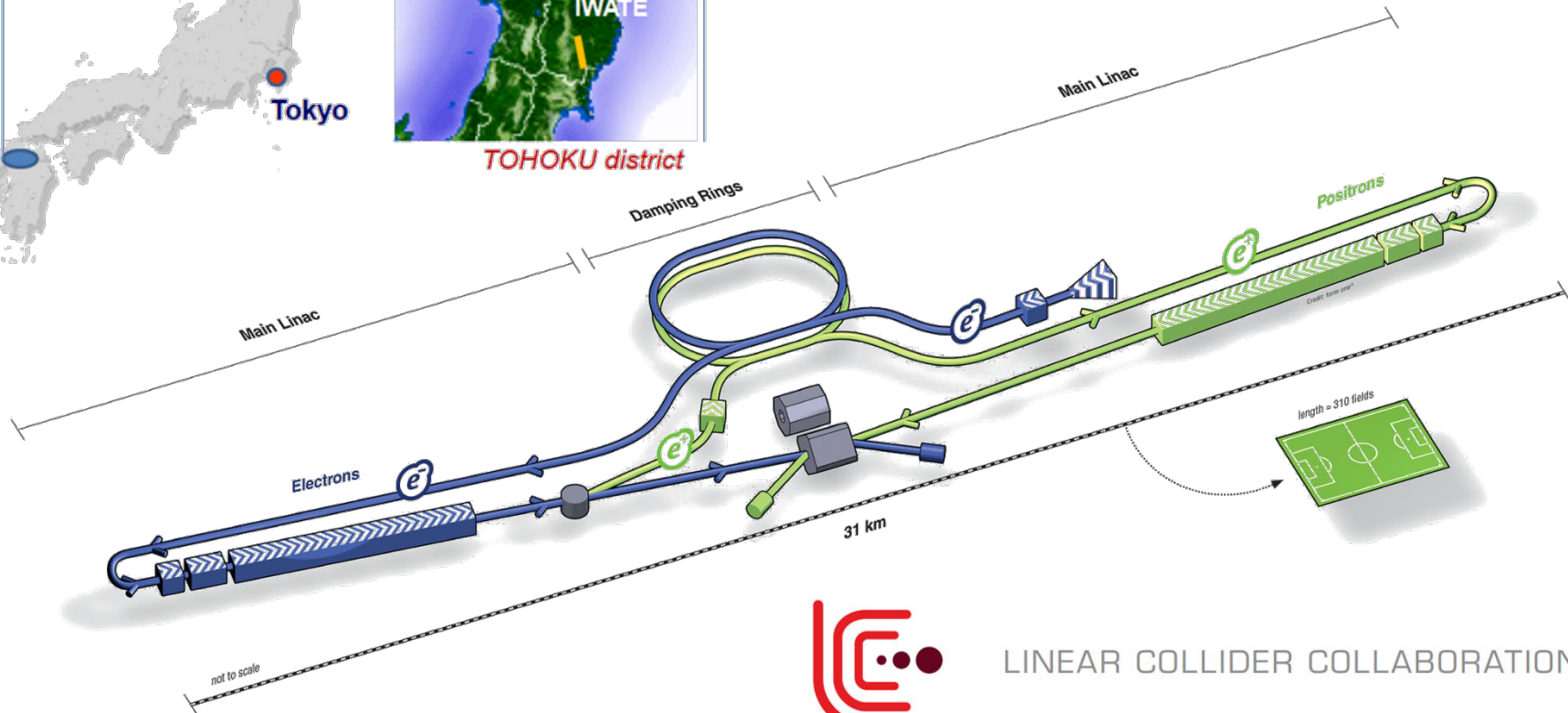
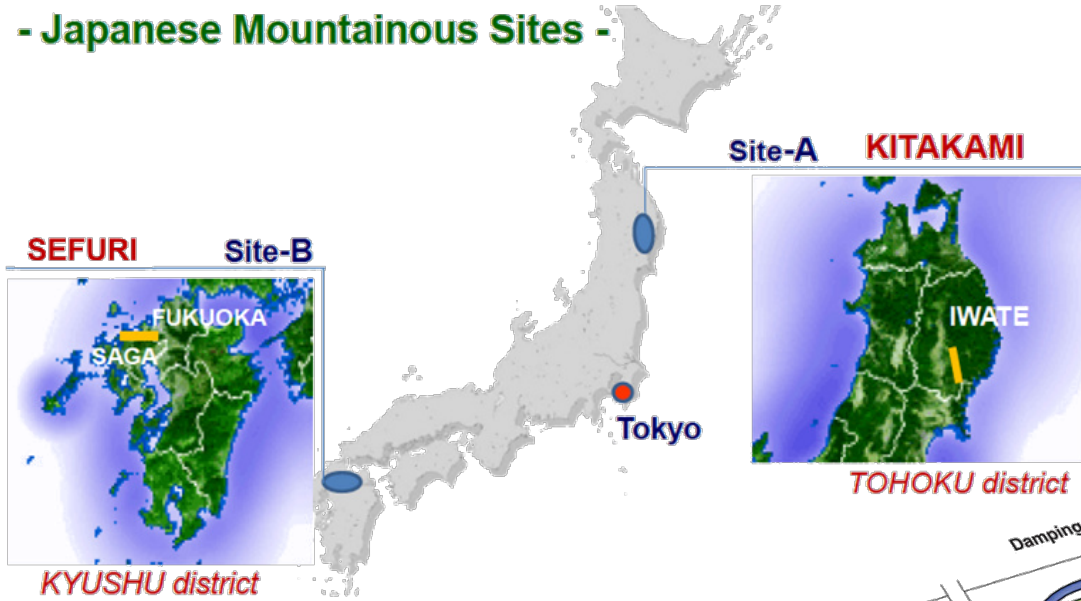
# DESY: particle physics at LHC





# DESY: particle physics at the ...

- Japanese Mountainous Sites -





## “DESY IT users”

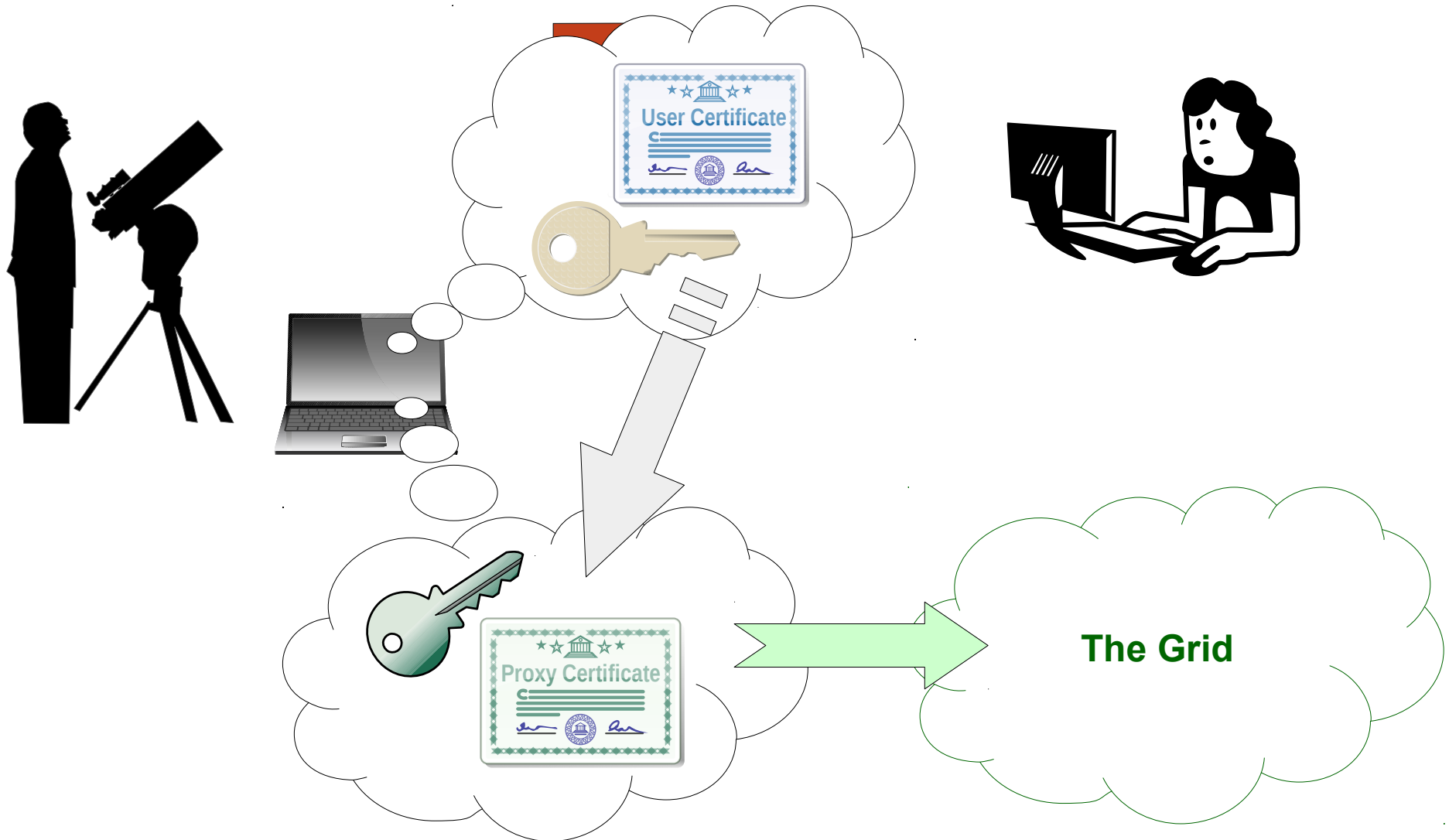
(people using DESY IT resources)

- people who come to DESY to use any of the photon science facilities,
- anyone in the supported LHC collaborations (worldwide),
- anyone in the ILC collaboration (worldwide),
- people our other collaborations: IceCube, Cherenkov Telescope Array (CTA), Belle II, ...

... oh, and some people who sit in offices at DESY.

---

# The grid solution: X.509 (user) certificates





# X.509 certificates: a huge success!

"It's been a global effort, a global success. It has only been possible because of the extraordinary achievements of the experiments, infrastructure and the **grid computing**."

**Rolf Heuer**, the Director General of CERN

---



# X.509 certificates: typical user reaction





# Federated Identity



Check who you are  
&  
Authorisation decision

Check who you are



Record information




Authorisation decision

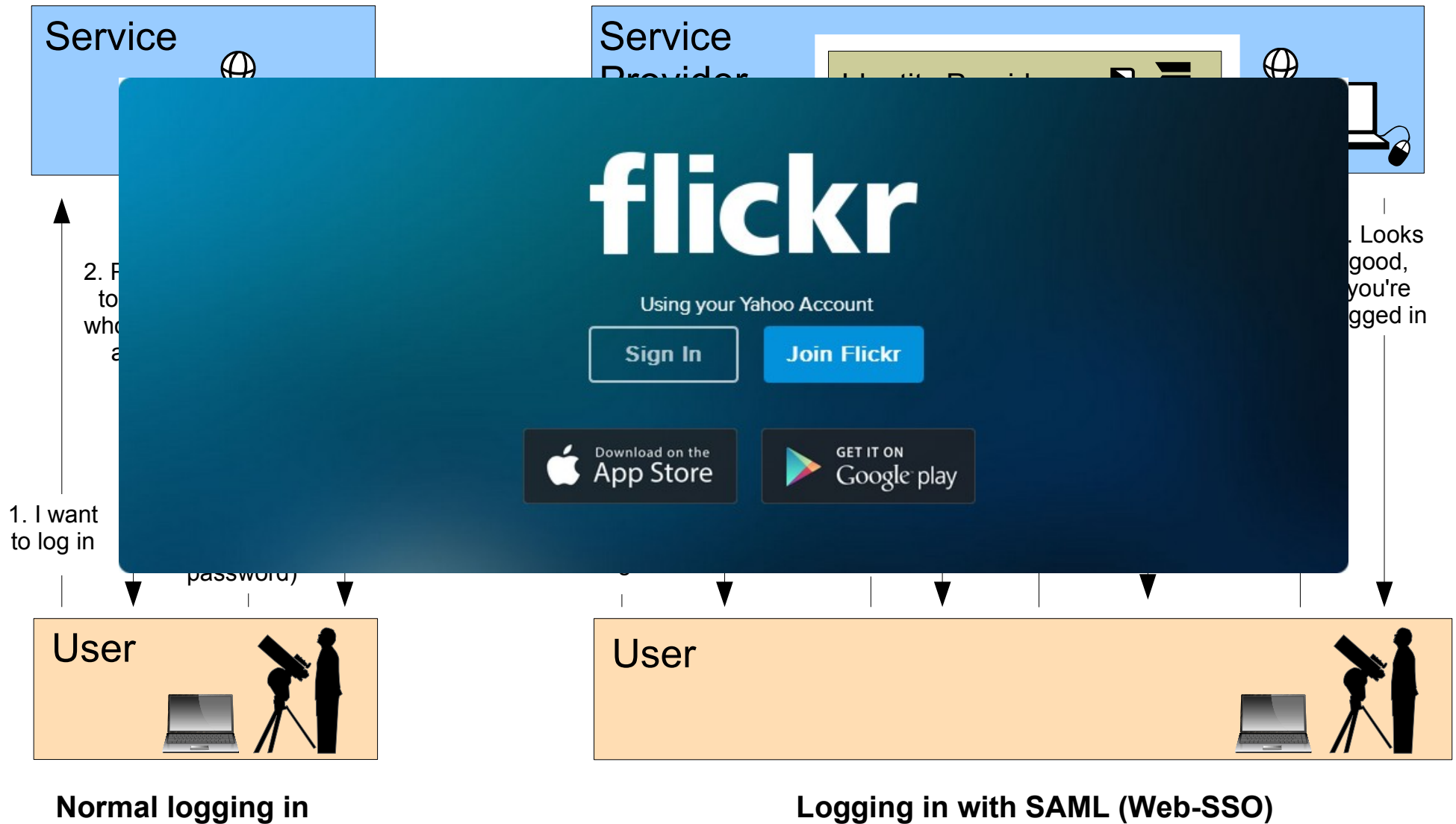


Identity Provider (IdP) 

Assertion

Service Provider (SP) 

# SAML Web Single Sign-On (Web SSO)

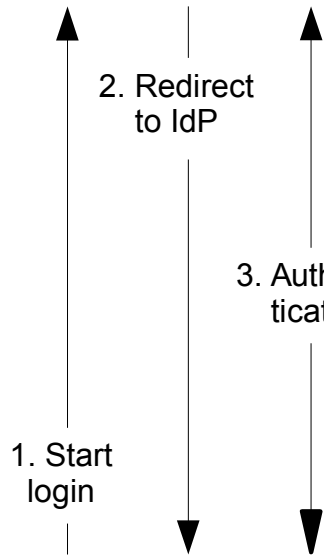




# “Where”

Service Provider (SP)

IdP (IdP)



User

SAML Web browser

DFN-AAI - Iceweasel

DFN-AAI

https://wayf.aai.dfn.de/DFN-AAI/wayf/WAYF

DFN  
Deutsches Forschungsnetz

## DFN-AAI

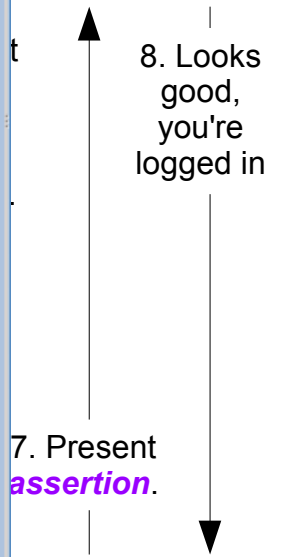
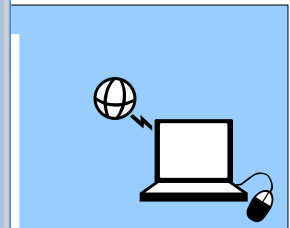
### Permanently set your Home Organization

On this page you can set a **default Home Organization** for this web browser. Setting a default Home Organization will hencefort redirect you directly to your Home Organization when you access AAI-Resources. Don't use this feature if you use several AAI accounts.

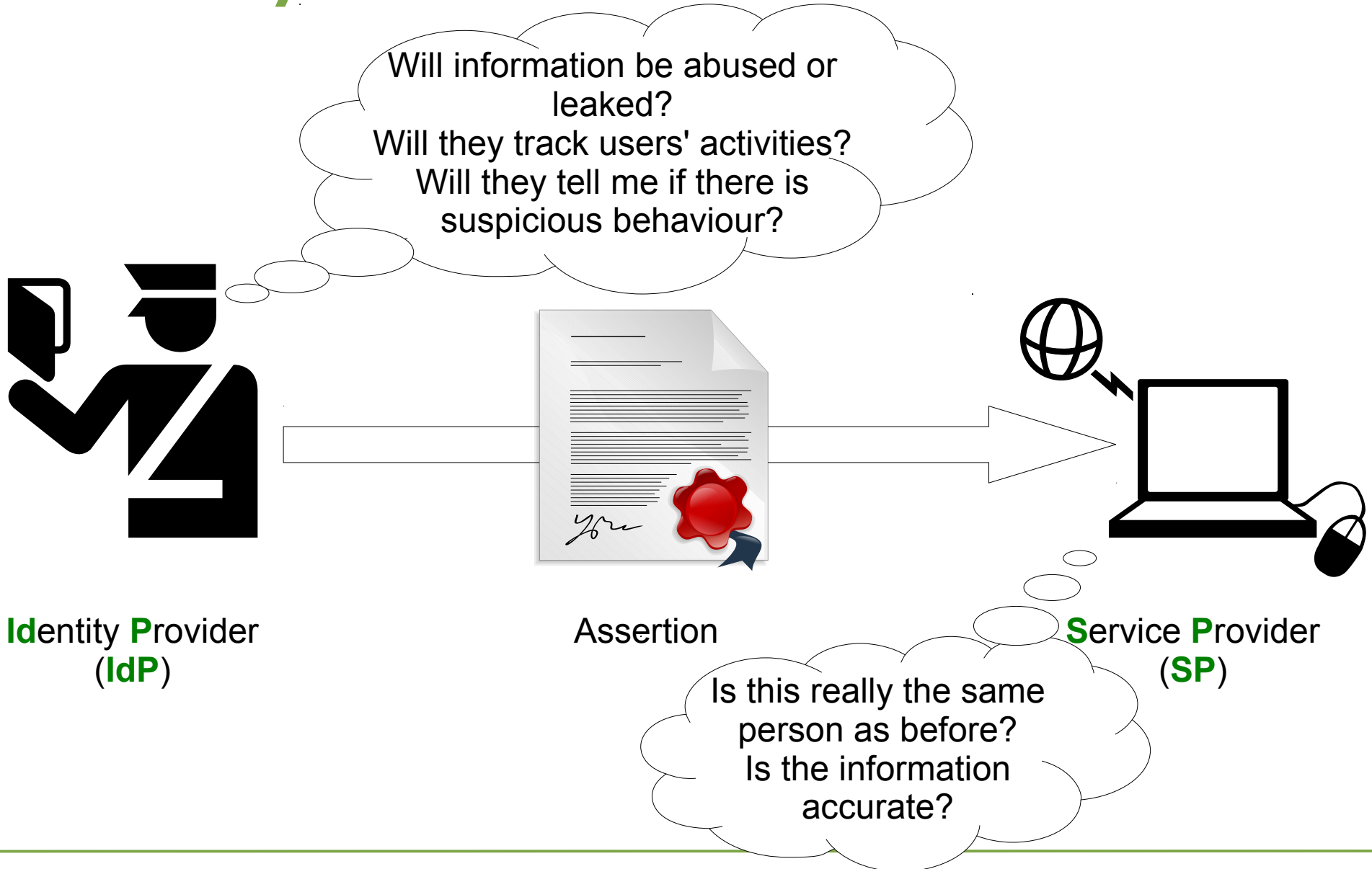
**Remember selection permanently and bypass WAYF from now on.**

Select the Home Organization you are affiliated with ... Save

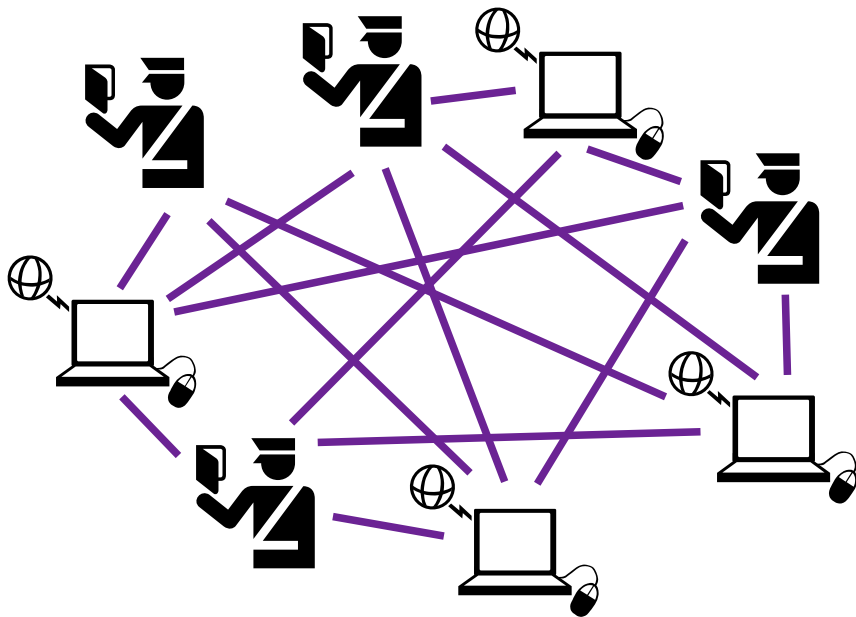
- DFN - Deutsches Zentrum fuer Edu- und Research-Inf...
- Europa-Universität Flensburg
- FH Worms
- Fachhochschule Düsseldorf
- Forschungszentrum Jülich GmbH
- Fraunhofer-Gesellschaft
- Freie Universität Berlin
- Friedrich-Schiller-Universität Jena
- GESIS - Leibniz-Institut für Sozialwissenschaften
- Georg-August Universität Göttingen
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
- HS-Harz
- HSU Hamburg
- HTWG Konstanz**
- Helmholtz Zentrum München
- Helmholtz-Zentrum Dresden-Rossendorf e.V.
- Helmholtz-Zentrum für Umweltforschung - UFZ
- HfWU Nürtingen-Geislingen
- Hochschule Aalen - Technik und Wirtschaft
- Hochschule Albstadt-Sigmaringen
- Hochschule Augsburg



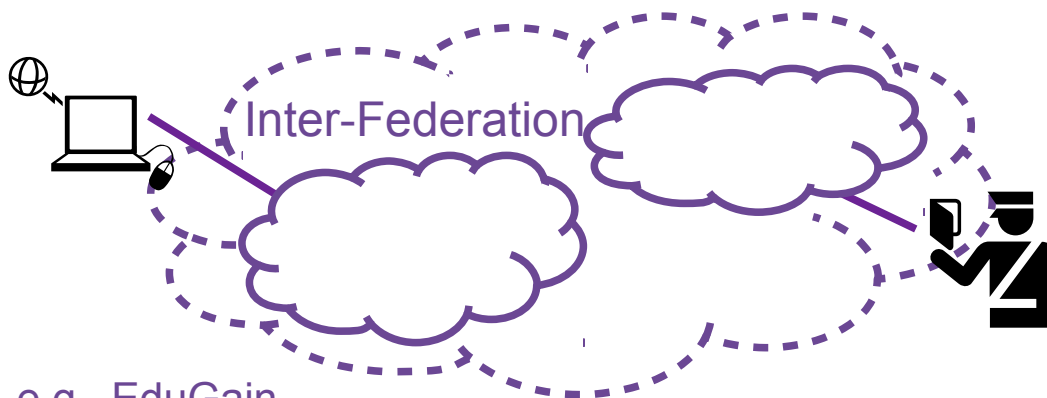
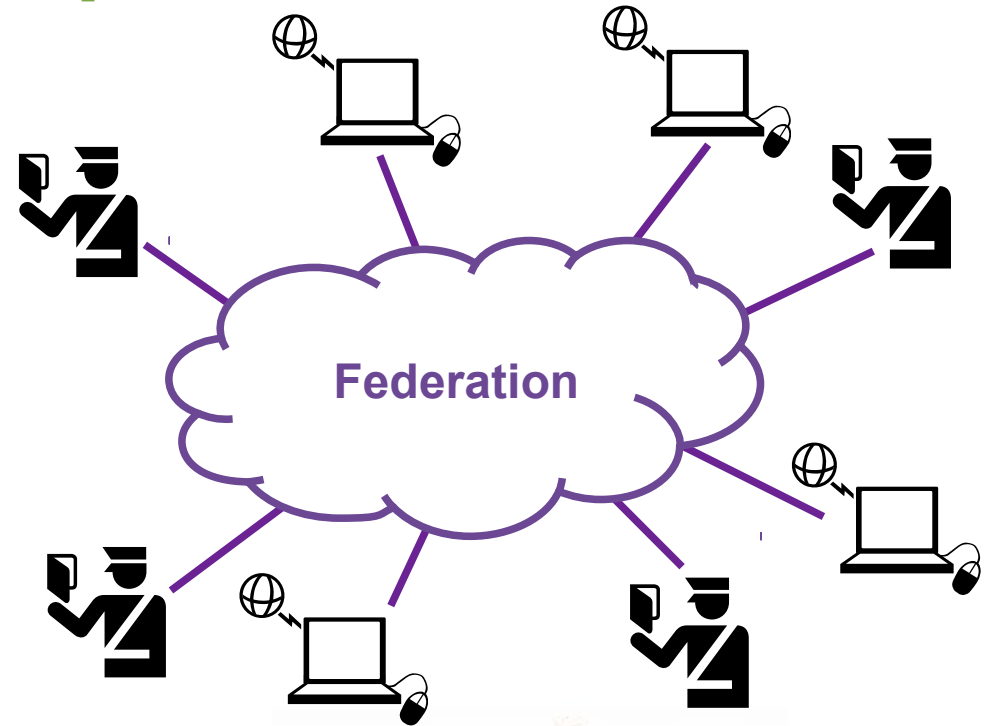
# Who do you trust?



# How to trust lots of people?



Point-to-point **trust** doesn't scale!



e.g., EduGain



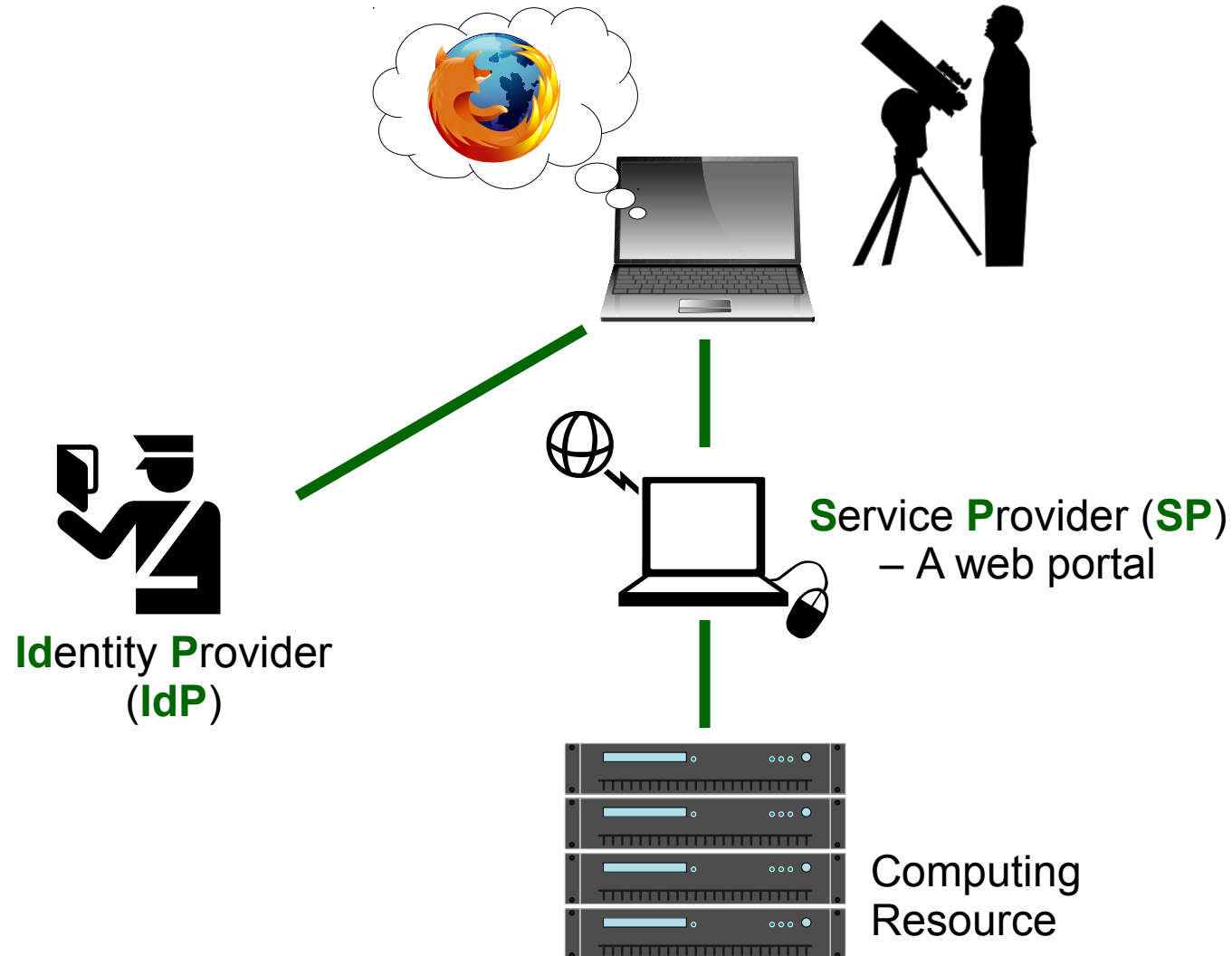


# What is Big Data?

- Three Vs: Volume, Velocity, Variety, (“Veracity”, Value, ...)
  - **Data storage and processing that is outside your comfort-zone.**
  - More scientific research now involves sifting through large amounts of data.
    - Particle Physics, Astronomy, Genomics, Biology, Medicine, ...
  - Often there is too much data to “just copy.”
    - Stored at specialise centres, like DESY.
  - Efficiency becomes increasingly important.
    - Need good algorithms and good bandwidth
  - Move the program to the data.
    - Need to access computers remotely
-

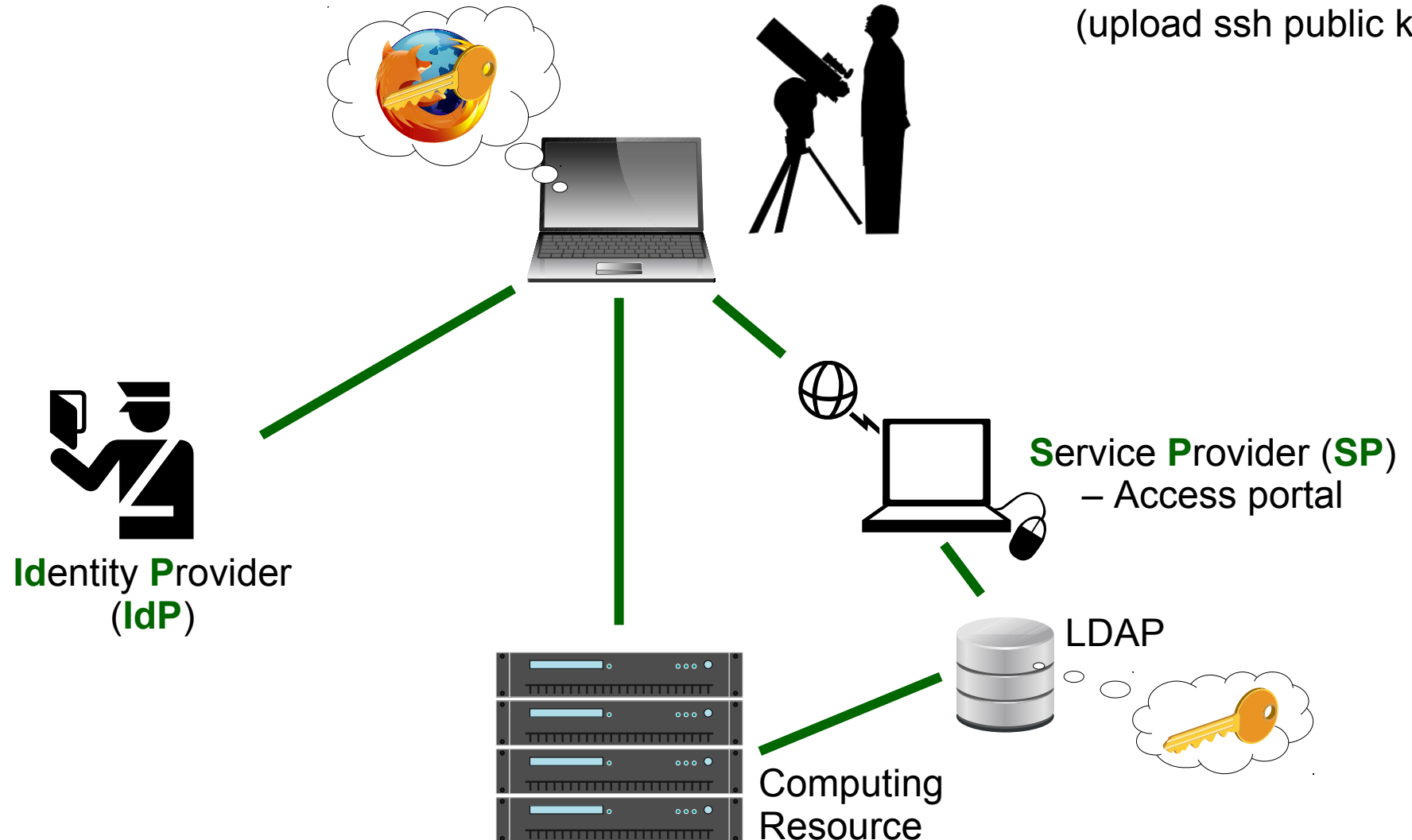
# Using (remote) computers

Web portal



# Using (remote) computers

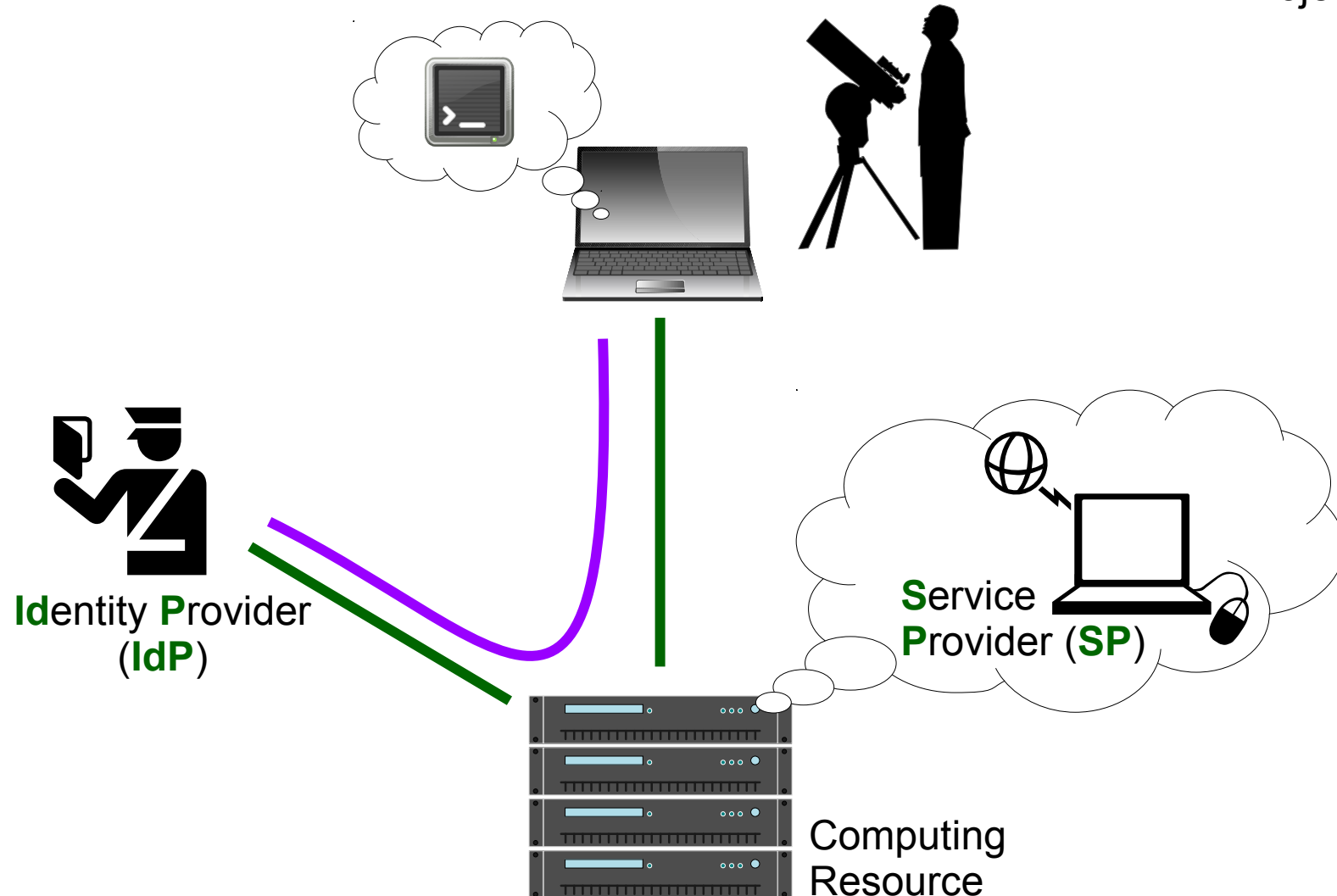
Substitute Credential  
(upload ssh public key)





# Using (remote) computers

Project Moonshot

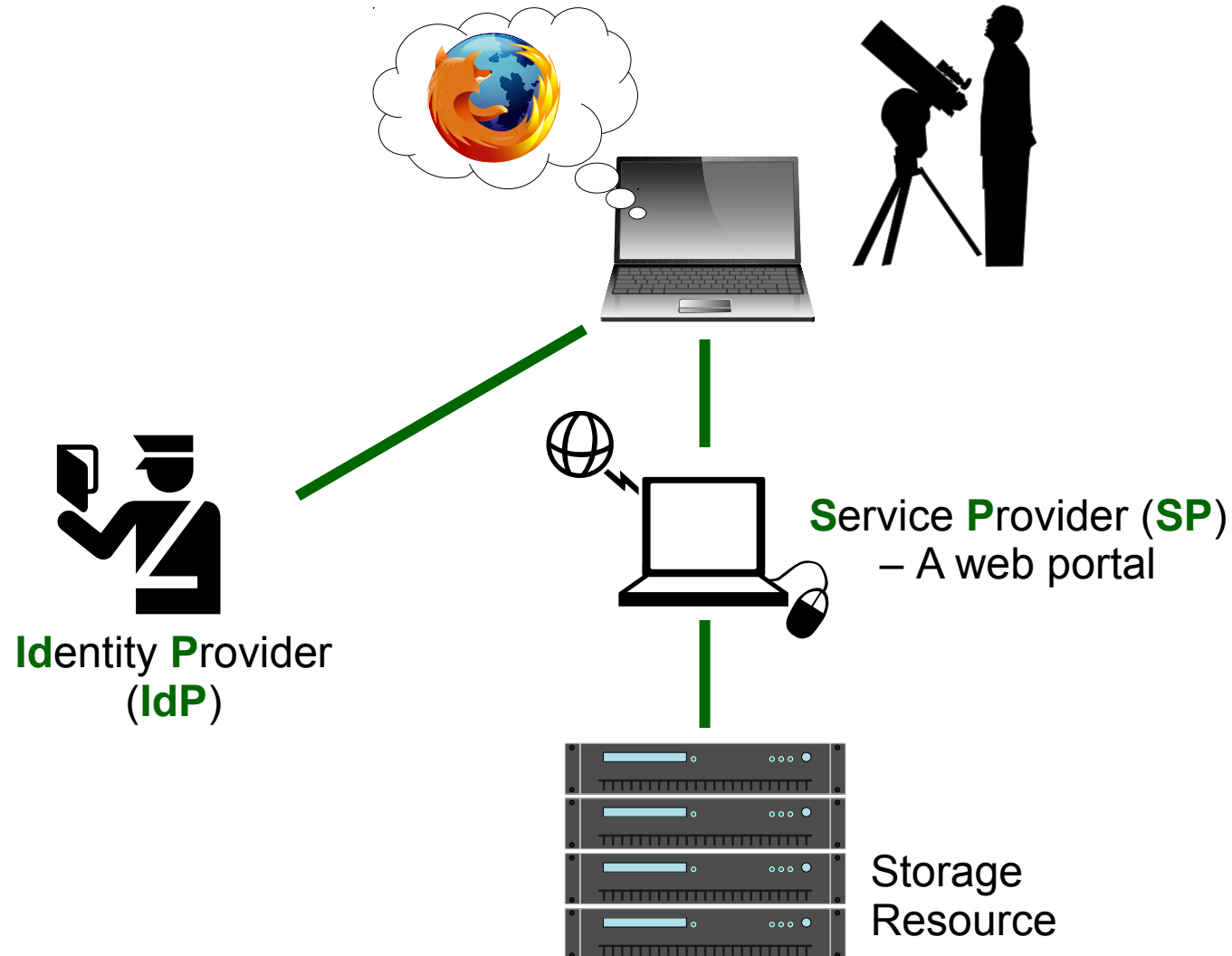


# Managing data with a Federated Identity

- Still need to manage the data:
    - Copy data back home,
    - Add annotation and other metadata,
    - Delete “bad data”,
    - Change permissions,
    - Manage data latency
-

# Managing (remote) data

Web portal



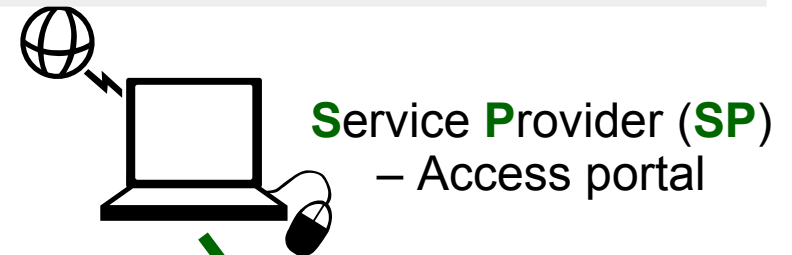


# Managing (remote) storage

Fetch Substitute Credential



**Token: Amazon AWS/S3 SAML support,  
X.509: SLCS, TCS, CI-Login, EMI STS, ...**

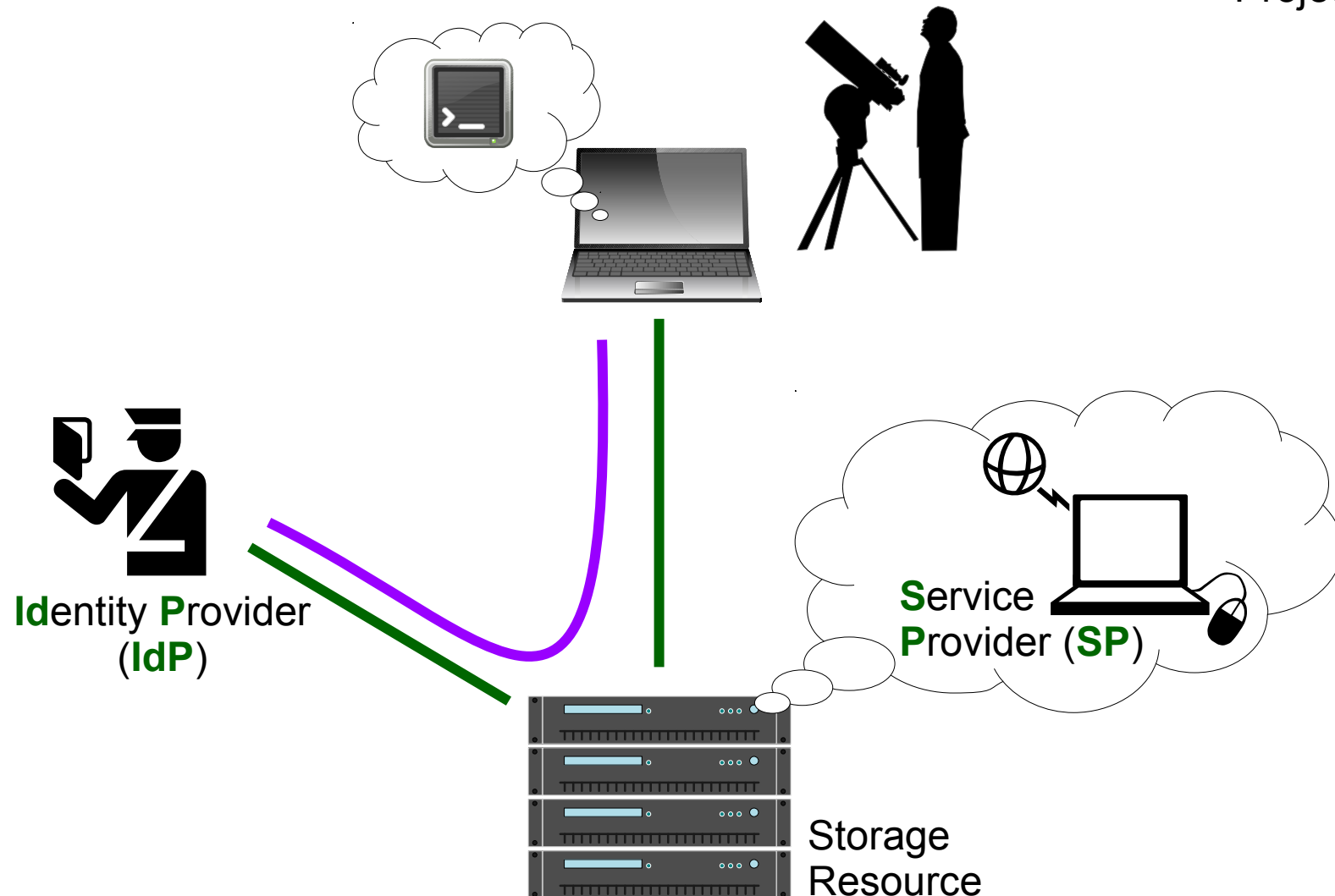


**Storage Resource**



# Managing (remote) storage

Project Moonshot



# Just scratching the surface...

- Multiple identities (e.g., the “nomadic user”),
  - Users with no home institute (the “homeless user”),
  - Group management,
  - Distributed authorisation,
  - Allocation / distributed quotes,
  - Accounting,
  - Discovering a user has departed,
  - Decommissioning policies,
  - Delegation,
  - Legal basis for releasing attributes,
  - Cross federation and inter-federation agreements,
  - ...
-



## Take home message(s)

- Federated Identity lets you **safely** use the same password.
- Your home institute needs to run a **SAML IdP** for this to work: make sure it does!
- Big Data will force Federated Identity into **new territory**:  
work has already started, more work is needed.



**Thanks for listening!**

---