

# Rucio Data Management system

---

Cédric Serfon for the Rucio team

# Introduction

---

- Questions that will be addressed in this talk :
  - What is Rucio ?
  - Why do I need a Data Management system ?
  - What would I gain using Rucio ?
  - Wouldn't it be an overkill to use Rucio ?
  - Who is using Rucio ?
  - What is the support model ?
- If you have questions during the talk, please ask

# What is Rucio ?

---

- Rucio is the Data Management system of the ATLAS experiment
- It was built using more than 10 years of experience in Data Management:
  - Designed from experience from the previous data management system DQ2
  - Integrate new features and technologies
- Modular, highly scalable, well supported
- Who is using Rucio ?
  - Used by ATLAS, [AMS](#) and [Xenon1T](#)
  - Being evaluated by other small and big HEP/Astro experiments (CMS, LIGO, IceCube, LSST...)
  - [Rucio community workshop](#) on March 1st-2nd 2018 to present Rucio to more collaboration/scientific communities

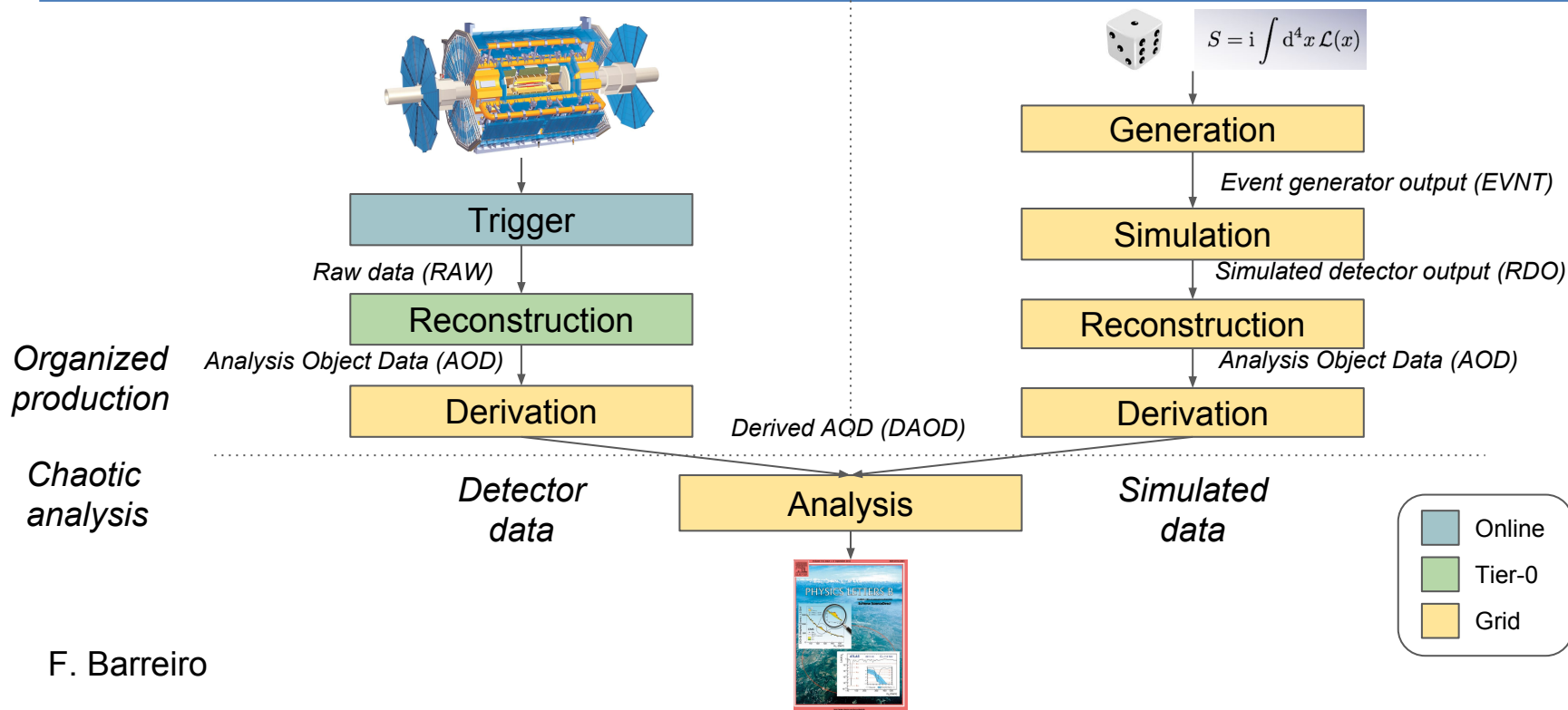
# Rucio main functionalities

---

- It provides many features (you are not forced to enable all) :
  - File and dataset catalog (logical definition and replicas)
  - Transfers between sites and staging capabilities
  - User Interface and Command Line Interface to allow user to download/upload/transfer their data
  - Extensive monitoring
  - Powerful policy engines (rules and subscriptions)
  - Bad file identification and recovery
  - Dataset popularity based replication
  - ...
- Rucio can be easily integrated with Workload and Workflow management system
  - With PanDA (>1M files/day)

More advanced features  
↓

# ATLAS workflows



F. Barreiro

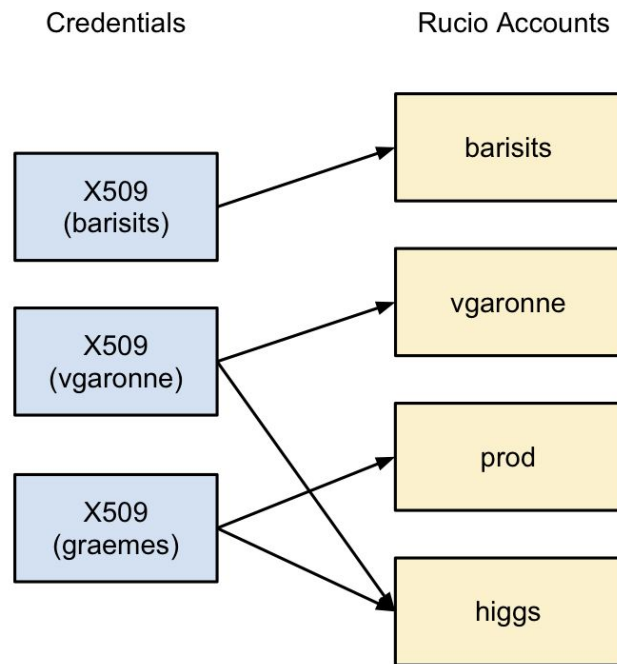
# Rucio in ATLAS

---

- Rucio stores all ATLAS data :
  - Data coming from the detector
  - Monte Carlo data
  - User data
- Rucio takes care of :
  - Ensuring the replication of files according to the replication policy specified by ATLAS
  - Replicate the data for other applications (e.g. panda) and for the end-users
  - Ensure file recovery
  - Staging data from TAPE
  - And plenty other things

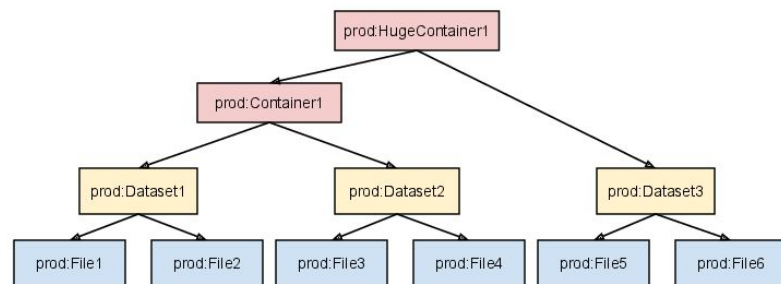
# Rucio main concepts - Accounts

- Each user who wants to use Rucio needs an account
- There are different types of accounts : user, service, group
- One user can have different accounts and can use credentials (X509, kerberos token, userpass, ssh) to connect to Rucio
- Permissions and quotas are based on accounts



# Rucio main concepts - DIDs

- The data stored in Rucio are identified by a Data Identifier (DID)
- There are different types of DIDs :
  - Files
  - Datasets : collection of files
  - Container : collection of dataset and/or container
- Each DID is composed of :
  - A scope : 25 characters to partition your data, e.g. data17, mc17
  - A name (up to 255 character)
- A name is unique within a scope



- DIDs hold a set of basic metadata e.g.
  - Bytes
  - Checksum (for files)
  - Number of events
  - Datatype



# Rucio main concepts - RSEs

---

- Rucio Storage Elements (RSEs) are logical entities of space
  - No software needed to run at the site
  - RSE names are arbitrary (e.g., "CERN-PROD\_DATADISK", "AWS\_REGION\_USEAST", ... )
  - Usually one RSE per site and storage data class
- RSEs collect all necessary metadata for a storage
  - protocols, hostnames, ports, prefixes, paths, implementations, ...
  - data access priorities can be set (e.g., to prefer a protocol for LAN access)
- RSEs can be tagged
  - Key/Value pairs (e.g., *country=UK, type=TAPE, support=brian@unl.edu*)
  - You can use RSE expressions to describe a list of RSEs (e.g. *country=UK&type=TAPE*)

# Rucio main concepts - Rules and subscriptions

---

- Replication rules :
  - Describe how a Data Identifier (DID) must be replicated on a list of Rucio Storage Elements (RSE)
  - e.g. : Make 2 replicas of dataset data15\_13TeV:mydatasetname on tier=1&disk=1
  - Rucio will create the minimum number of replicas to optimise storage space, minimise the number of transfers and automate data distribution
- Subscriptions :
  - Replication policies based on Data Identifiers metadata, for Data Identifiers that will be produced in the future
  - e.g. : Make 2 replicas of datasets with scope=data15\_13TeV and datatype=AOD on tier=1&disk=1

# Search DIDs & show details

<https://rucio-ui.cern.ch/search?scope=user.bdouglas&name=undefined>

ATLAS Rucio UI

Monitoring ▾

Reports ▾

user.bdouglas

Search

Using account: tbeerman ▾

Other Monitoring ▾

Help ▾

You are here: Search [All DIDs in scope user.bdouglas]

Rucio Version: 0.2.11

Show 

10 ▾

 entries

Search: 

dataset

Type	DID	Parent
DATASET	<a href="#">user.bdouglas:user.bdouglas.0226201326.408875.lib_.002231</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.0401160328.393826.lib_.002233</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.1224161342.665999.lib_.002200</a>	
CONTAINER	<a href="#">user.bdouglas:user.bdouglas.GRIF-IRFU_PHYS-TOP-datasets-to-possibly-move-04OCT11-0836</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.mc14_8TeV.117050.PowhegPythia_P2011C_ttbar.recon.AOD.e1727_s1933_s1911_r5591_tid01494881_00_der1412624703</a>	
CONTAINER	<a href="#">user.bdouglas:user.bdouglas.NIKHEF-ELPROD_PHYS-TOP_datasets_to_move_15Oct12_1212</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.physics_Egamma.SMWZd3pdExample.NTUP_SMWZ.f406_m991_p716</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.physics_Muons.SMWZd3pdExample.NTUP_SMWZ.f406_m991_p716</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.rucio_cache_test.26Sep14_1630</a>	
DATASET	<a href="#">user.bdouglas:user.bdouglas.testds_10Nov14_1519</a>	

Showing 1 to 10 of 31 entries (filtered from 21,975 total entries)

Previous

1

234Next

# WebUI: Data Discovery, details, transfers, etc.

<https://rucio-ui.cern.ch/search?scope=user.bdouglas&name=undefined>

ATLAS Rucio UI

Monitoring

Reports

user.bdouglas

Search

Using account: tbeerman

Other Monitoring

Help

You are here: Search [All DIDs in scope user.bdouglas]

Show 10 entries

ATLAS Rucio UI

Monitoring

Reports

scope OR scoper: Search

Using account: tbeerman

Other Monitoring

Help

You are here: Rules for Subscription [ddmin, Stress\_test\_02, Replicating]

Rucio Version: 0.2.11

Type

D

DATASET

US

DATASET

US

DATASET

US

CONTAINER

MT

DATASET

US

CONTAINER

TC

DATASET

US

DATASET

US

DATASET

US

DATASET

US

DATASET

US

Type

D

Rules

Show 10 entries

Search:

Name	Creation Date	Locks OK	Locks Replicating	Locks Stuck
step14.10010.automatix_stream.reco.n.ESD.913	Tue, 03 Feb 2015 04:33:16 UTC	156	4	0
step14.10034.automatix_stream.reco.n.ESD.208	Wed, 04 Feb 2015 03:32:33 UTC	152	8	0
step14.10059.automatix_stream.reco.n.ESD.561	Wed, 04 Feb 2015 01:24:29 UTC	152	8	0
step14.1008.automatix_stream.reco.n.ESD.974	Tue, 03 Feb 2015 08:46:43 UTC	153	7	0
step14.10104.automatix_stream.reco.n.ESD.696	Wed, 04 Feb 2015 03:45:57 UTC	122	38	0
step14.10130.automatix_stream.reco.n.ESD.55	Tue, 03 Feb 2015 16:29:32 UTC	148	12	0
step14.10213.automatix_stream.reco.n.ESD.816	Tue, 03 Feb 2015 01:40:38 UTC	149	11	0
step14.1024.automatix_stream.reco.n.ESD.618	Tue, 03 Feb 2015 04:56:29 UTC	146	14	0
step14.1029.automatix_stream.reco.n.ESD.51	Wed, 04 Feb 2015 08:07:50 UTC	81	79	0
step14.10395.automatix_stream.reco.n.ESD.703	Wed, 04 Feb 2015 03:26:46 UTC	135	25	0

Showing 1 to 10 of 1,969 entries

Previous

1

2

3

4

5

...

197

Next

Account Usage Overview (in TB)

Dataset	Available Quota (TB)	Used Space (TB)	Free Space (TB)
perf-egamma	842	589	262
perf-flavtag	1,048	772	276
perf-ldtracking	130	131	0
perf-jets	1,538	1,287	250
perf-muons	403	248	155
perf-tau	1,715	1,035	689
phys-beauty	597	420	176
phys-exotics	717	591	126
phys-gener	191	119	22
phys-hi	648	479	168
phys-higgs	3,045	3,729	316
phys-sm	3,084	2,405	678
phys-susy	3,485	2,111	1,373
phys-top	2,576	1,952	623

Available Quota Used Space Free Space

© 2012-2015 European Organisation for Nuclear Research (CERN)

Wed Feb 04 2015 13:10:05 GMT+0100 (CET)

Showing 1 to 10 of 31 entries (filtered from 21,975 total entries)

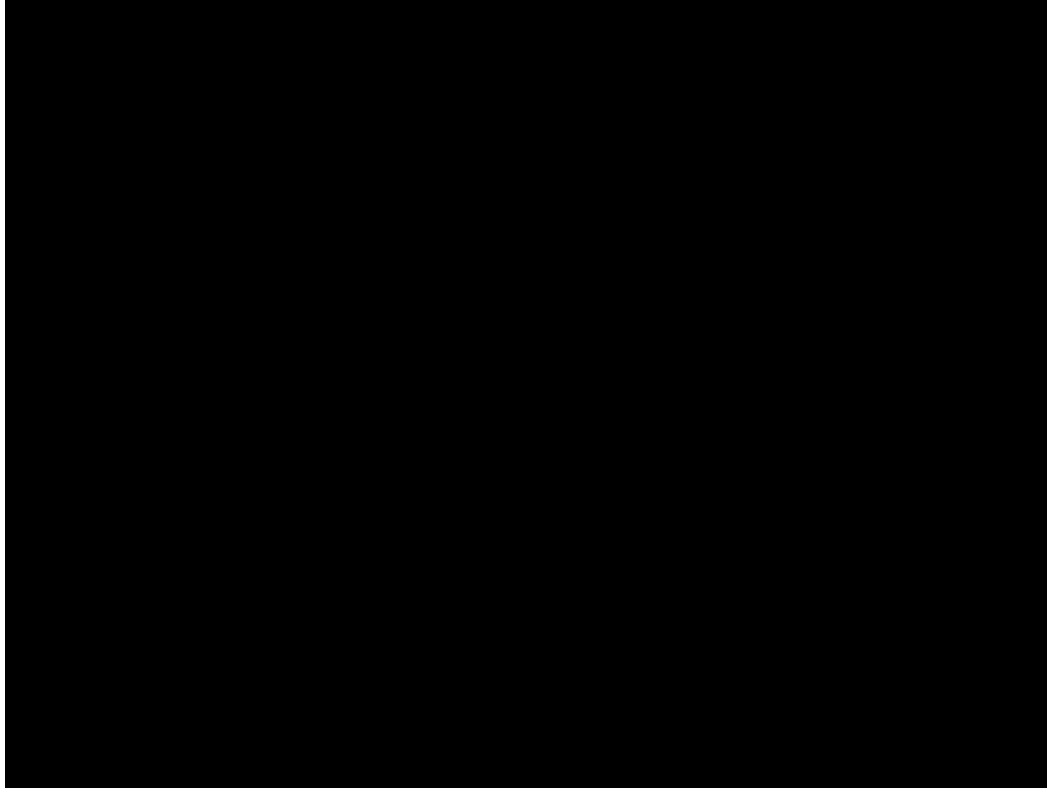
© 2012-2015 European Organisation for Nuclear Research (CERN)

Wed Feb 04 2015 10:19:37 GMT+0100 (CET)



# Rucio demo : CLI

---



# Why using an advanced system like Rucio ?

---

- Automation of some tasks :
  - Staging :
    - Panda asks Rucio to stage a dataset from TAPE
    - Rucio takes care of the staging and send a notification to Panda as soon as the files are on DISK
  - Data distribution :
    - If you plan to operate more than one site, you can specify to Rucio what type of data you want to replicate outside CERN
  - File recovery :
    - If you lose data on one site (e.g. TAPE damaged), Rucio can automatically recover it from another place
- Well interfaced with PanDA
- Easy access for the end-user to their data

# How easy is it to deploy/maintain Rucio ?

---

- A lot of work was done during the last months to ease the deployment of Rucio.
- Two deployment models possible :
  - Based on openstack infrastructure + configuration management via puppet
    - Current deployment model for ATLAS at CERN
    - ATLAS puppet templates available and can be reused by another VO
  - Based on Docker + docker-compose or Kubernetes. Being evaluated by ATLAS.
    - Should be even simpler than the previous model
- In the future, there might be the possibility to get Rucio as a Service

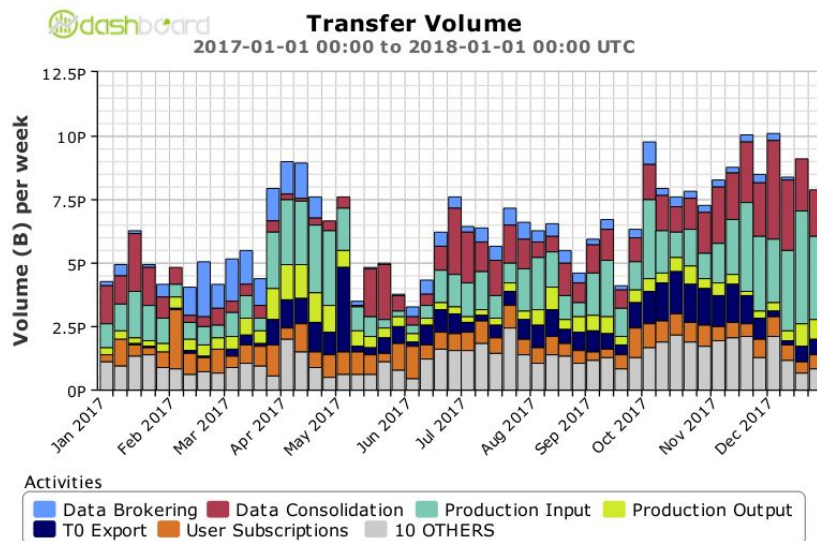


# Rucio in ATLAS

Backend: Oracle, ~1B files/340 PB, ~120 sites, 3000 users

## Big instance

- ATLAS DDM central team operates with only 2 FTEs + shifters
  - Identify problems and communicate with the sites
  - Provide feedback to Rucio developers (ops driven d
  - Provide user support
  - Evolve the replication policies (e.g. number of
  - Configure and run Rucio services



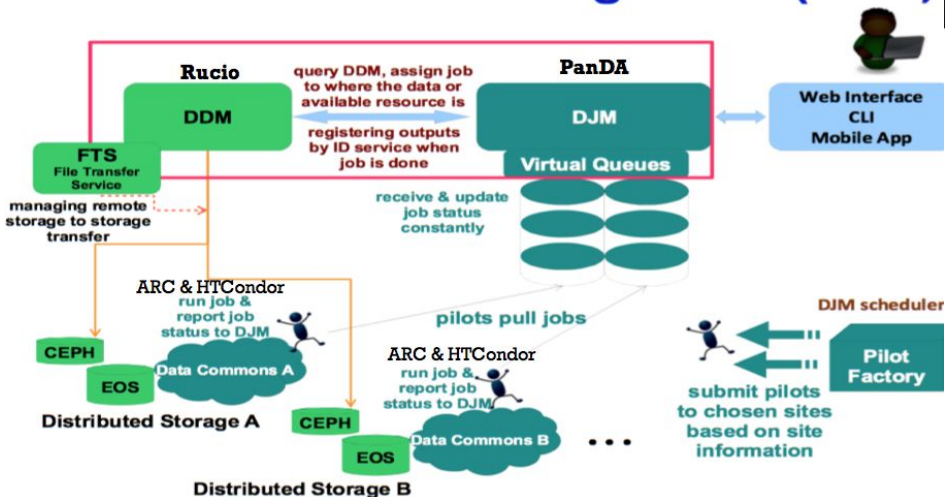
# Alpha Magnetic Spectrometer (AMS)



Backend: Mysql, ~ M files, ~10 sites, TW (ASGC) collaboration

Medium instance

## Distributed Job Management (DJM)



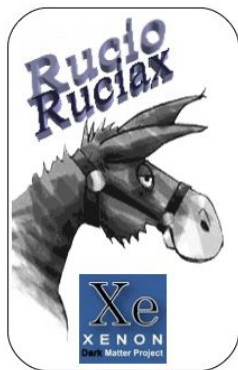
## Web Interface

The screenshot shows the **DICOS Web Interface** with a navigation bar (About, Job Submission, Job Monitoring, Data Management, Wiki, API, Contact, Terminal) and a user profile (Chi Hsun Wu). The **Data Management** section includes buttons for **Upload**, **Delete**, **Download**, **Detach**, and **Search**. Below these is a file browser for **twgrid-user-chwu** showing a directory tree with files like **5\_data**, **job**, **JobOutput**, and various log files. A **Get link** button is present, and a **Direct Links** dialog box is open, displaying a list of links. A red text overlay **Drag to attach** is at the bottom.

# XENON1T

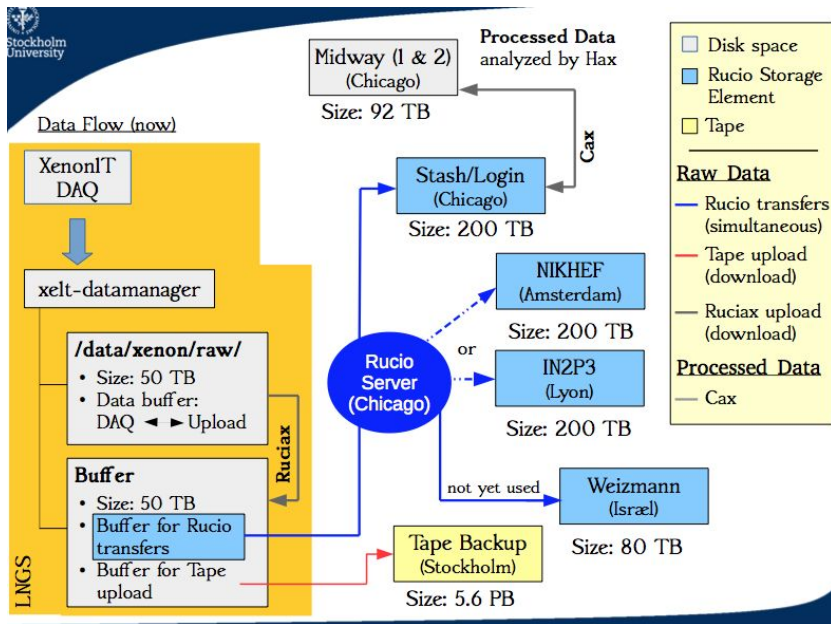
Backend: MariaDB, ~K files, ~10 sites, collaboration with University of Chicago and Stockholm

Small instance



## → Documentation & Installation notes:

- [Rucio Installation Notes](#)
- [Xenon1T/MWT2 Installation Notes](#)



# Support

---

- Support provided by the Rucio team
- To reach us :
  - [rucio-dev@cern.ch](mailto:rucio-dev@cern.ch) : Only developers
  - <https://rucio.slack.com/messages/#support/> : All developers are on it + many people of the other VO using or evaluating Rucio (~33 people). This is the recommended way to ask for help

# Conclusion

---

- Rucio is not only dedicated to big collaborations, but can also serve smaller ones
- Can simplify/automate some operations
- Close integration with Panda
- Easy interface for the end-users to access data

# More information

---

Website <http://rucio.cern.ch>

Documentation <https://rucio.readthedocs.io>



Repository <https://github.com/rucio/>



Continuous Integration <https://travis-ci.org/rucio/>



Images <https://hub.docker.com/r/rucio/>



docker

Online support <https://rucio.slack.com/messages/#support/>



Developer contact [rucio-dev@cern.ch](mailto:rucio-dev@cern.ch)