ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Event Driven Processing and Data Management

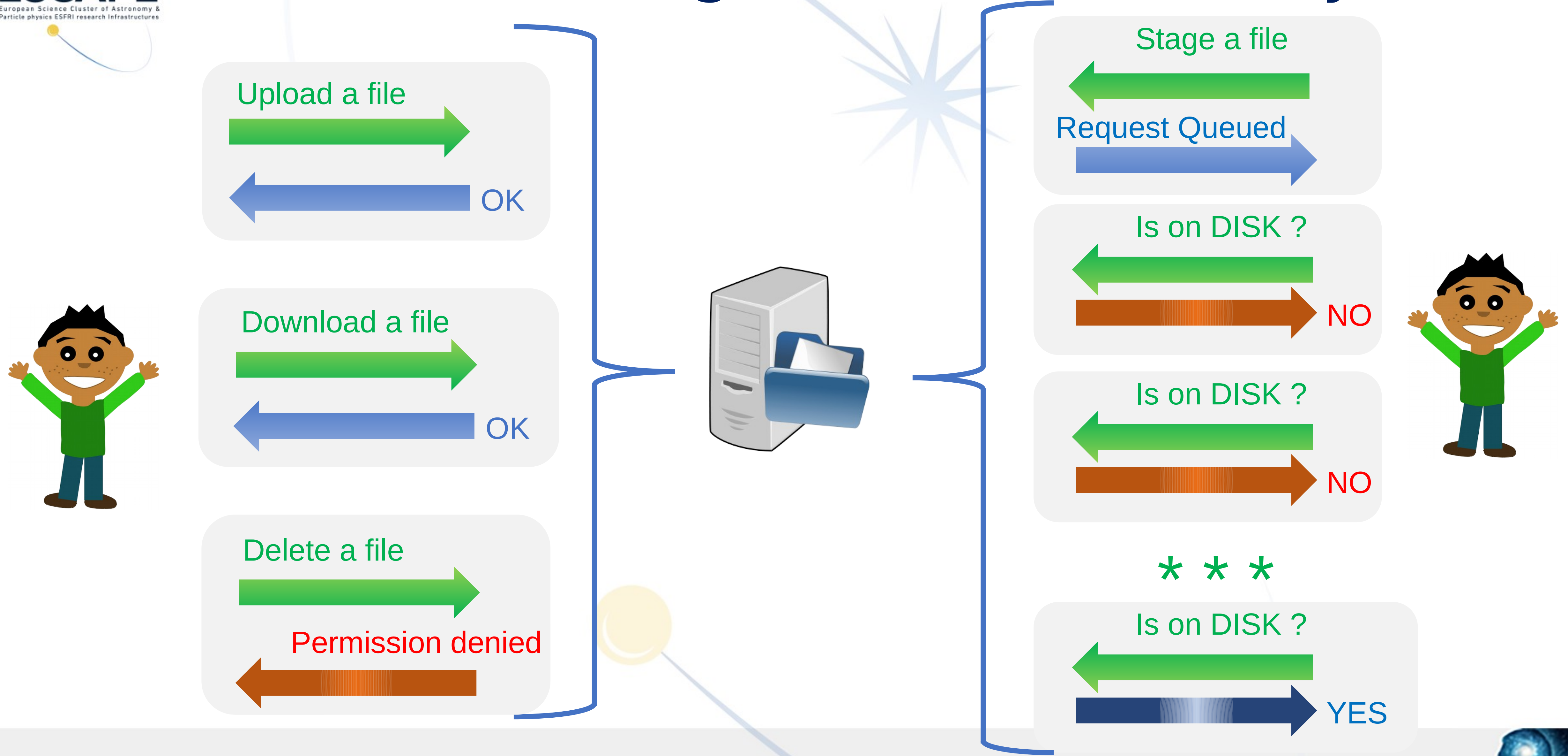
Paul Millar

Joint ESCAPE WP2/WP5 workshop

(with material donated by Marica Antonacci, Patrick Fuhrmann and Michael Schuh)

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

How storage is used currently



Upload a file

OK

Download a file

OK

Delete a file

Permission denied

Stage a file

Request Queued

Is on DISK ?

NO

Is on DISK ?

NO

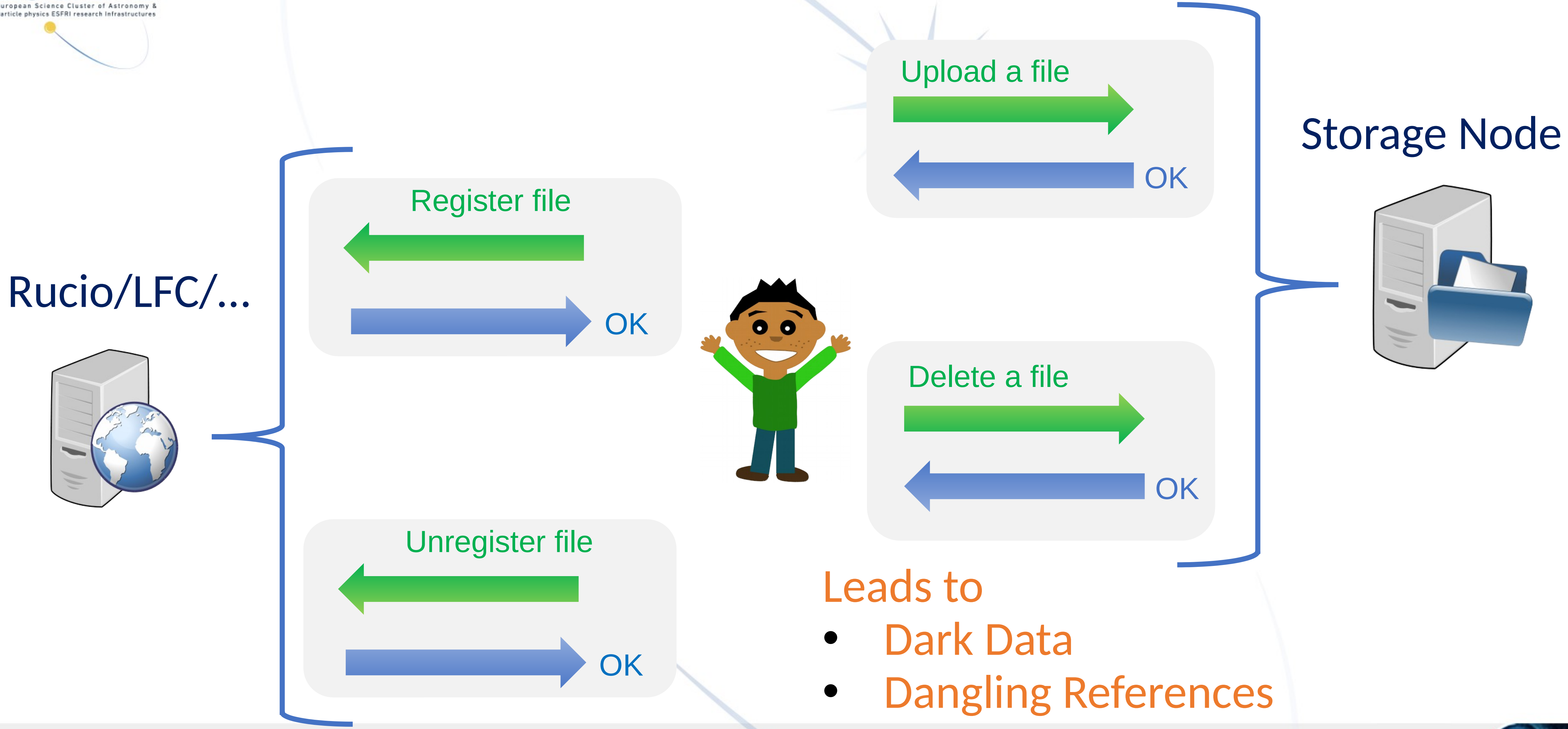
* * *

Is on DISK ?

YES

Funded by the European Union's
Horizon 2020 - Grant N° 824064

This may lead to problems…

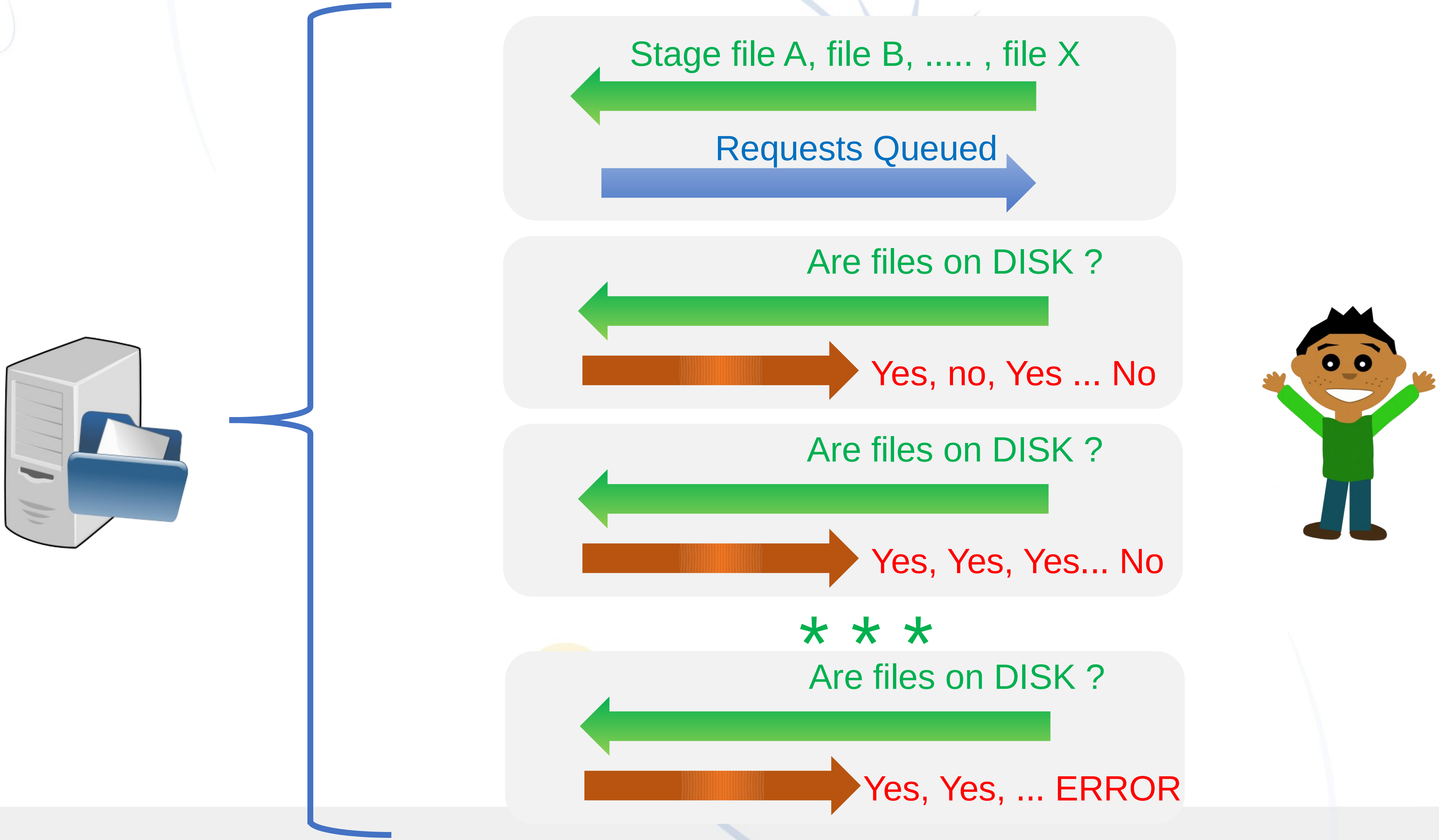


Leads to

- Dark Data
- Dangling References

Funded by the European Union's
Horizon 2020 - Grant N° 824064

This may cause more problems



Stage file A, file B, ... , file X

Requests Queued

Are files on DISK ?

Yes, no, Yes ... No

Are files on DISK ?

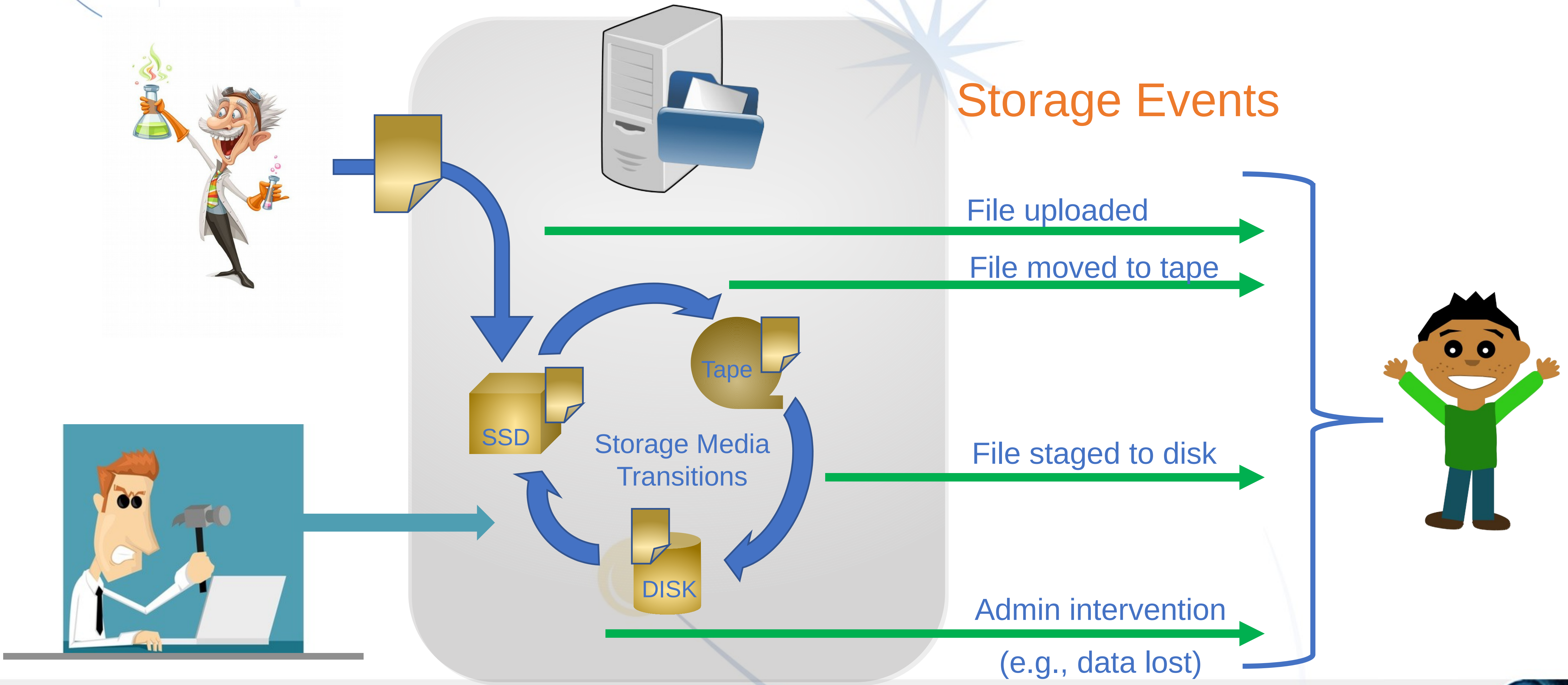
Yes, Yes, Yes... No

* * *

Are files on DISK ?

Yes, Yes, ... ERROR

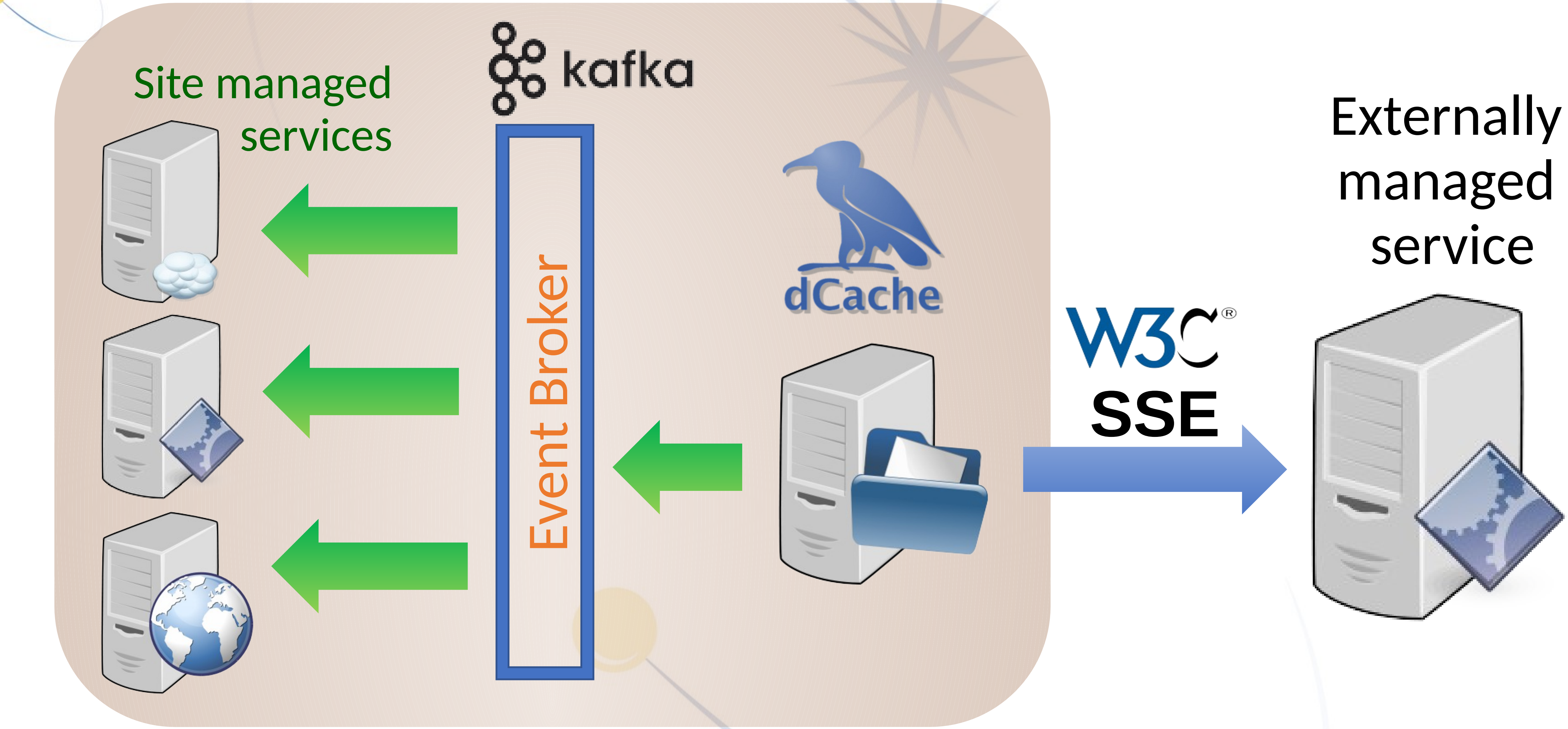
New way of interacting: storage events



Funded by the European Union’s
Horizon 2020 - Grant N° 824064

dCache implementation

dCache Storage Events: Kafka and SSE



Funded by the European Union's Horizon 2020 - Grant N° 824064

Cheat sheet: Kafka vs SSE

	kafka	W3C® SSE
Availability since...	dCache v4.1	dCache v5.0
Standard …	Software package	Protocol
What events does it see?	dCache billing events	inotify
Main benefit	Easy integration	Built-in security
“Catch-up” storage	Memory & disk	Memory-only (currently)
Target audience	Site-level integration	Events for users

Use-cases and demonstrators

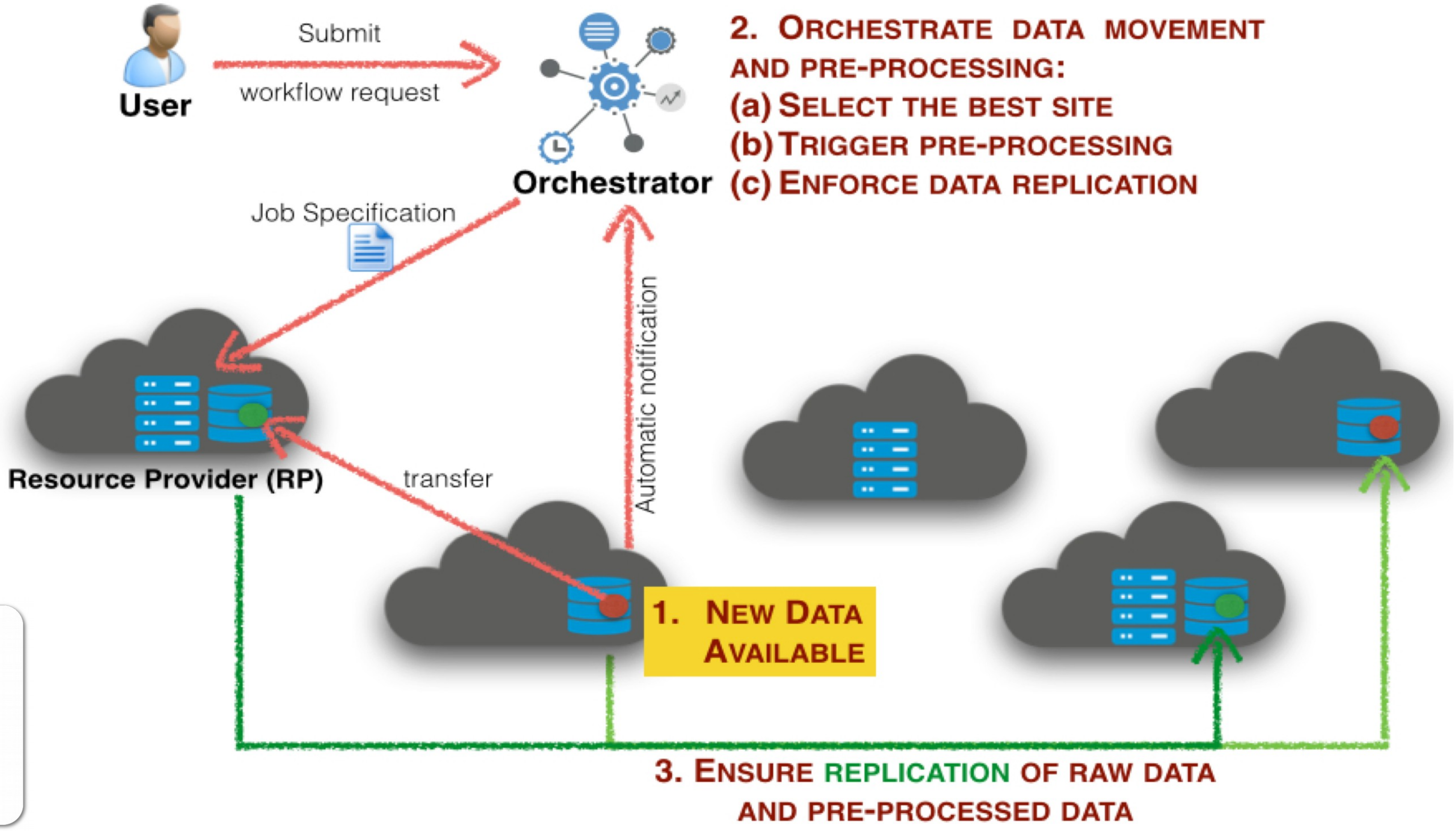
Use-cases and demonstrators

INDIGO-Orchestrator

& automated data processing

Funded by the European Union's
Horizon 2020 - Grant N° 824064

INDIGO Orchestrator (SSE)



2. Orchestrate data movement and pre-processing:
(a) Select the best site
(b) Trigger pre-processing
(c) Enforce data replication

1. New Data Available

3. Ensure replication of raw data and pre-processed data

Use-cases and demonstrators

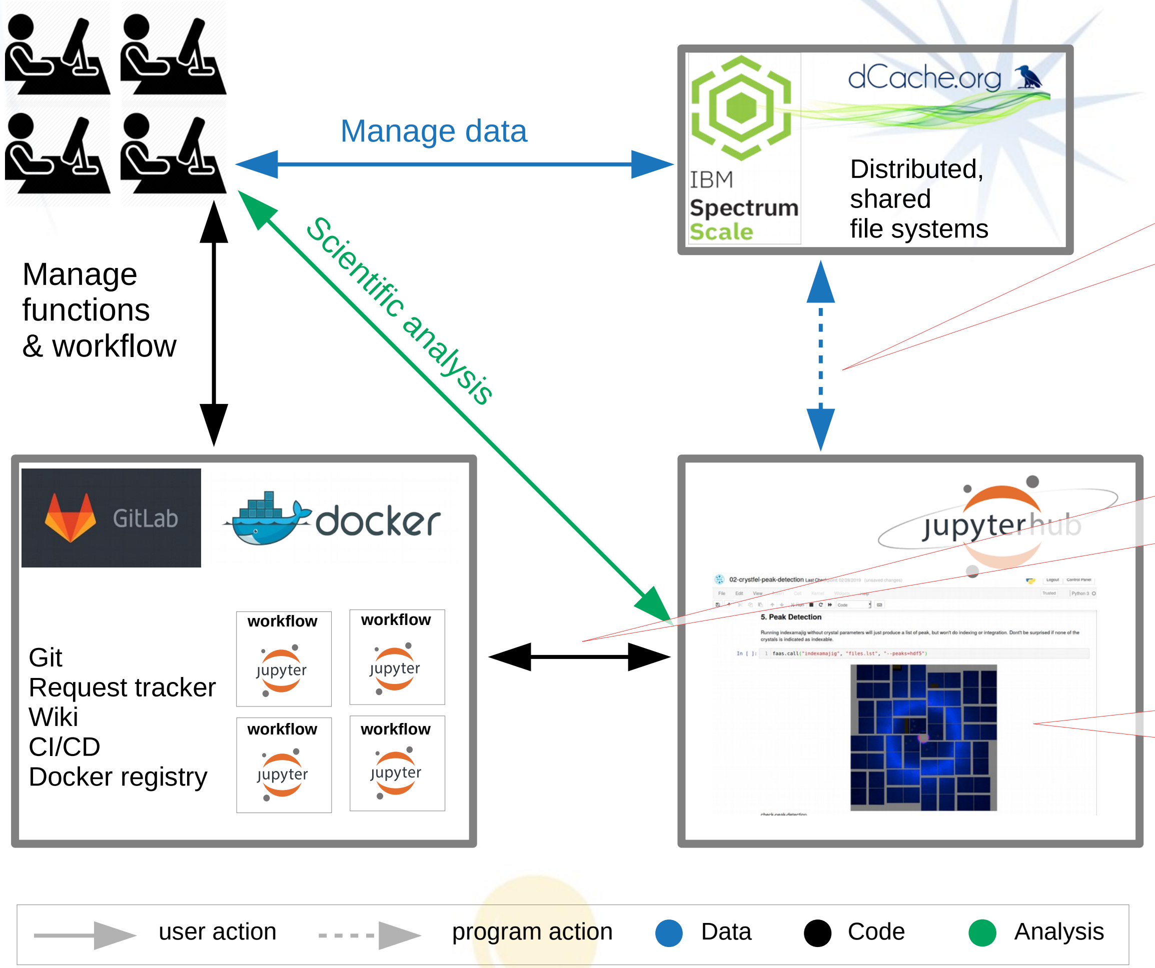
EU-XFEL

Analysis and automation

pilot platform pilot

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Jupyter Notebooks on HPC cluster



- Data on shared file systems mounted on HPC host (NFS)
- uid, gid mapping

Sync git and Jupyter

- Mybinder
- Nbgitpuller
- git

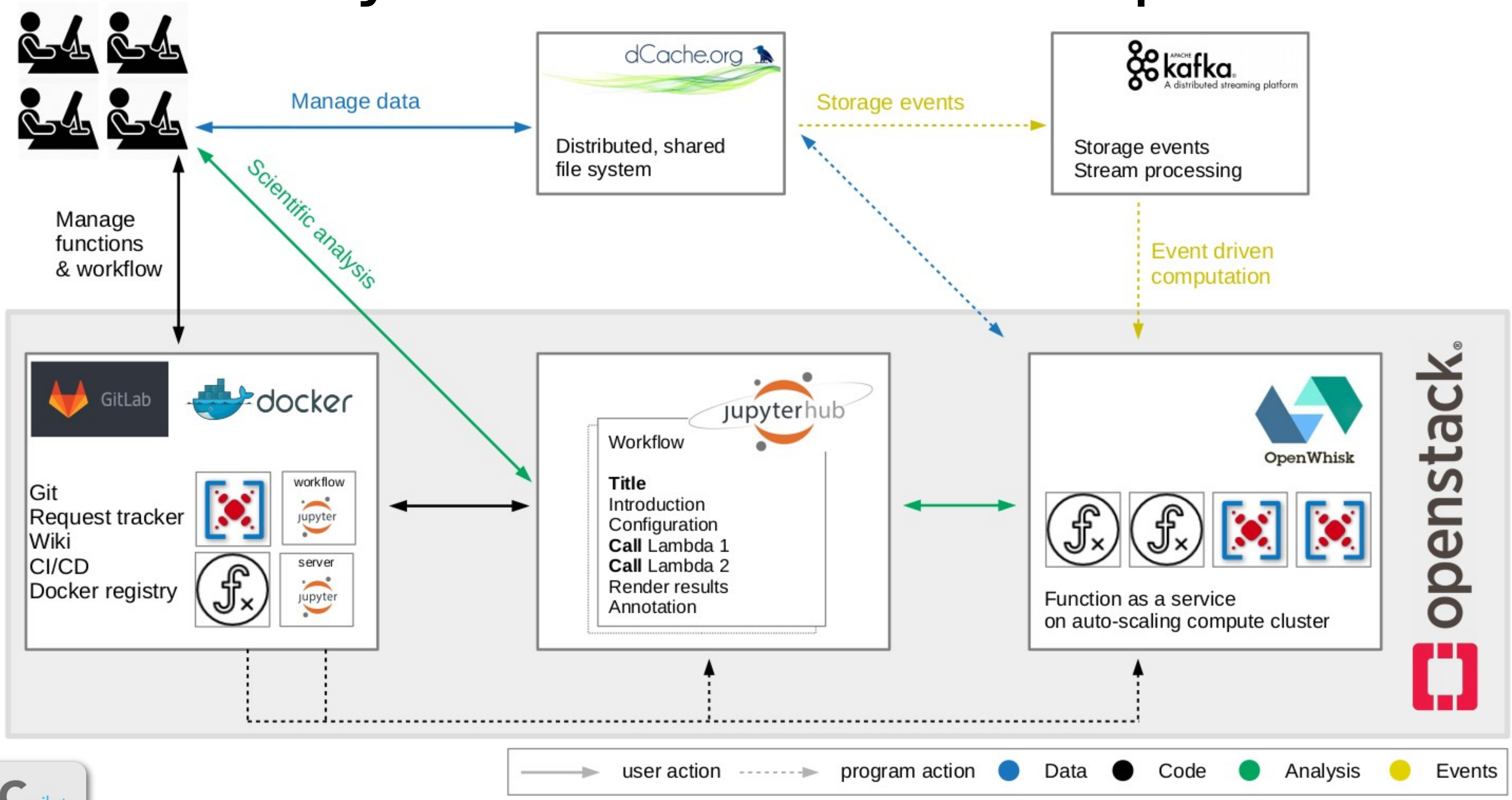
Reproducibility

- Functions
- Results

Contact: eosc-pan-info@desy.de
Icons: flaticon.com (freepik/prettycon)

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Analysis and automation platform



Manage data

Distributed, shared file system

Storage events

Storage events
Stream processing

Event driven computation

Manage functions & workflow

Scientific analysis

Git
Request tracker
Wiki
CI/CD
Docker registry

Workflow

Title
Introduction
Configuration
Call Lambda 1
Call Lambda 2
Render results
Annotation

Function as a service on auto-scaling compute cluster

→ user action ⇢ program action ● Data ● Code ● Analysis ● Events

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

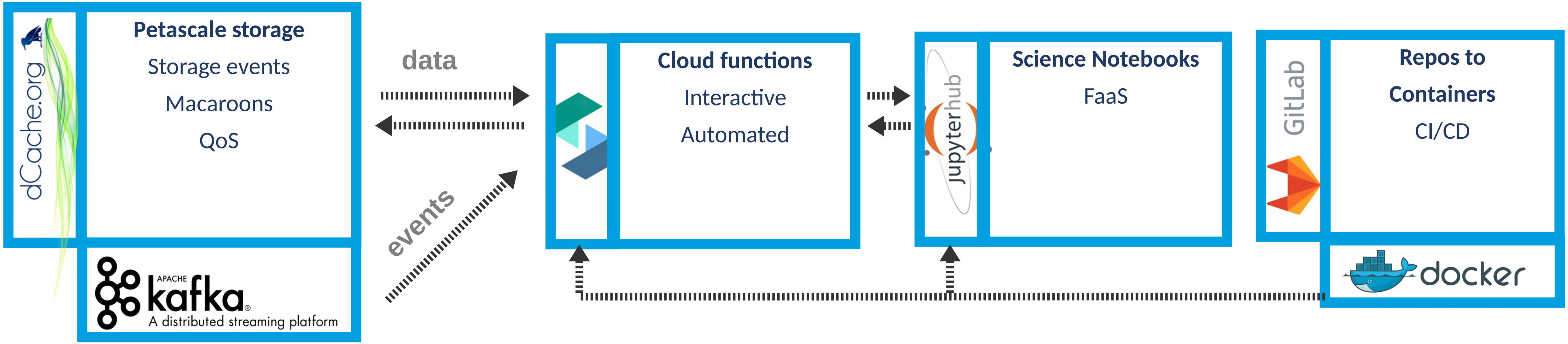
Analysis and automation platform

Single namespace in multi-clouds.

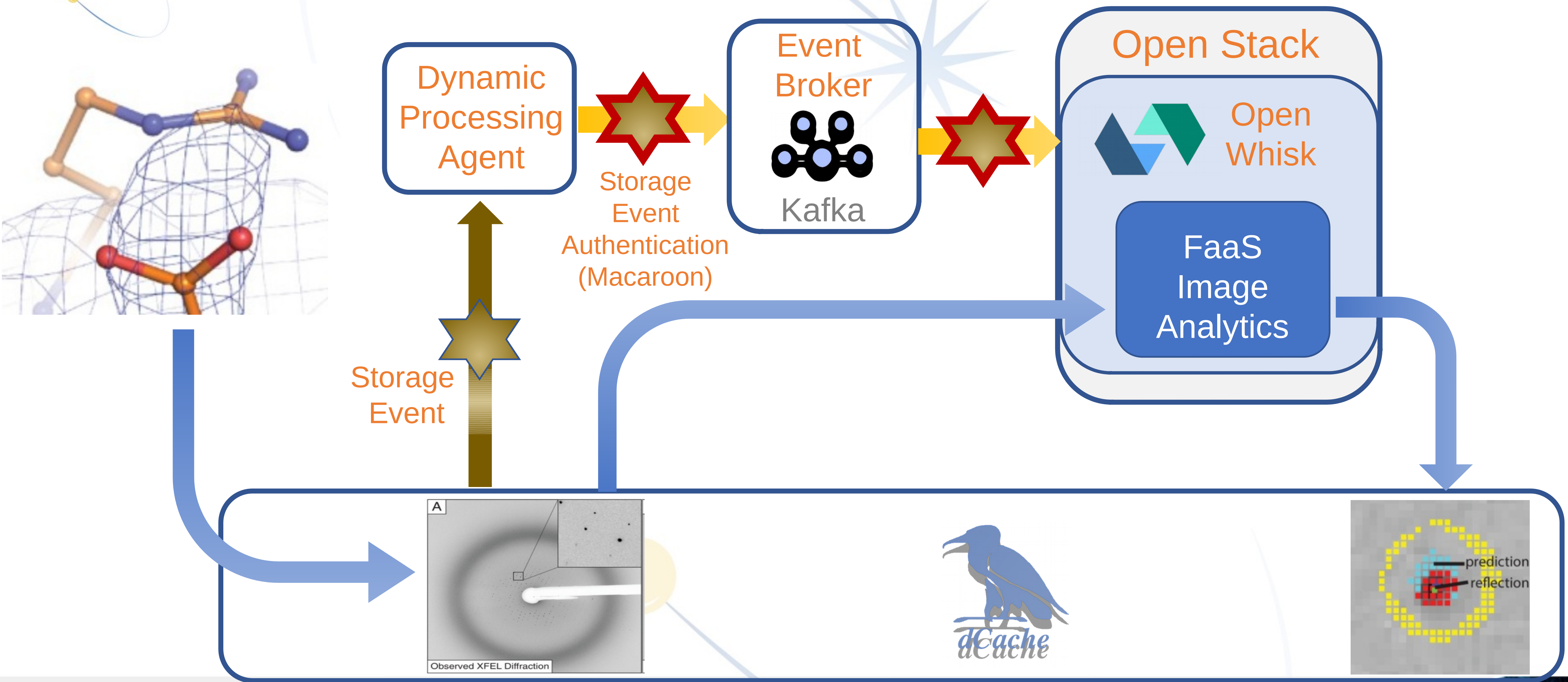
Function-as-a-Service in Science Notebooks and in automation.

Jupyter Notebooks in user-defined environments.

Just push code it builds, goes live and scales.



System/integration view



Dynamic Processing Agent

Storage Event Authentication (Macaroon)

Event Broker

Kafka

Open Stack

Open Whisk

FaaS Image Analytics

Storage Event

Funded by the European Union's
Horizon 2020 - Grant N° 824064

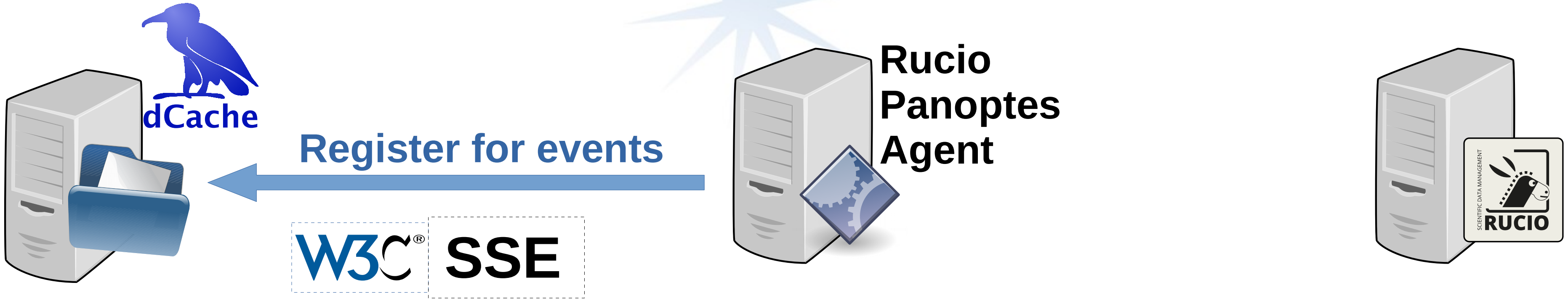
Use-cases and demonstrators

EISCAT 3D

Automated replication pilot

Funded by the European Union's
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication



dCache

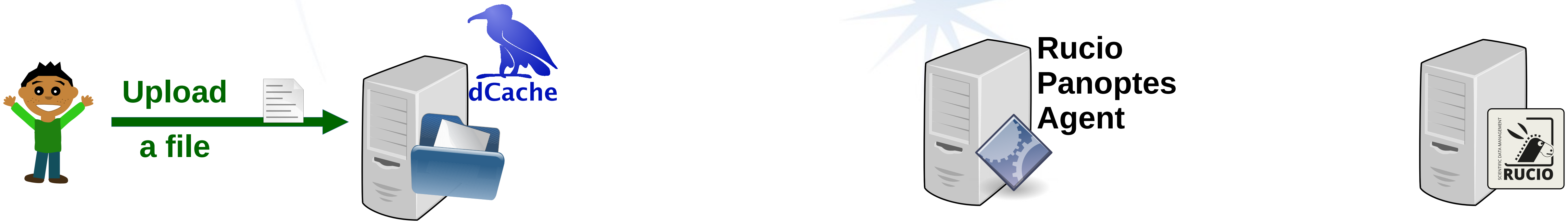
Register for events

W3C SSE

Rucio
Panoptes
Agent

Funded by the European Union's
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication



Upload

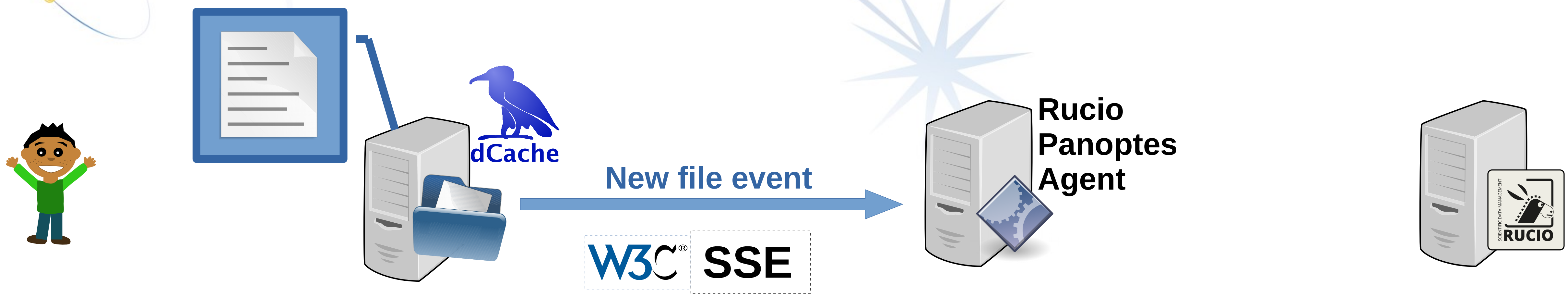
a file

dCache

Rucio
Panoptes
Agent

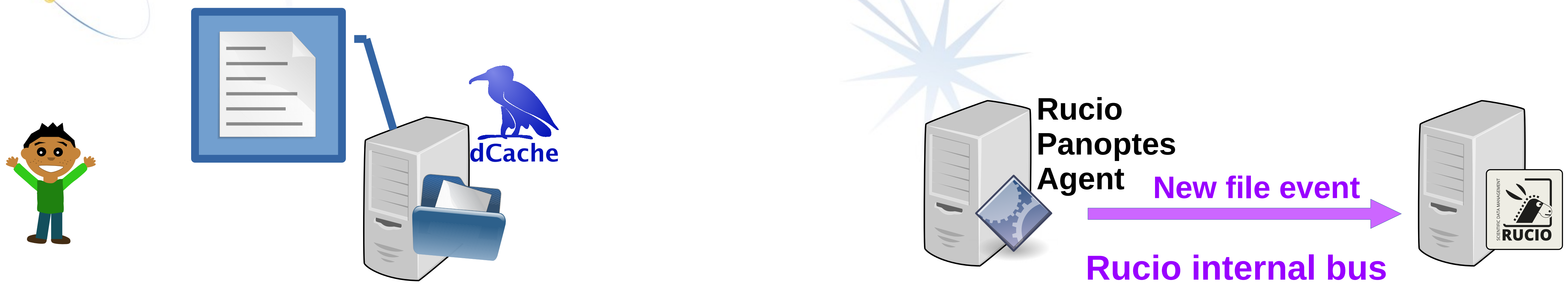
Funded by the European Union's
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication



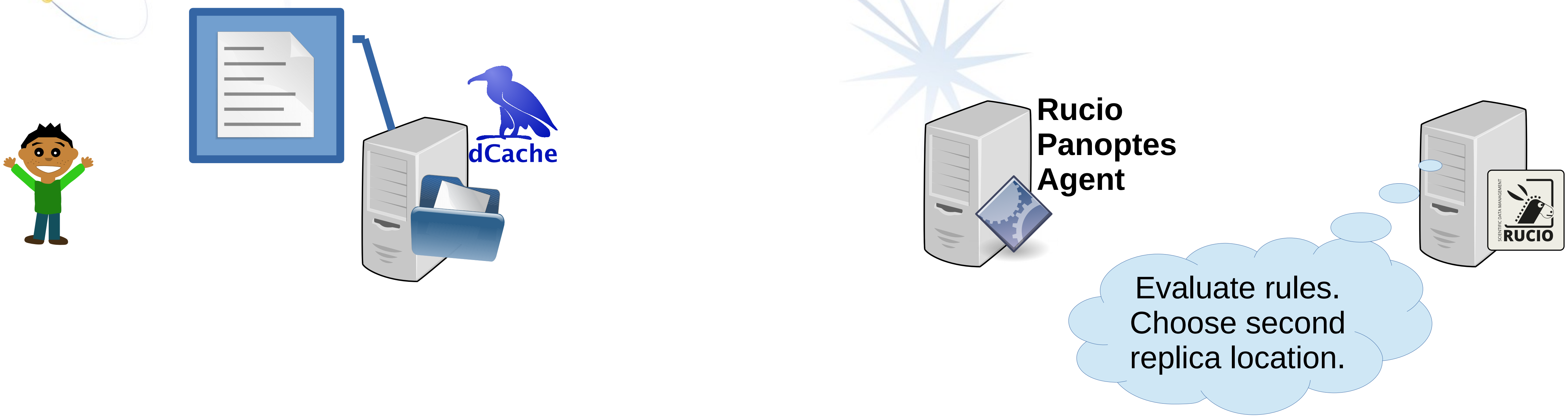
Funded by the European Union's
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication



Funded by the European Union’s
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication

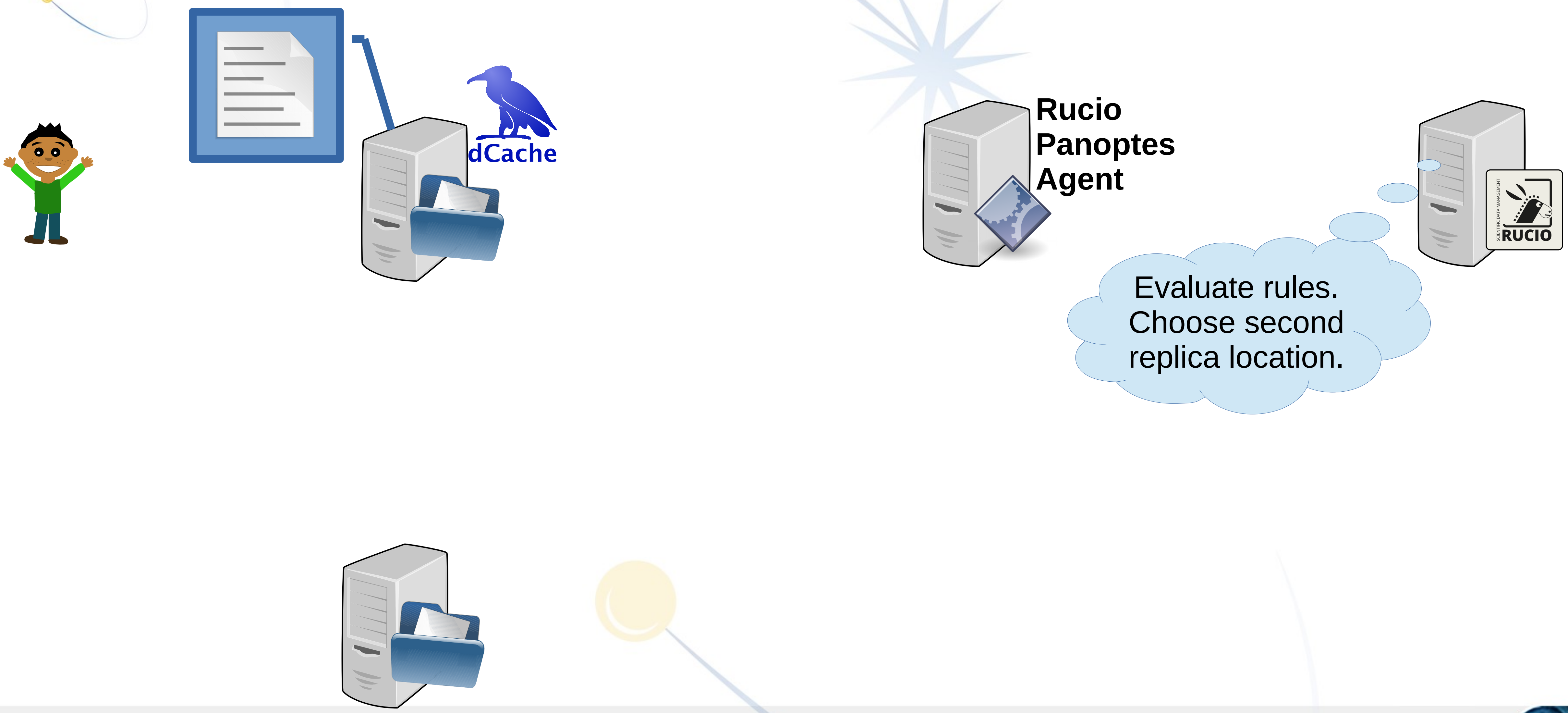


Rucio
Panoptes
Agent

Evaluate rules.
Choose second
replica location.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

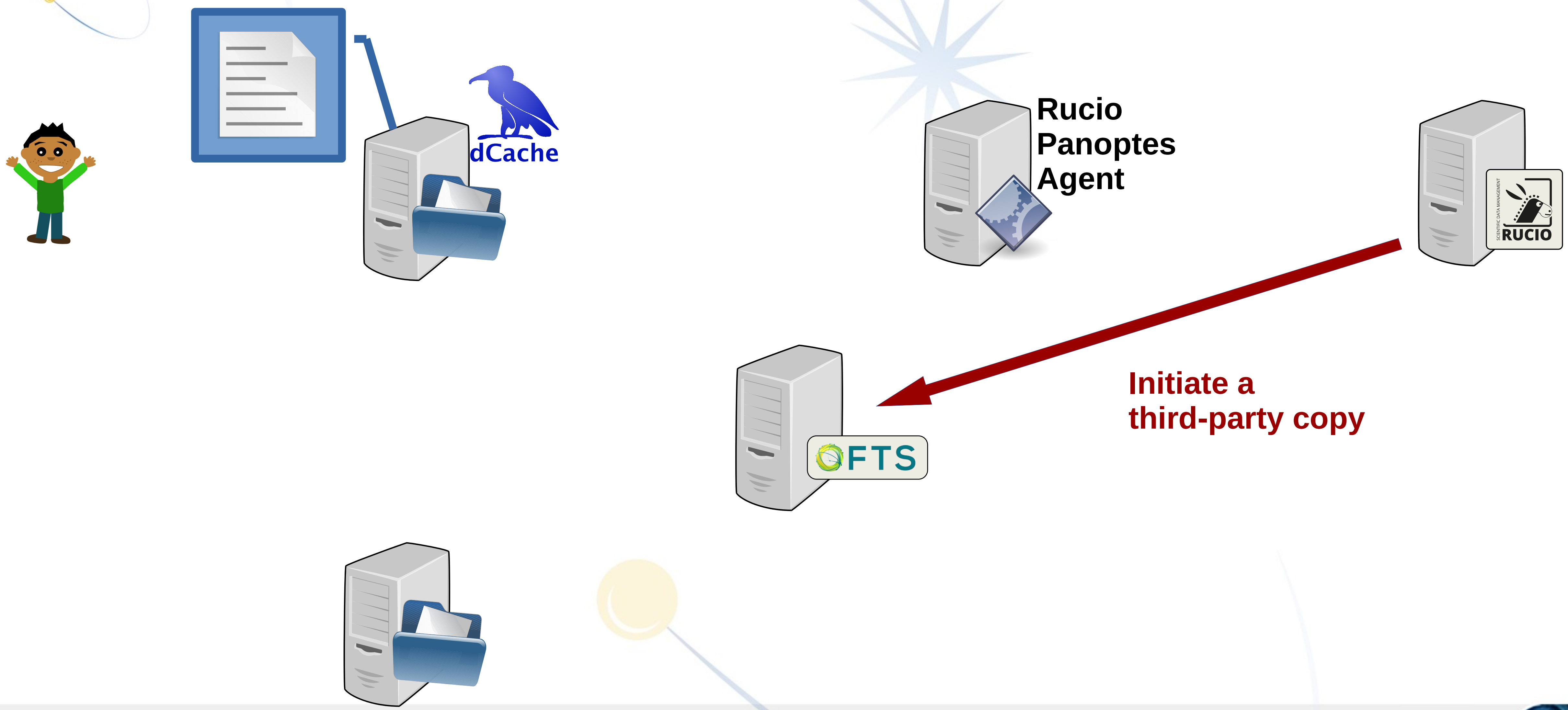
EISCAT_3D: automated replication



Evaluate rules. Choose second replica location.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

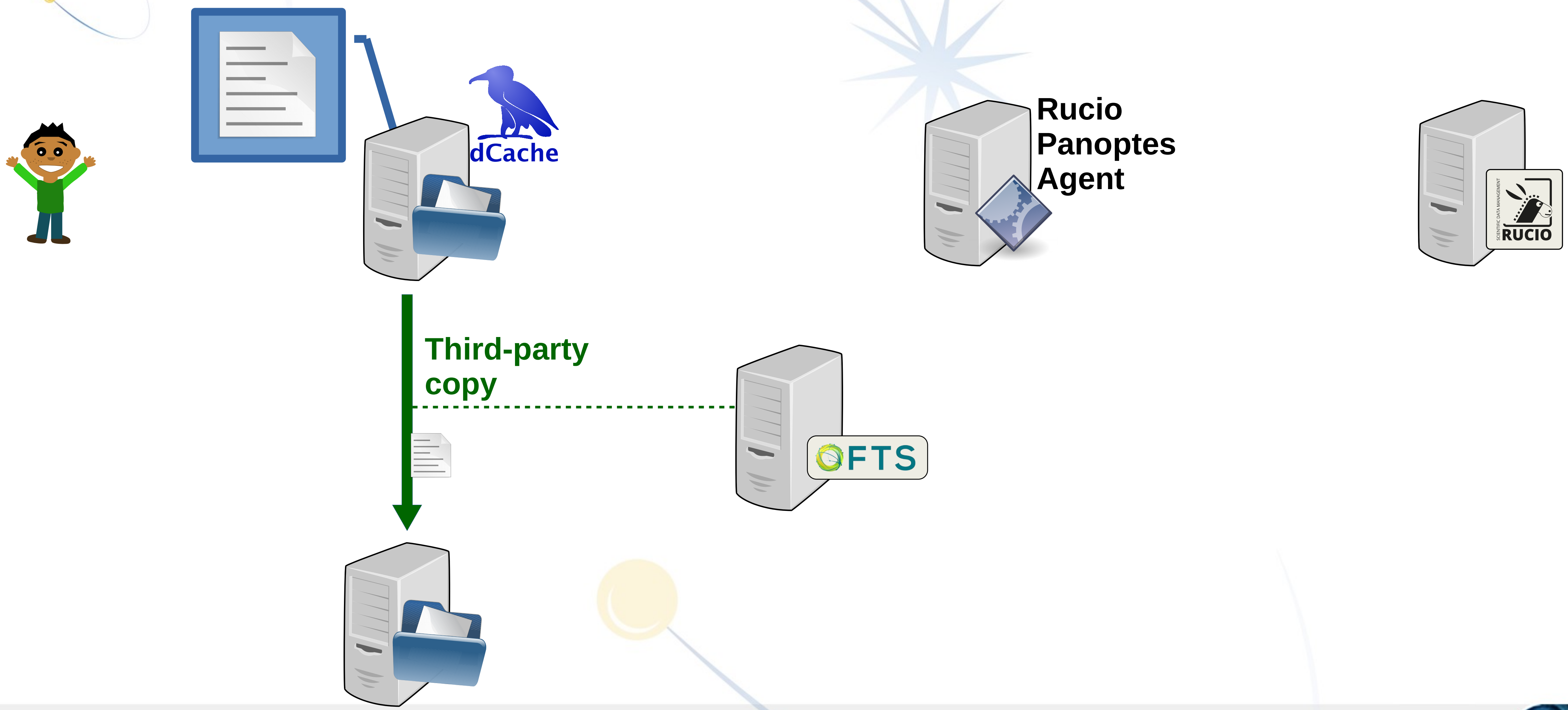
EISCAT_3D: automated replication



Initiate a third-party copy

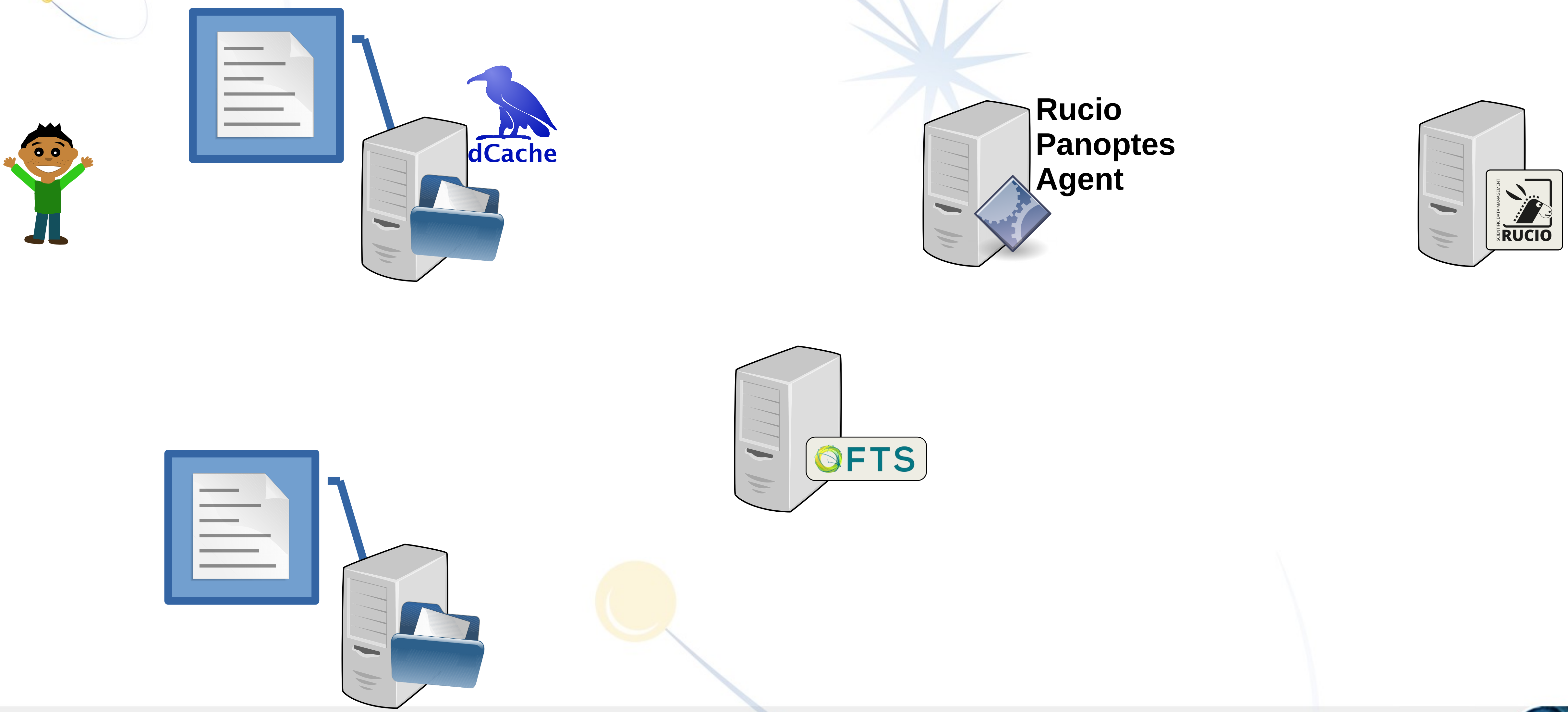
Funded by the European Union's
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication



Funded by the European Union's
Horizon 2020 - Grant N° 824064

EISCAT_3D: automated replication



Rucio
Panoptes
Agent

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Use-cases and demonstrators

Fermilab & WLCG

Increased tape staging efficiency

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Increased tape staging efficiency

- One of the closely correlated metric for tape inefficiency is **remount count**: loading same tape two (or more) times in close succession.

  Load tape once. Read all relevant data. Move on.

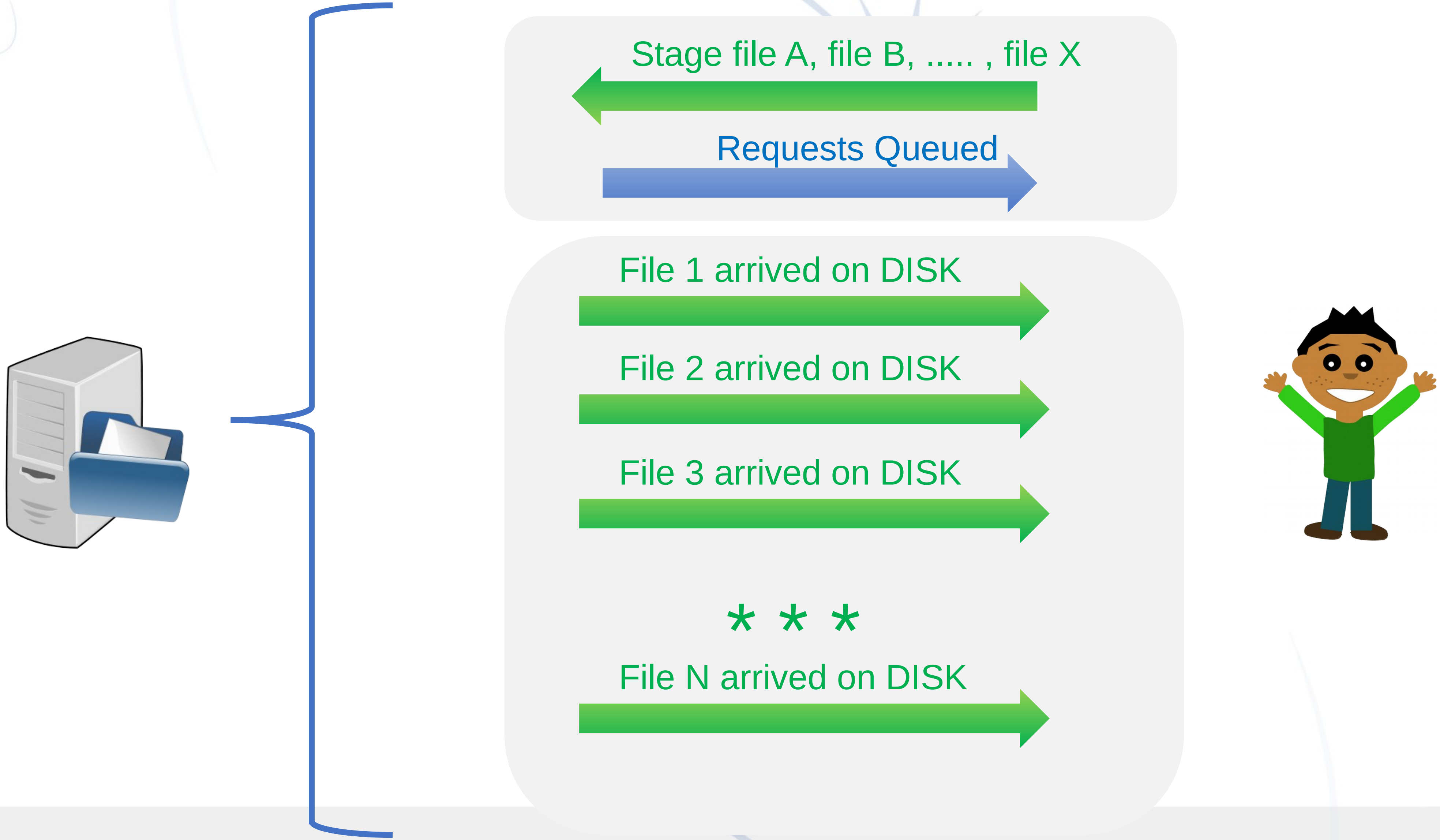
- To reduce likelihood, the tape system should be given (ideally) **all pending requests**.

  This allows the tape system to reorder requests.

- **Problem**: polling overhead from repeatedly checking status.
- **Solution**: storage events to discover when files have been staged
  - Allow clients to request all files (almost without limit).
  - Low latency: jobs/transfers may start as soon as file is available.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Increased tape staging efficiency



Stage file A, file B, ... , file X

Requests Queued

File 1 arrived on DISK

File 2 arrived on DISK

File 3 arrived on DISK

* * *

File N arrived on DISK

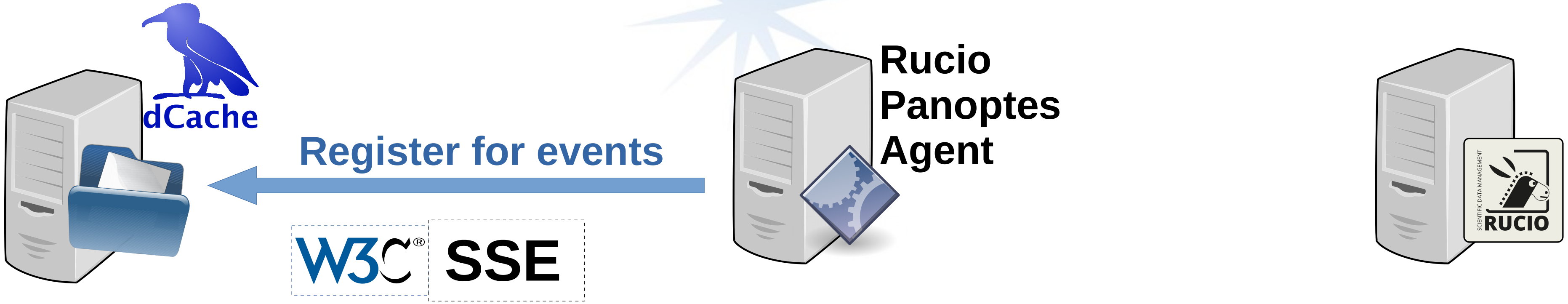
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Use-cases and demonstrators

Handling data-loss:
Automated data-replica recovery

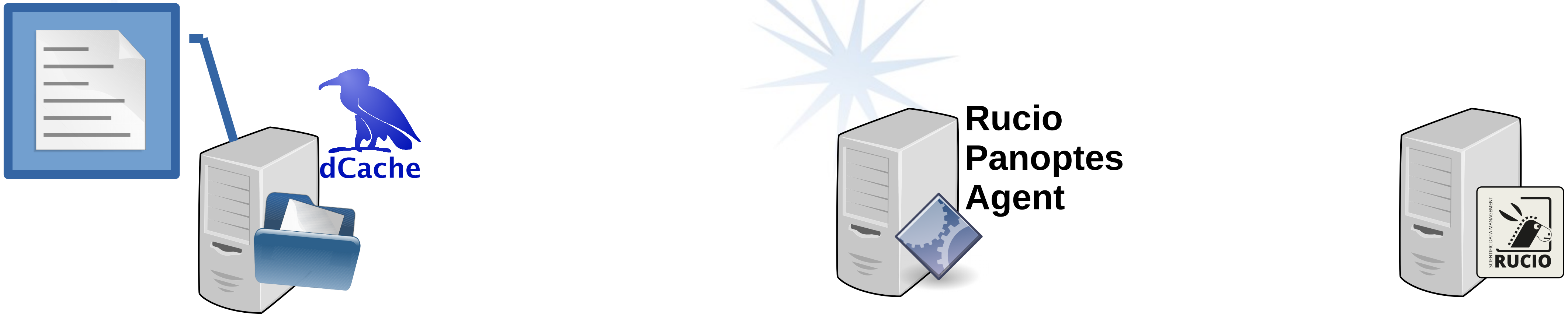
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Automatic data-replica recovery



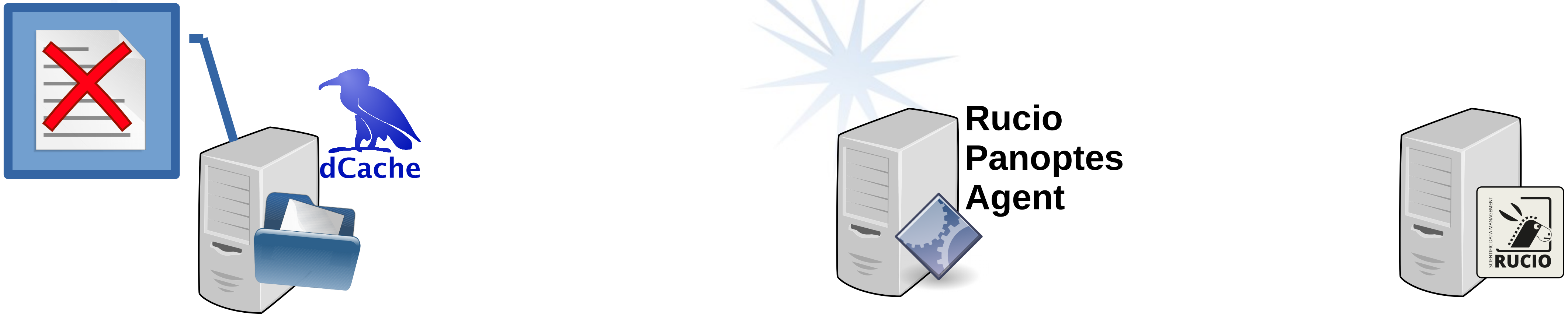
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Automatic data-replica recovery

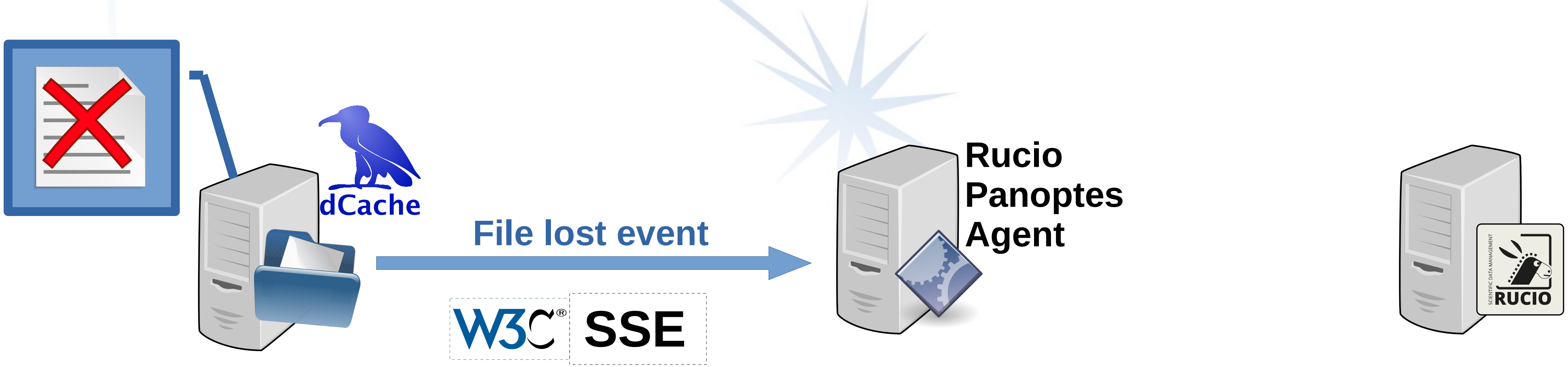


Rucio
Panoptes
Agent

Automatic data-replica recovery

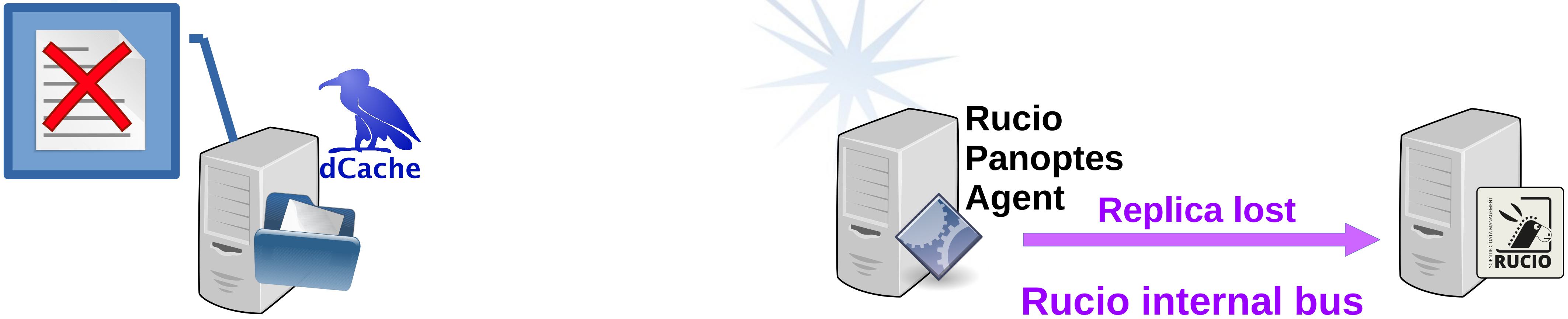


Automatic data-replica recovery



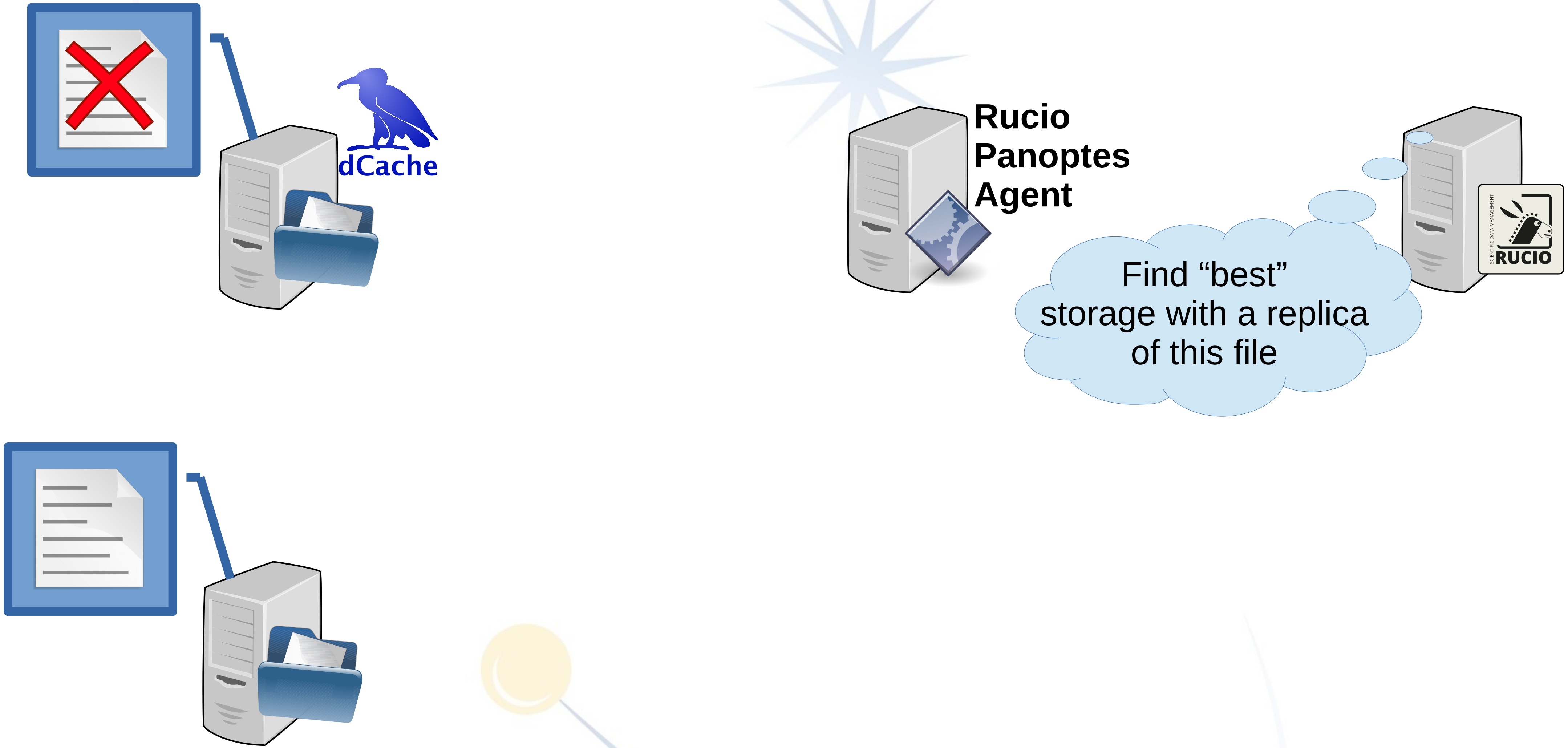
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Automatic data-replica recovery



Funded by the European Union's
Horizon 2020 - Grant N° 824064

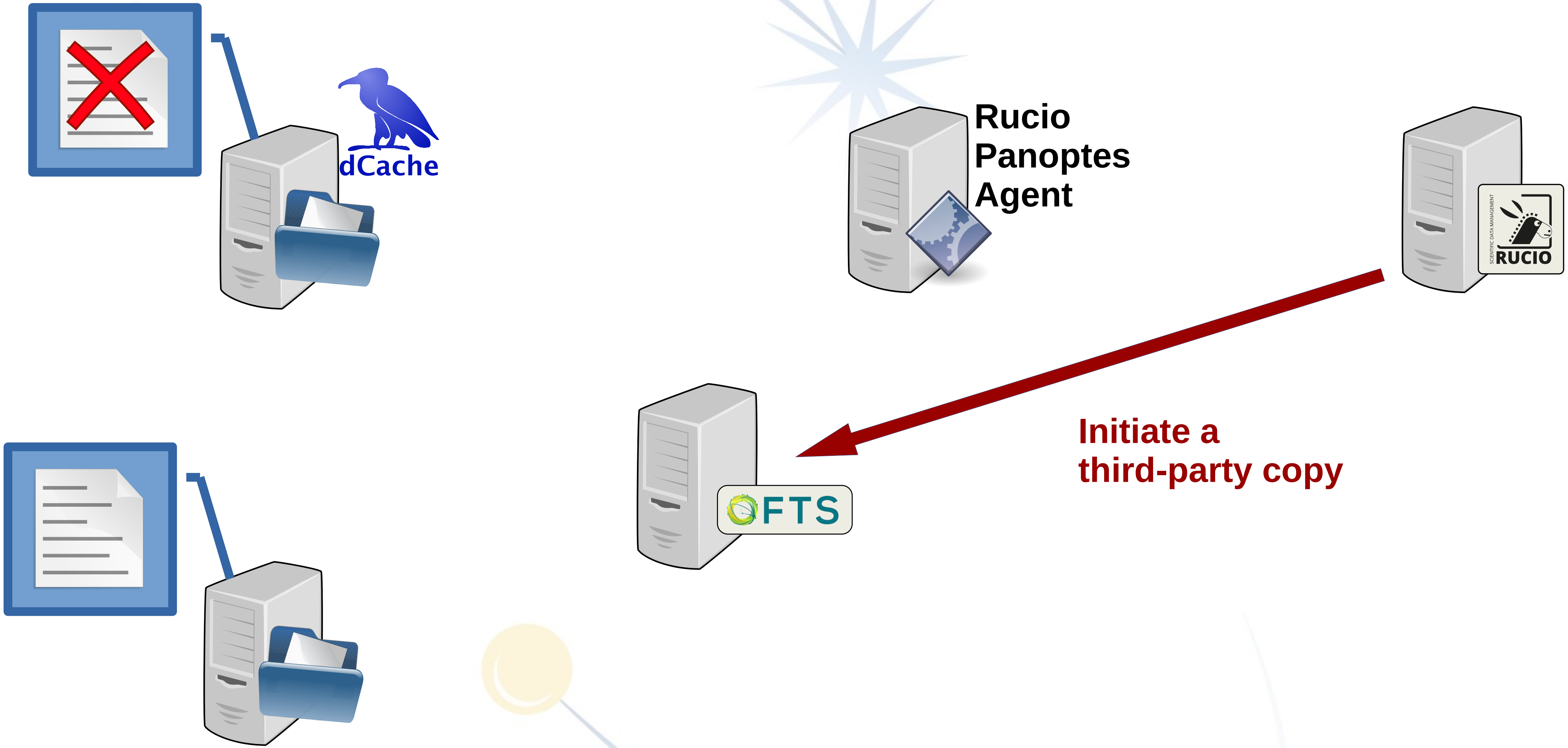
Automatic data-replica recovery



Find “best” storage with a replica of this file

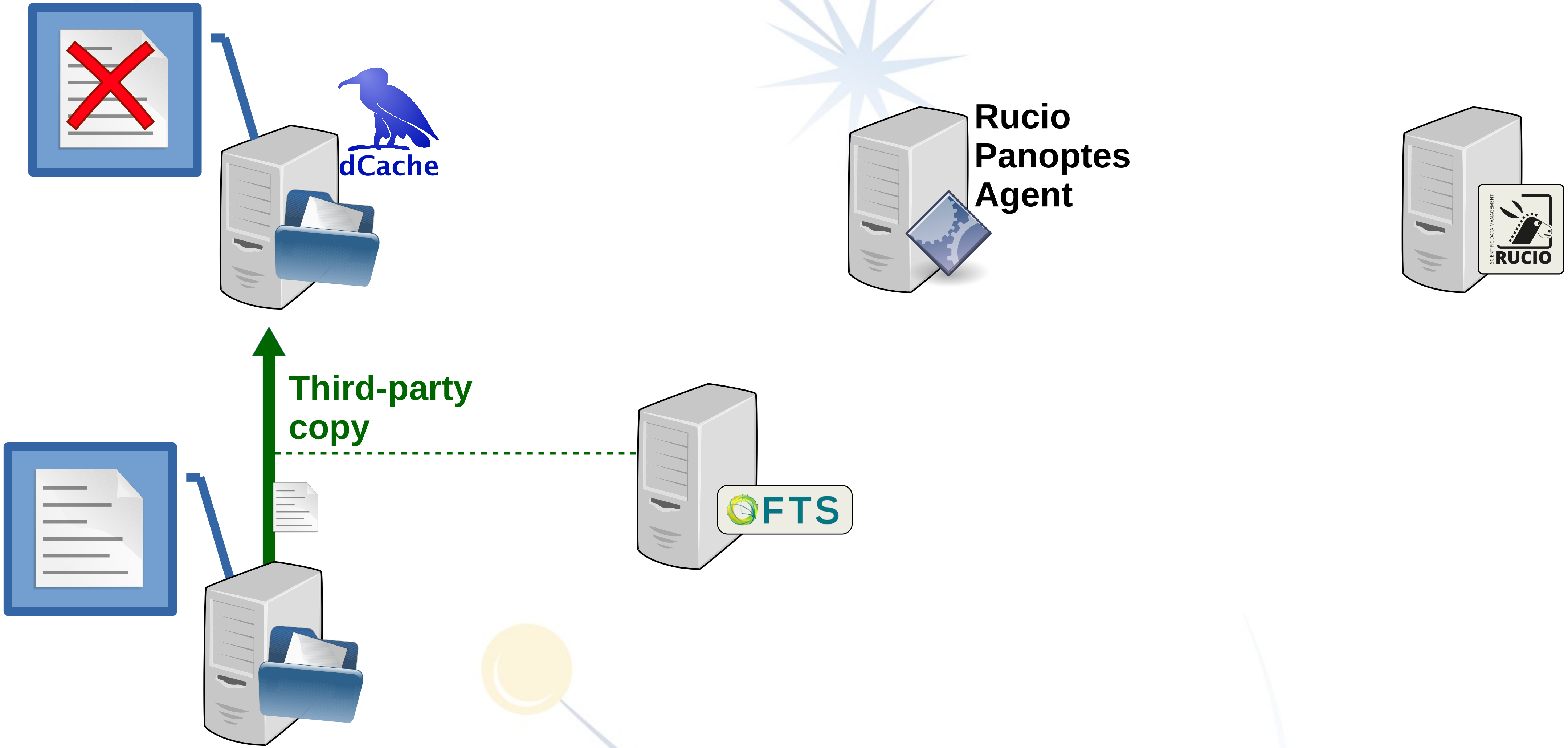
Funded by the European Union’s Horizon 2020 - Grant N° 824064

Automatic data-replica recovery



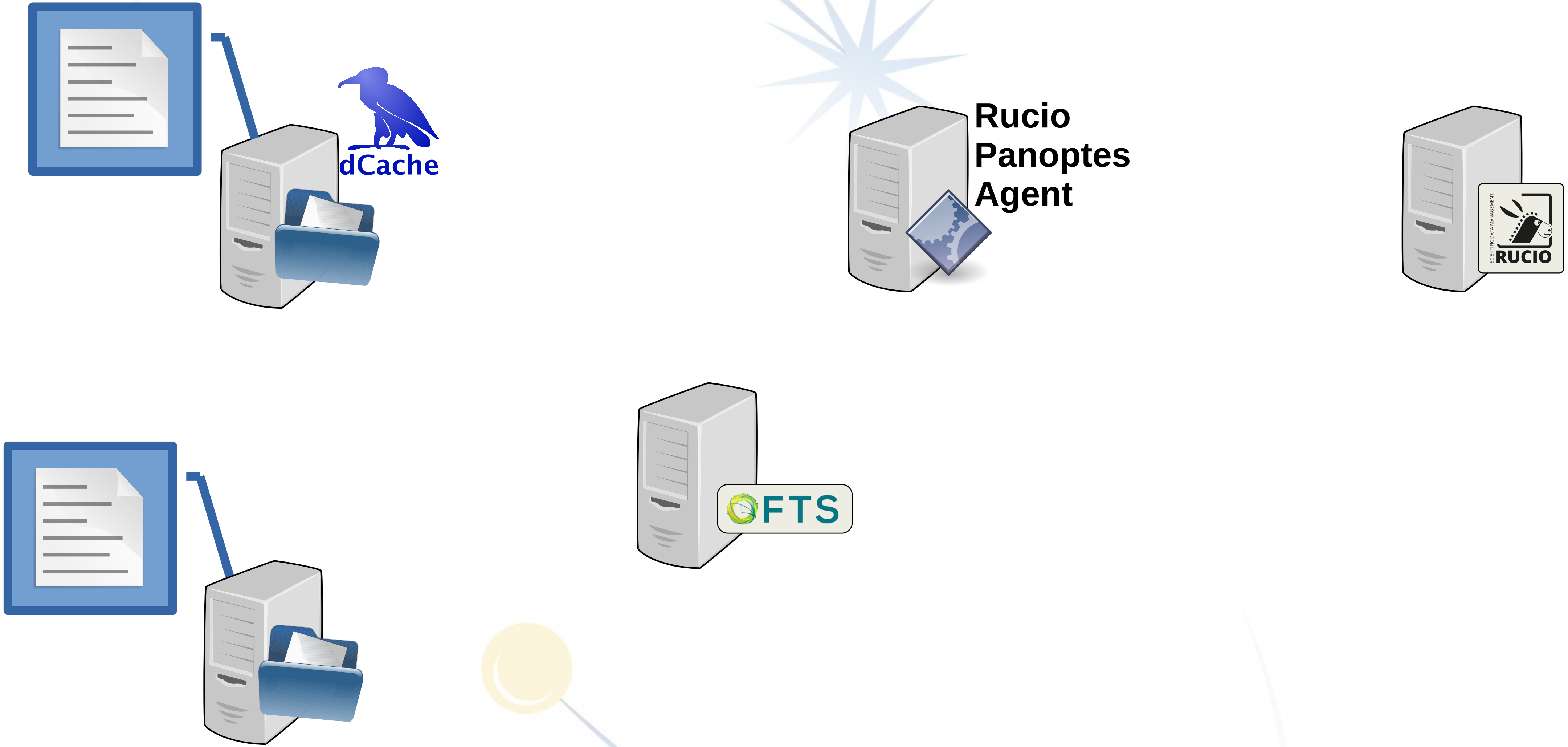
Initiate a
third-party copy

Automatic data-replica recovery



Funded by the European Union's
Horizon 2020 - Grant N° 824064

Automatic data-replica recovery



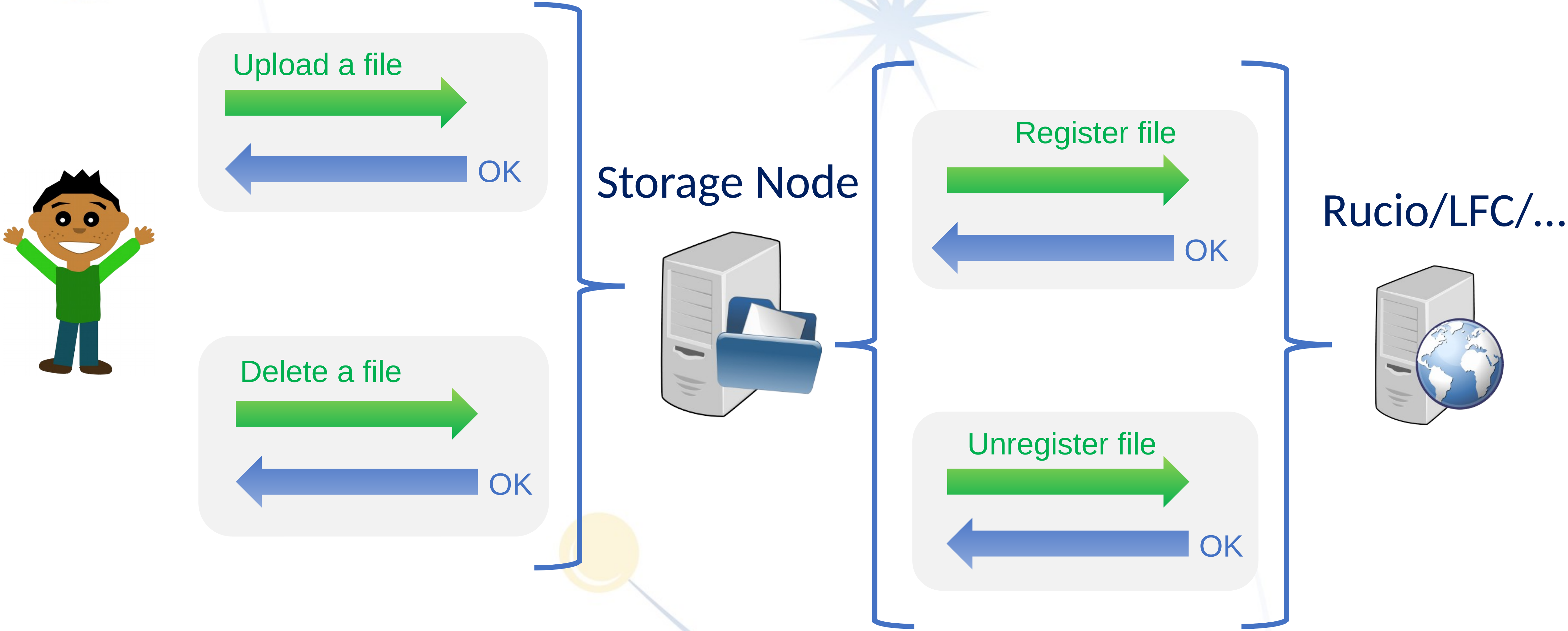
Thanks for listening!

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Bonus material

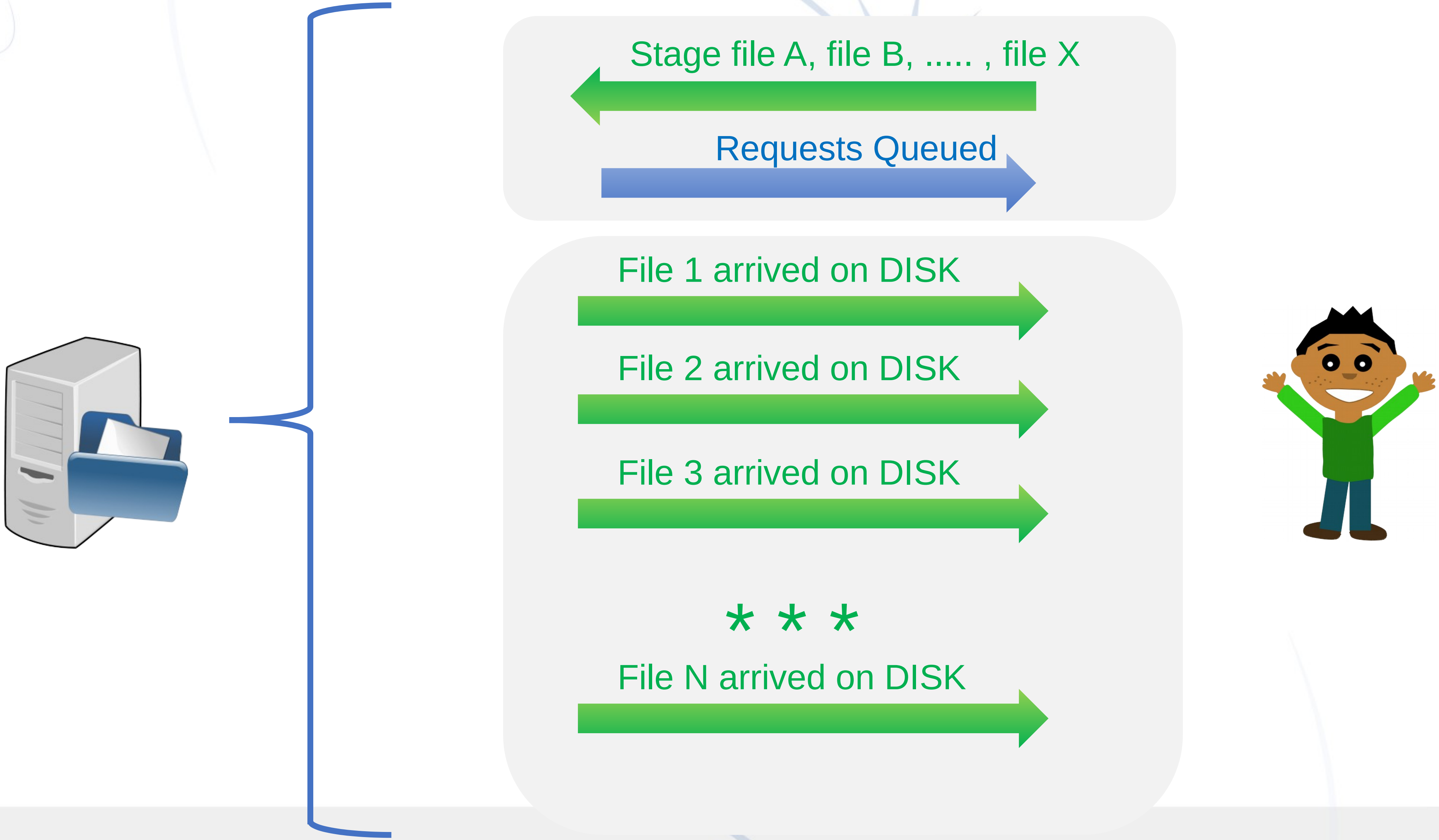
Funded by the European Union's
Horizon 2020 - Grant N° 824064

New solutions to old problems



Funded by the European Union's
Horizon 2020 - Grant N° 824064

New solutions to old problems



Stage file A, file B, ... , file X

Requests Queued

File 1 arrived on DISK

File 2 arrived on DISK

File 3 arrived on DISK

* * *

File N arrived on DISK

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Comparison: events in industry…

Amazon Lambda

Image Thumbnail Creation

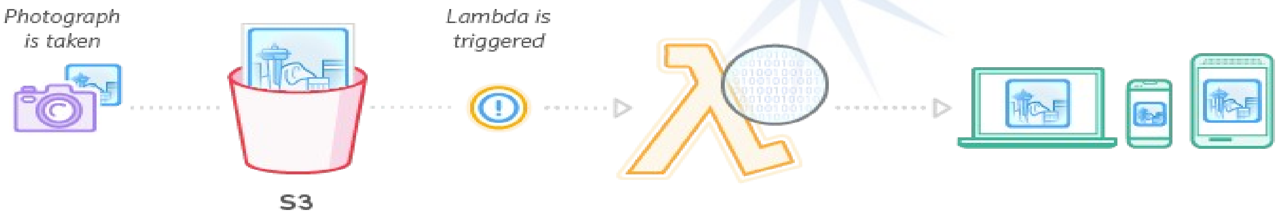
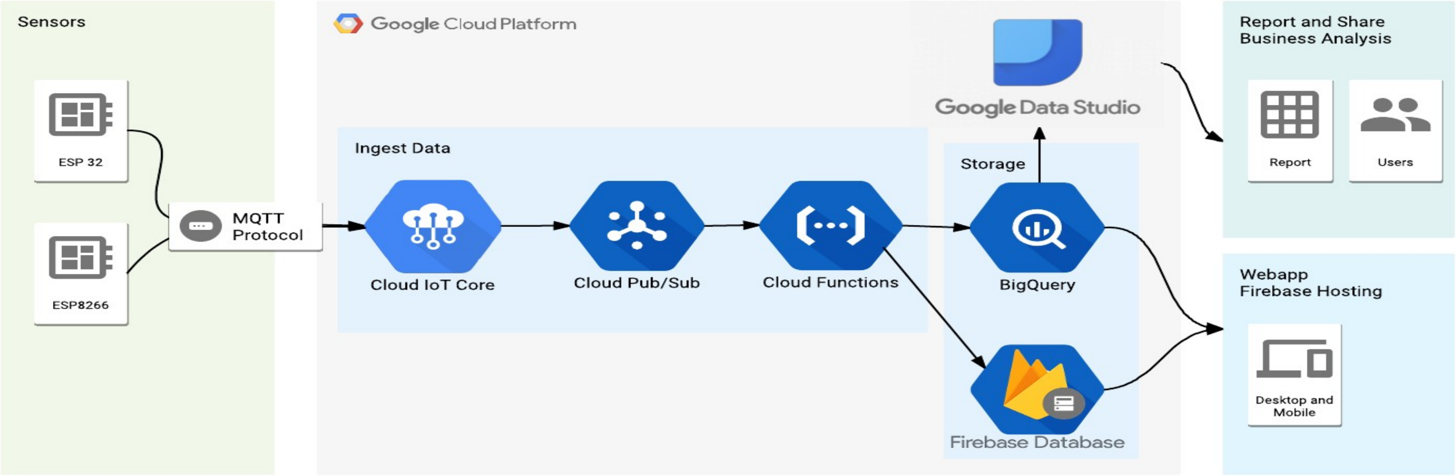


Photo is uploaded into S3 bucket

Lambda runs image resizing code to generate Web, mobile and tablet sizes

Google Cloud Platform



Funded by the European Union's
Horizon 2020 - Grant N° 824064

Comparison: events in Open-Source

STORM

Apache Storm is a distributed stream processing computation framework written predominantly in the Clojure programming language.

Spark

OpenWhisk

APACHE kafka®
A distributed streaming platform

samza

Samza allows you to build stateful applications that process data in real-time from multiple sources including Apache Kafka.

APACHE nifi

Apache NiFi is a software project from the Apache Software Foundation designed to automate the flow of data between software systems.

Kubeless

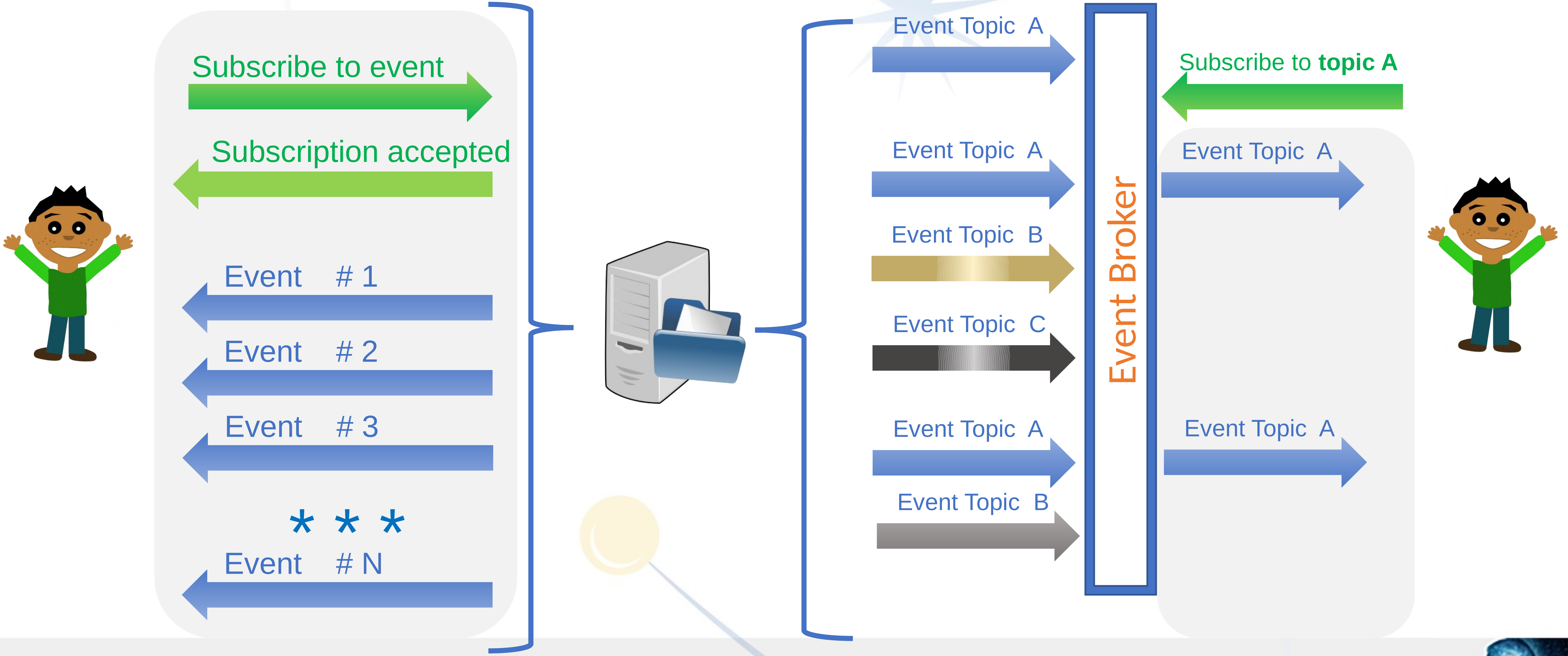
Kubeless is a Kubernetes-native serverless framework that lets you deploy small bits of code (functions) without having to worry about the underlying infrastructure.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

New way of interacting: storage events

Direct event delivery

Brokered event delivery



Subscribe to event

Subscription accepted

Event # 1

Event # 2

Event # 3

* * *

Event # N

Event Topic A

Event Topic A

Event Topic B

Event Topic C

Event Topic A

Event Topic B

Event Broker

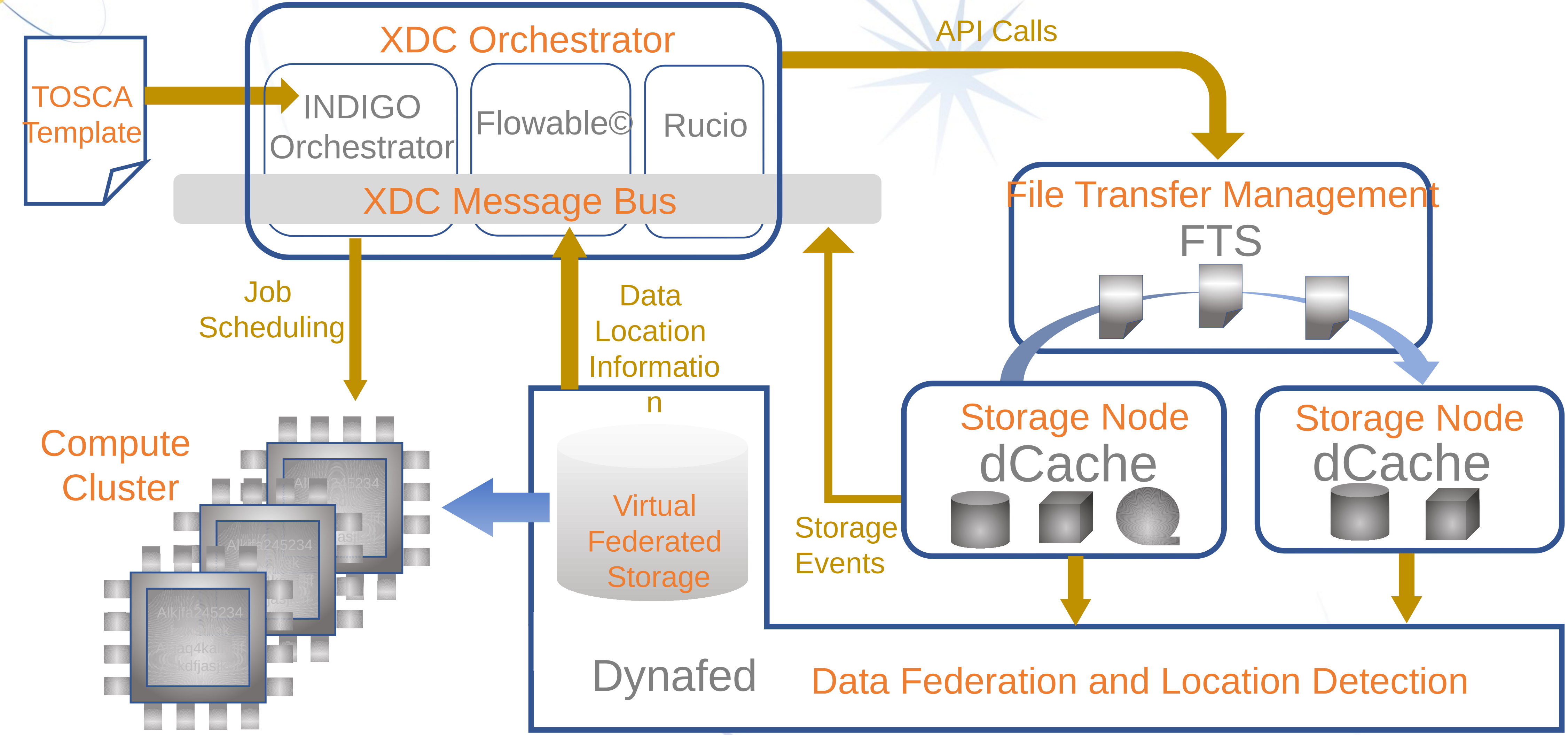
Subscribe to **topic A**

Event Topic A

Event Topic A

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Full XDC Layout using Storage Events



Funded by the European Union's Horizon 2020 - Grant N° 824064