Wavefier

A prototype for real time Gravitational Wave transient signal classifier

EGO - Elena Cuoco (*Scientific Supervisor*)

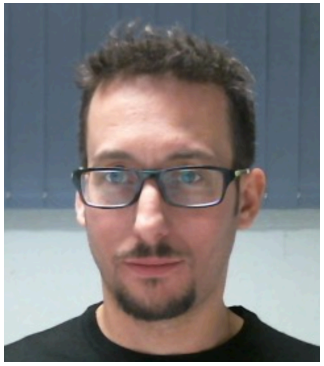
Trust-IT - Alessandro Staniscia, Emanuel Marzini, Filip Morawski, Alessandro Petrocelli, Silvana Muscella

H2020-ASTERICS project brings together for the first time scientists and communities from astronomy, astrophysics, particle astrophysics& big data. http://www.asterics2020.eu

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

About us

Elena Cuoco
European Gravitational Observatory



Alessandro Staniscia
Temporary Consultant for Trust-IT Service

- Ligo/Virgo Co-chair of Machine Learning group
- ESCAPE project scientist https://projectescape.eu
- Chair of CA17137: www.g2net.eu
- 10% Freelancer and Temporary Consultant
- 80% Scrum Muster and Team Member on NAVIONICS A Garmin Company
- 10% Dreamer

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

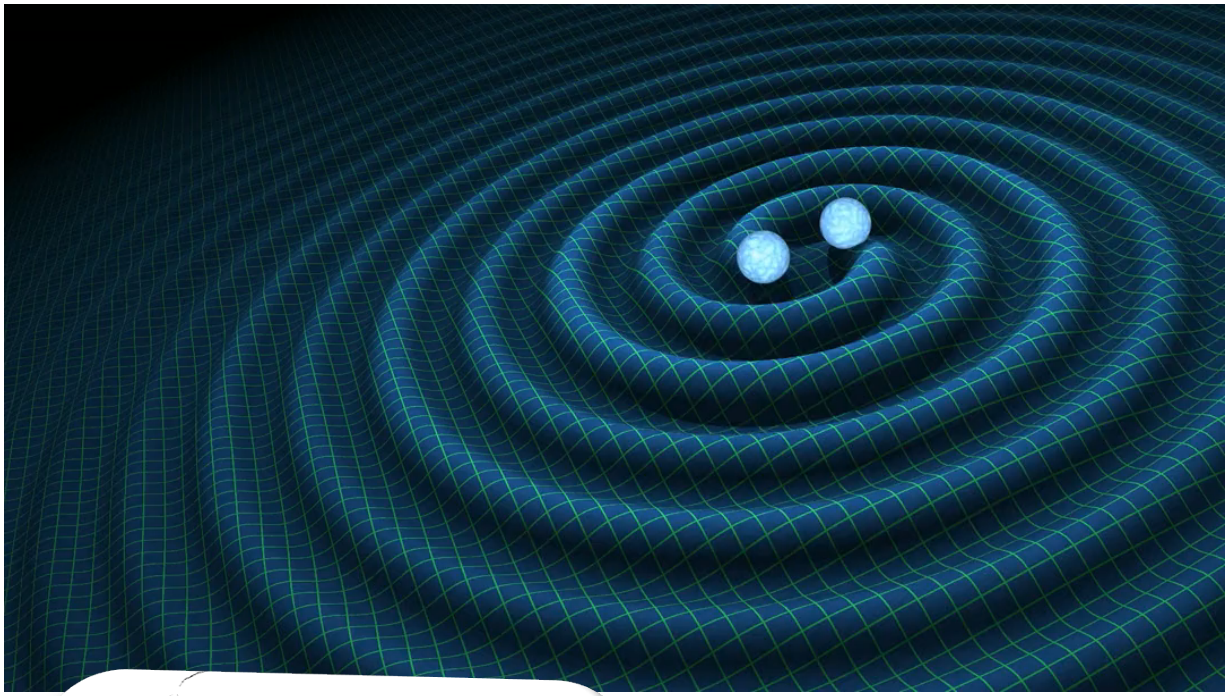
Outline

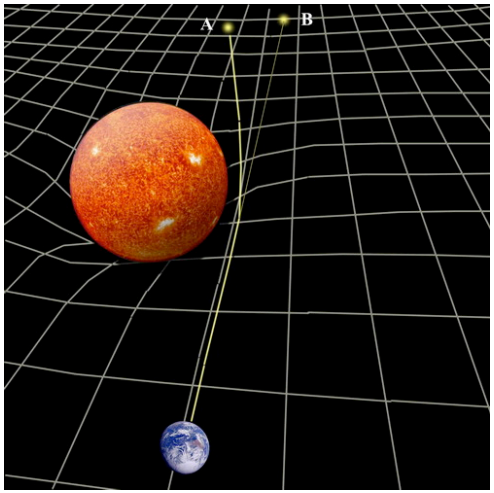
I. *Gravitational waves in a nutshell*
II. *Why real time analysis*
III. *From In-time analysis to Wavefier prototype*
IV. *The technical / implementation aspects*
V. *What's next*

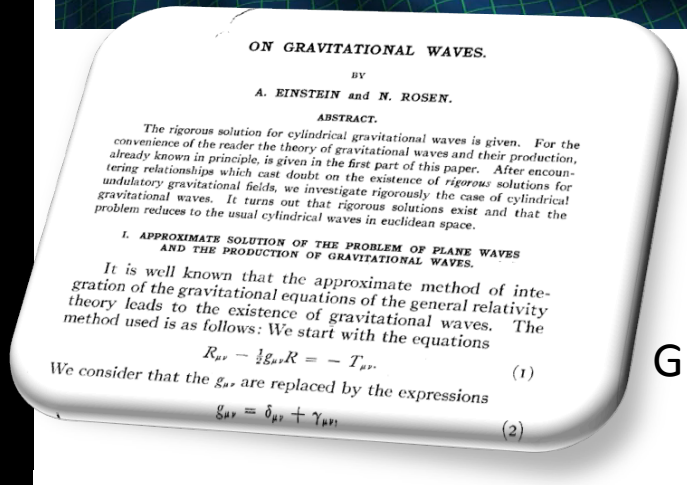
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

I. Gravitational waves in a nutshell

What are Gravitational Waves (GWs)?







ON GRAVITATIONAL WAVES.

BY

A. EINSTEIN and N. ROSEN.

ABSTRACT.

The rigorous solution for cylindrical gravitational waves is given. For the convenience of the reader the theory of gravitational waves and their production, already known in principle, is given in the first part of this paper. After encountering relationships which cast doubt on the existence of rigorous solutions for undulatory gravitational fields, we investigate rigorously the case of cylindrical gravitational waves. It turns out that rigorous solutions exist and that the problem reduces to the usual cylindrical waves in euclidean space.

I. APPROXIMATE SOLUTION OF THE PROBLEM OF PLANE WAVES AND THE PRODUCTION OF GRAVITATIONAL WAVES.

It is well known that the approximate method of integration of the gravitational equations of the general relativity theory leads to the existence of gravitational waves. The method used is as follows: We start with the equations

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = -T_{\mu\nu}. \qquad (1)$$

We consider that the $g_{\mu\nu}$ are replaced by the expressions

$$g_{\mu\nu} = \delta_{\mu\nu} + \gamma_{\mu\nu}, \qquad (2)$$

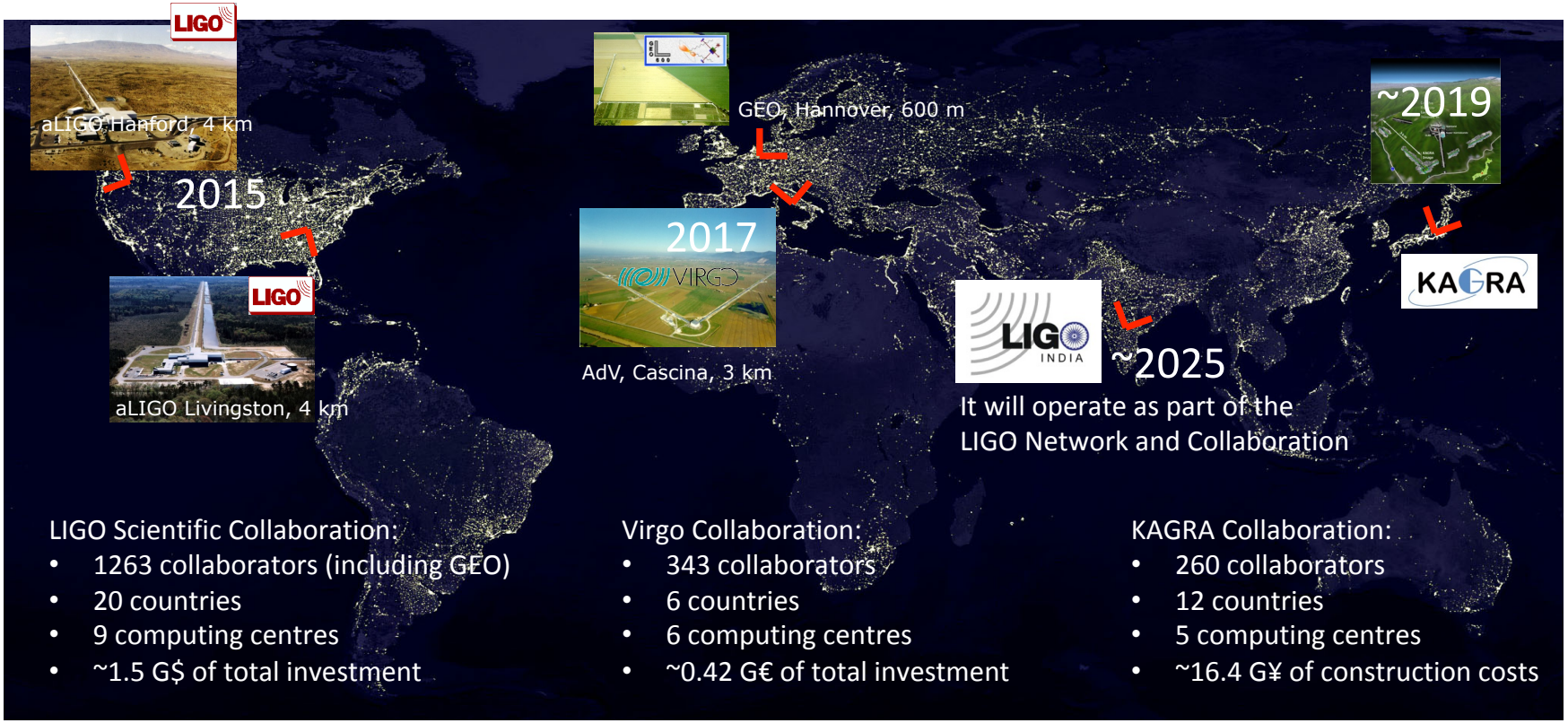
Gravitational Waves (1916)

General Relativity (1915)

$$\mathbf{G}_{\mu\nu} = \frac{8\pi G}{c^4}\mathbf{T}_{\mu\nu}$$

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

International collaboration

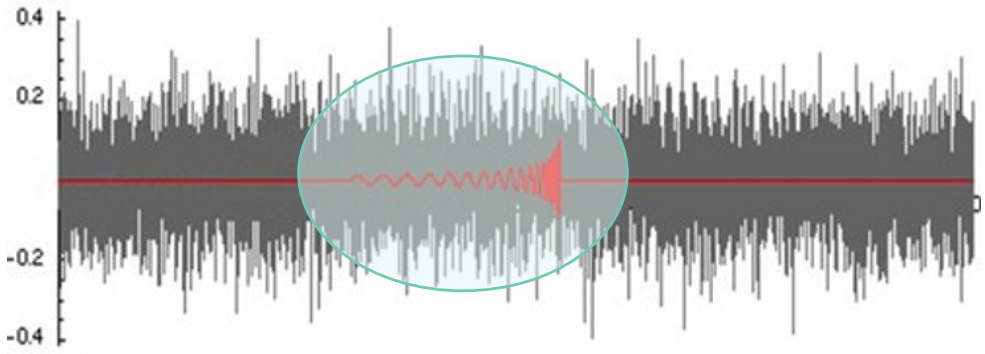


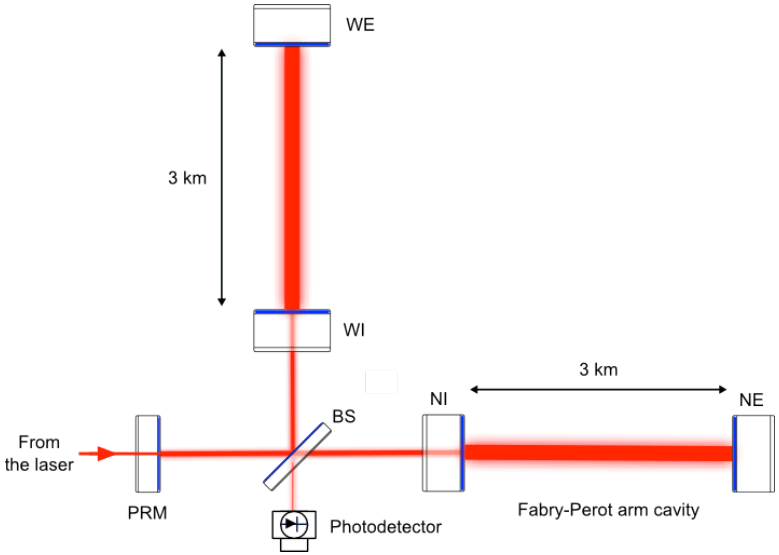
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

LIGO/Virgo data

Time series sequences… **noisy time series** with low amplitude GW signal buried in

Data flux stream: 50MB/s





H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

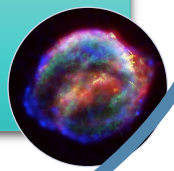
Which kind of astrophysical sources?

- Known waveform
  - Transient signal

Compact binary coalescence (CBC)

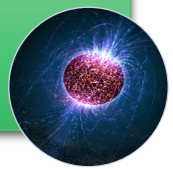
- Unknown waveform
  - Transient signal

Core Collapse Supernovae (CCSN)



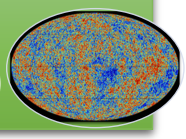
- Known waveform
  - Persistent signal

Continuous Waves (CW)



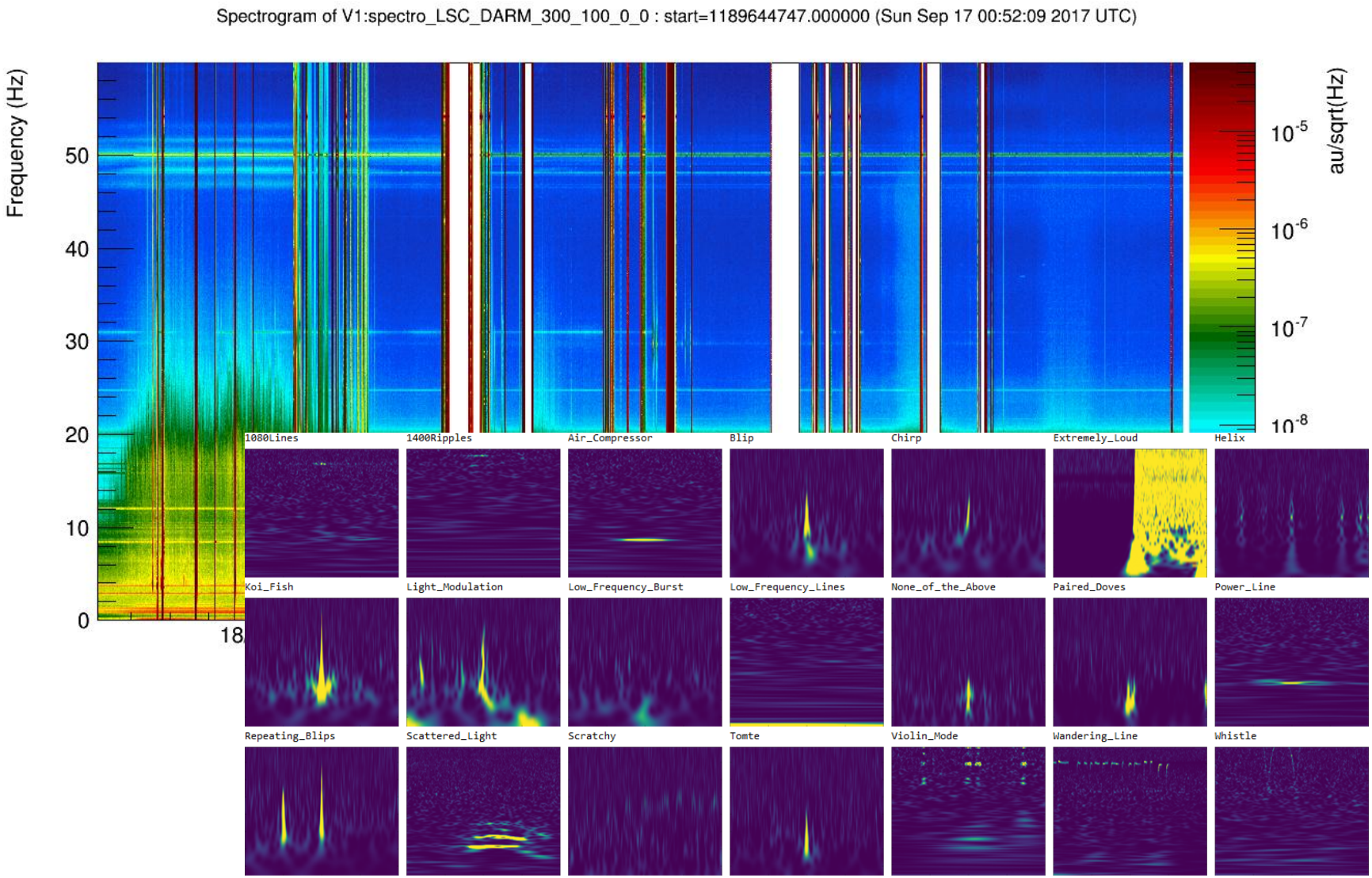
- Unknown model
  - Persistent signal

Stochastic Background (SB)



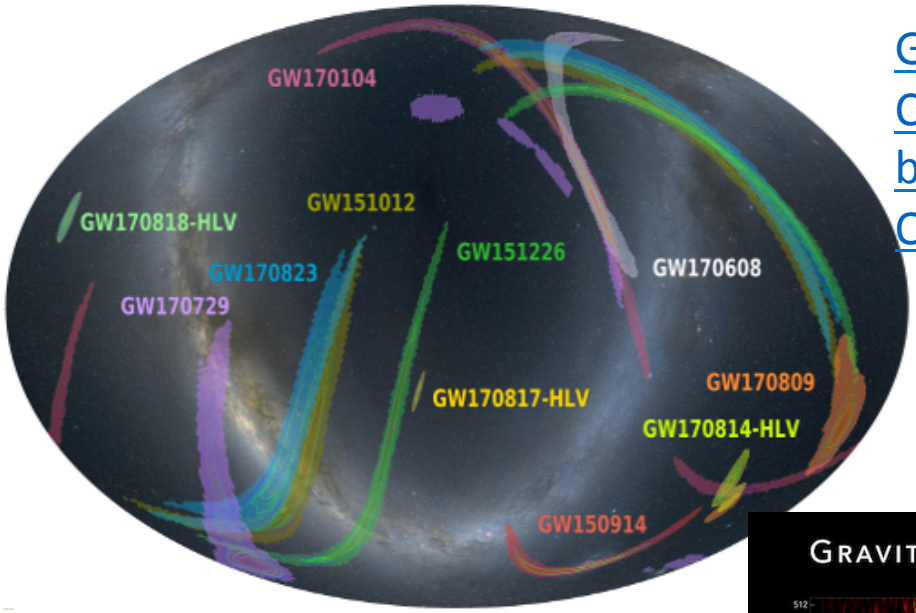
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Which kind of noise?

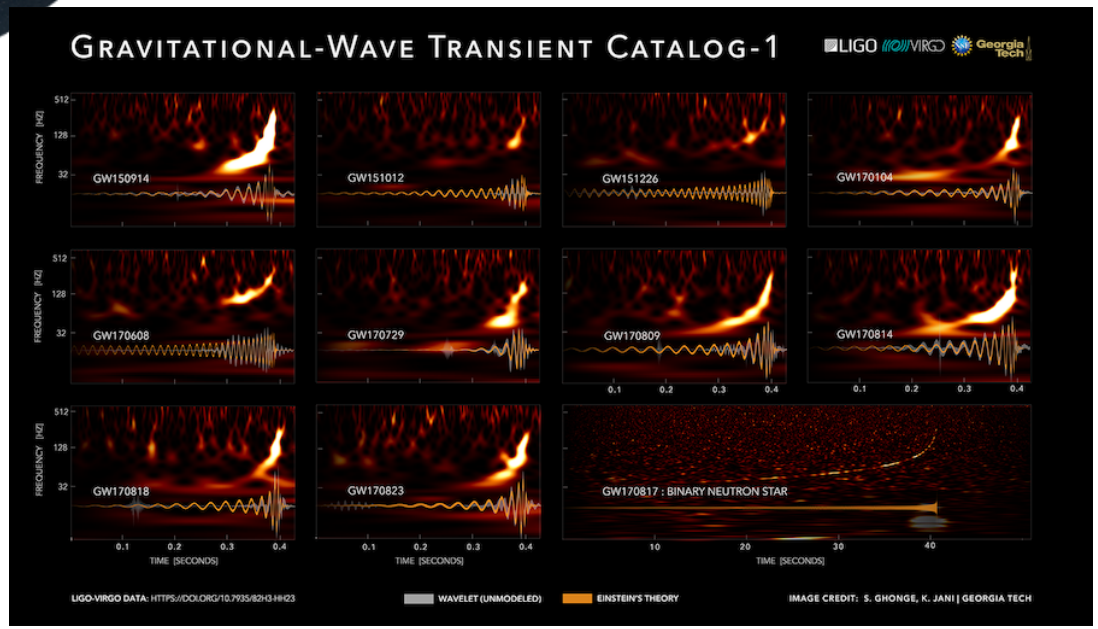


H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

The first GW catalog (O1/O2 run)



[GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs](arxiv.org/abs/1811.12907) arxiv.org/abs/1811.12907



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

II

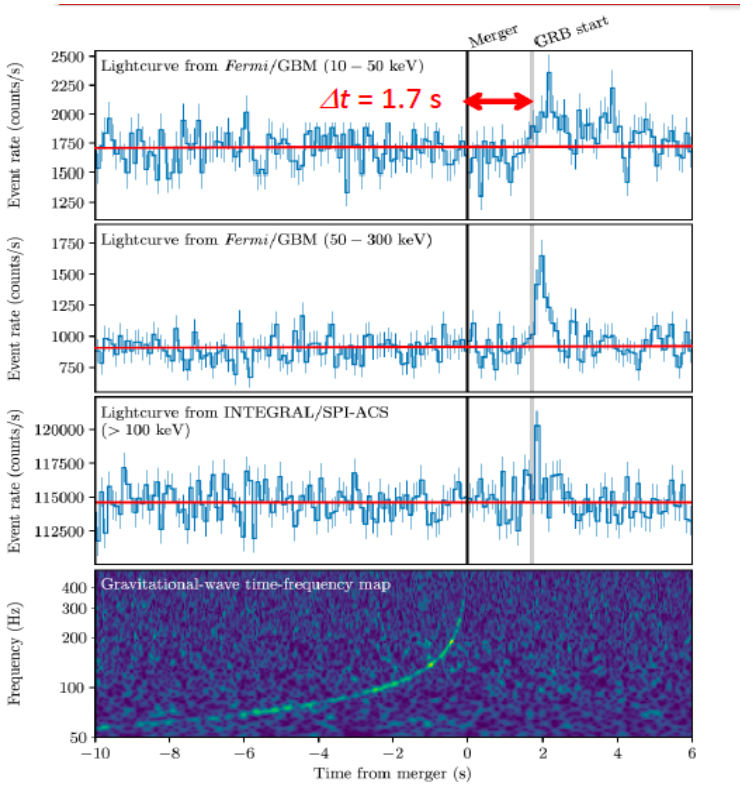
Why real time analysis

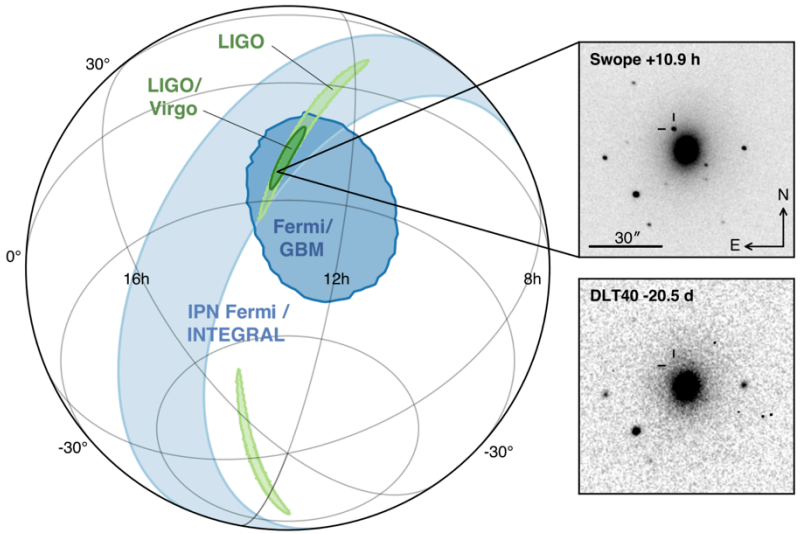
Use of Machine learning

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

17 August 2017, 12:41:04 UT
The MultiMessenger Astronomy

II. Why real time analysis

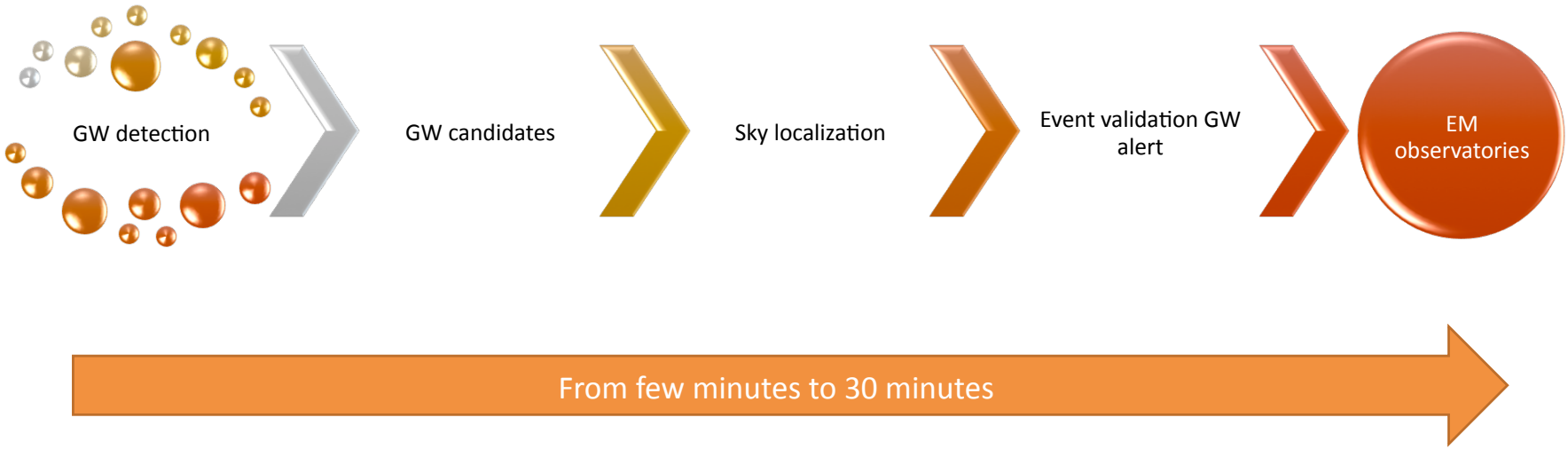




DOI:10.1103/PhysRevLett.119.161101.

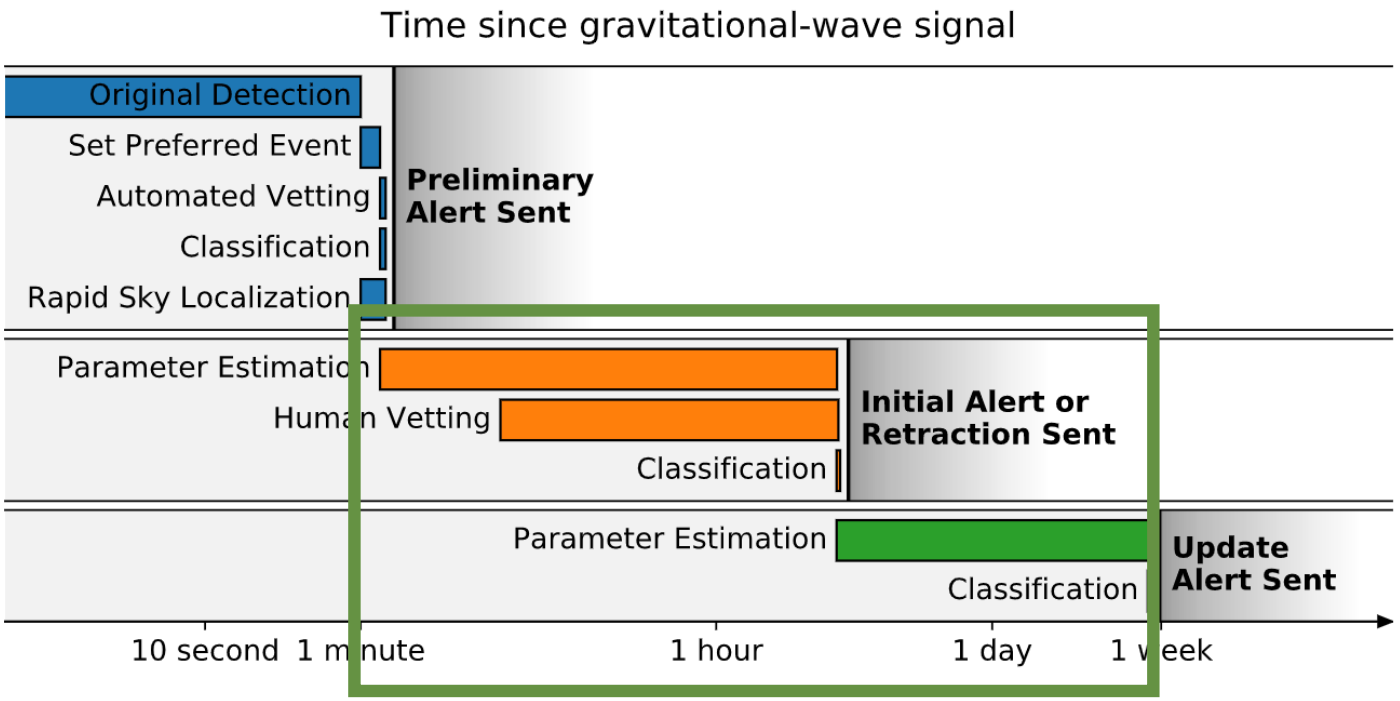
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Low Latency data analysis



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

GW alert system



https://emfollow.docs.ligo.org/userguide/index.html

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

O3 event rate ~1/week. Up to now 32 events!

GraceDB — Gravitational Wave Candidate Event Database

HOME | SEARCH | LATEST | DOCUMENTATION | LOGIN

Latest — as of 8 July 2019 13:15:27 UTC

Test and MDC events and superevents are not included in the search results by default; see the query help for information on how to search for events and superevents in those categories.

Query:

Search for: Superevent

Search

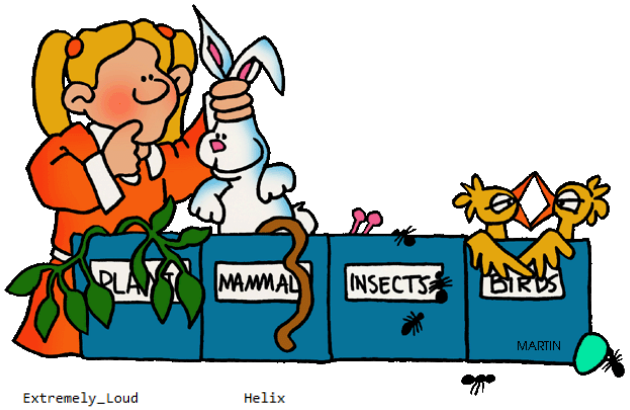
UID	Labels	t_start	t_0	t_end	FAR (Hz)	Created (UTC)
S190707q	ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1246527223.118398	1246527224.181226	1246527225.284180	5.265e-12	2019-07-07 09:33:44 UTC
S190706ai	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1246487218.321541	1246487219.344727	1246487220.585938	1.901e-09	2019-07-06 22:26:57 UTC
S190701ah	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1246048403.576563	1246048404.577637	1246048405.814941	1.916e-08	2019-07-01 20:33:24 UTC
S190630ag	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1245955942.175325	1245955943.179550	1245955944.183184	1.435e-13	2019-06-30 18:52:28 UTC
S190602aq	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1243533584.081266	1243533585.089355	1243533586.346191	1.901e-09	2019-06-02 17:59:51 UTC
S190524q	ADVNO SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1242708743.678669	1242708744.678669	1242708746.133301	6.971e-09	2019-05-24 04:52:30 UTC
S190521r	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1242459856.453418	1242459857.460739	1242459858.642090	3.168e-10	2019-05-21 07:44:22 UTC
S190521g	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1242442966.447266	1242442967.606934	1242442968.888184	3.801e-09	2019-05-21 03:02:49 UTC
S190519bj	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1242315361.378873	1242315362.655762	1242315363.676270	5.702e-09	2019-05-19 15:36:04 UTC
S190518bb	ADVNO SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1242242376.474609	1242242377.474609	1242242380.922655	1.004e-08	2019-05-18 19:19:39 UTC
S190517h	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1242107478.819517	1242107479.994141	1242107480.994141	2.373e-09	2019-05-17 05:51:23 UTC
S190513bm	ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1241816085.736106	1241816086.869141	1241816087.869141	3.734e-13	2019-05-13 20:54:48 UTC
S190512at	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1241719651.411441	1241719652.416286	1241719653.518066	1.901e-09	2019-05-12 18:07:42 UTC
S190510g	ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1241492396.291636	1241492397.291636	1241492398.293185	8.834e-09	2019-05-10 03:00:03 UTC
S190503bf	ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1240944861.288574	1240944862.412598	1240944863.422852	1.636e-09	2019-05-03 18:54:26 UTC
S190426c	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1240327332.331668	1240327333.348145	1240327334.353516	1.947e-08	2019-04-26 15:22:15 UTC
S190425z	ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK	1240215502.011549	1240215503.011549	1240215504.018242	4.538e-13	2019-04-25 08:18:26 UTC
S190421ar	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1239917953.250977	1239917954.409180	1239917955.409180	1.489e-08	2019-04-21 21:39:16 UTC
S190412m	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1239082261.146717	1239082262.222168	1239082263.229492	1.683e-27	2019-04-12 05:31:03 UTC
S190408an	PE_READY ADVOK SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK GCN_PRELIM_SENT	1238782699.268296	1238782700.287958	1238782701.359863	2.811e-18	2019-04-08 18:18:27 UTC
S190405ar	ADVNO SKYMAP_READY EMBRIGHT_READY PASTRO_READY DQOK	1238515307.863646	1238515308.863646	1238515309.863646	2.141e-04	2019-04-05 16:01:56 UTC

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

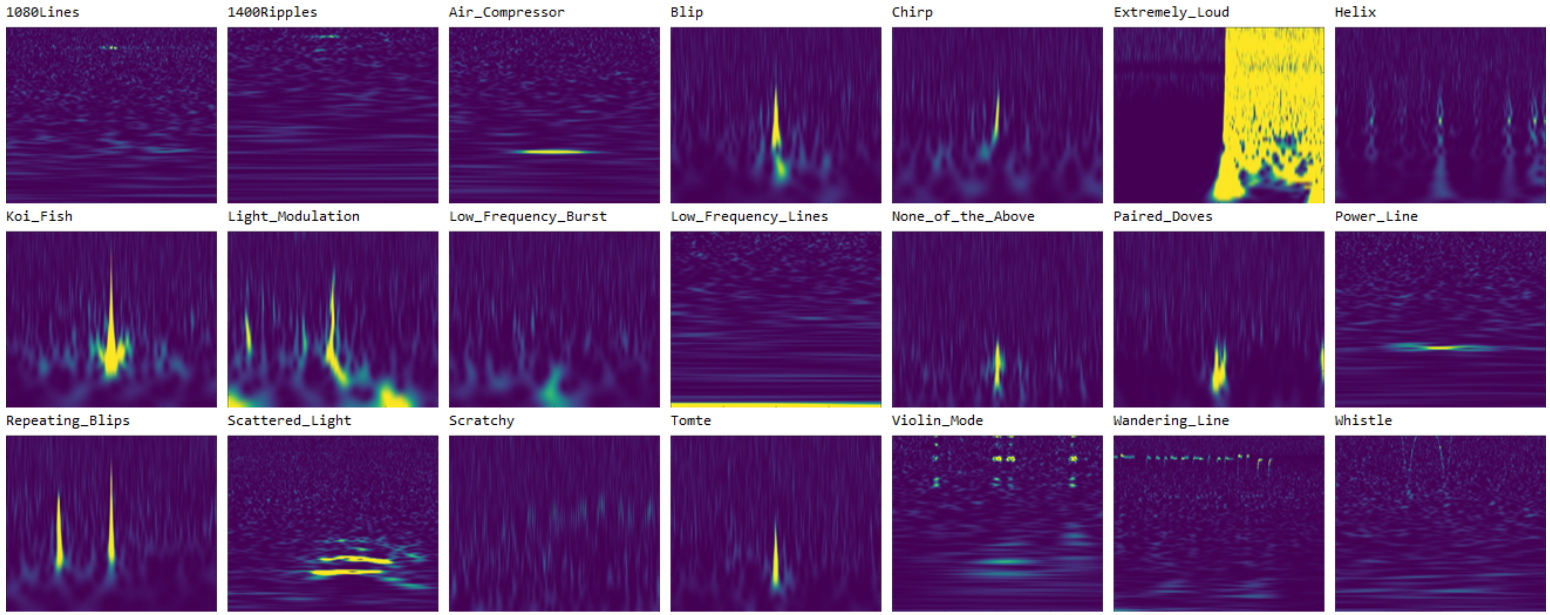
Transient Signal Classification using Machine Learning

If we are able to classify the noise events,

we can clean the data in a fast way



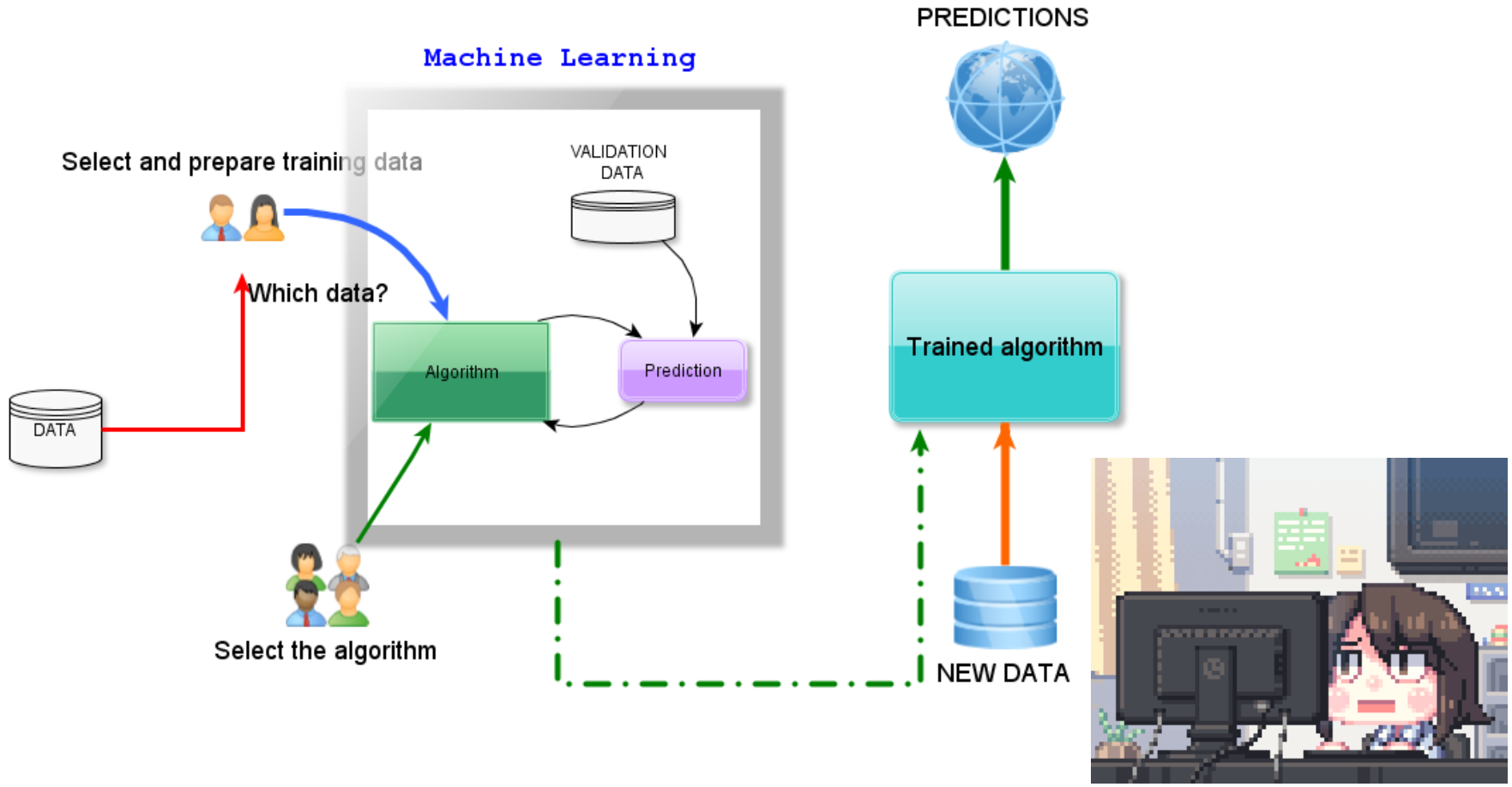
Gravity Spy, Zevin et al (2017)



https://www.zooniverse.org/projects/zooniverse/gravity-spy

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Artificial Intelligence workflow



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Machine learning implemented pipeline

Supervised

- We need a labeled training data set
- We want to disentangles 'glitches' due to noise from 'transient signals' due to astrophysical signals

Unsupervised

- Extract signal features
- We try to catch main relations among the detected signals to create cluster of classes

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

III

From In-time analysis to Wavefier prototype

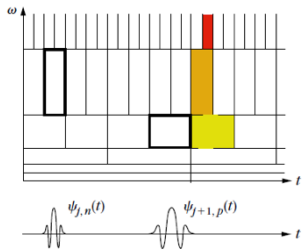
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

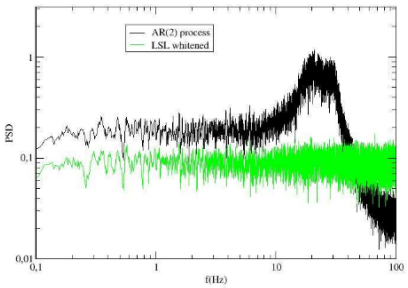
ETG*: Wavelet Detection Filter (WDF) workflow

* ETG= Event Trigger Generator

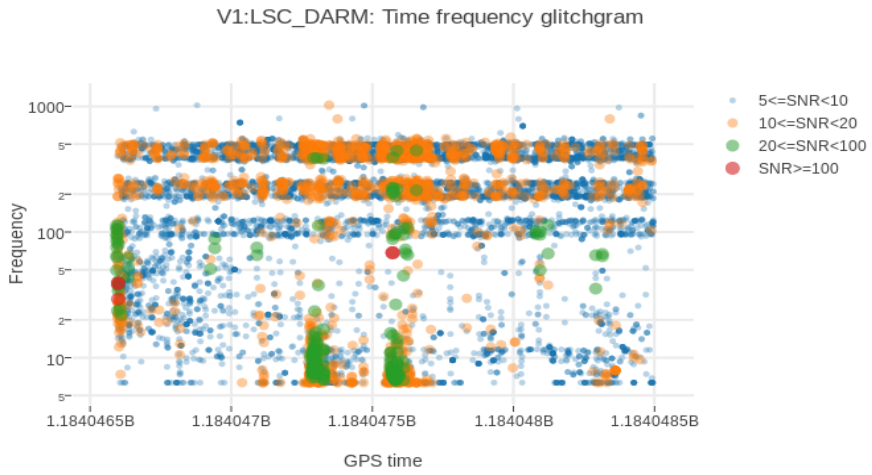
$$x_i = h_i + n_i, \quad i = 0, 1, \ldots N-1,$$

$$Wf(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{b}} \psi^* \left(\frac{t-a}{b} \right) \quad dt.$$



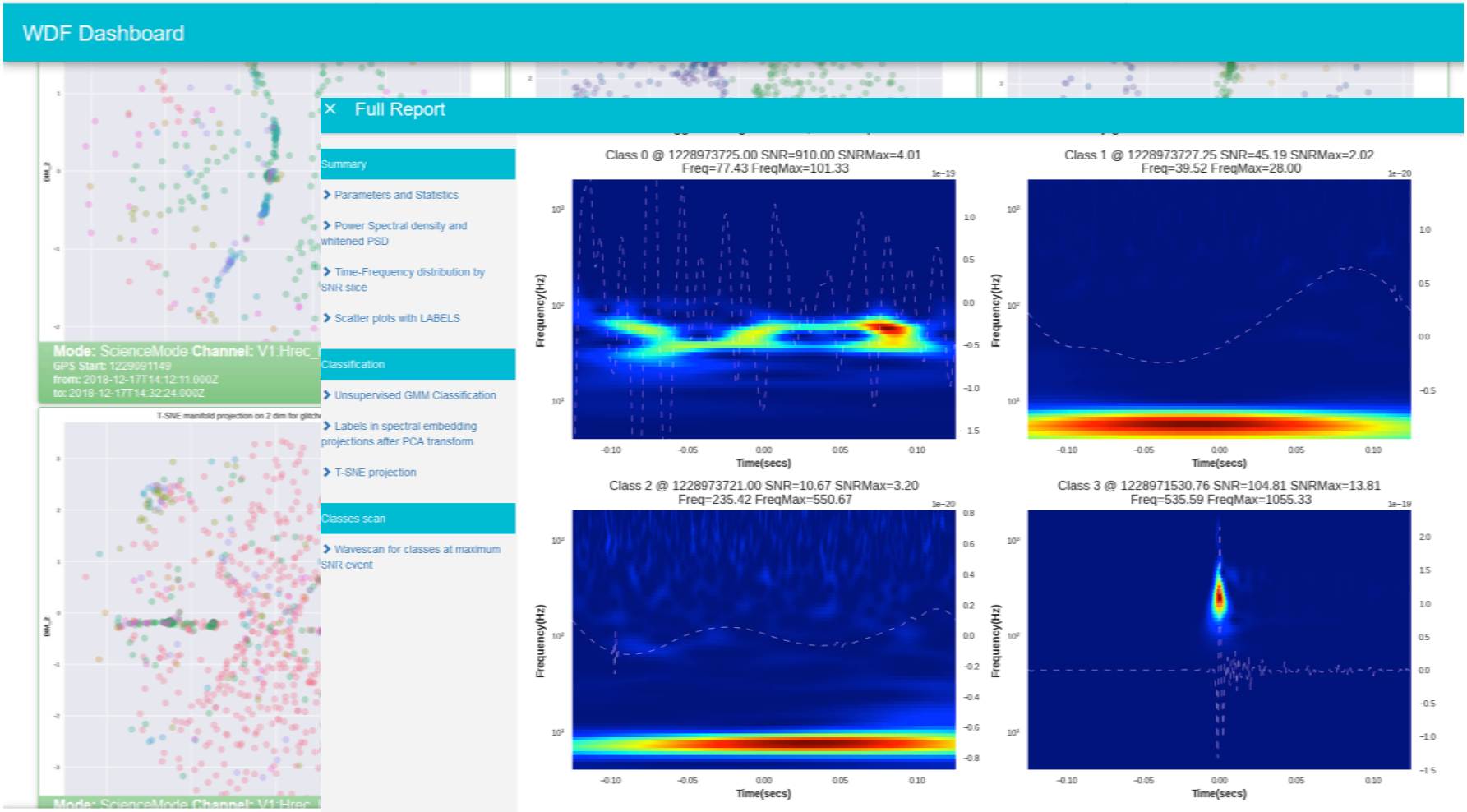


$$\hat{h}_i = W^{-1}(t[W(x_i)]).$$



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

In-Time Analysis running @VIRGO



https://wdf.virgo-gw.eu/

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Why Wavefier and Key Objectives

- Setup a prototype for a real time pipeline for the **detection of transient signals and their automatic classification**
- Best practices for software management
- Test different software architecture solutions to prototype a scalable pipeline for big data analysis in GW context.
- Interoperability and access to data and services
- ICT services supporting research infrastructures
- Use of data in network infrastructures and services
- Big data and Machine Learning solution easy to plug-in
- Test on cluster infrastructures

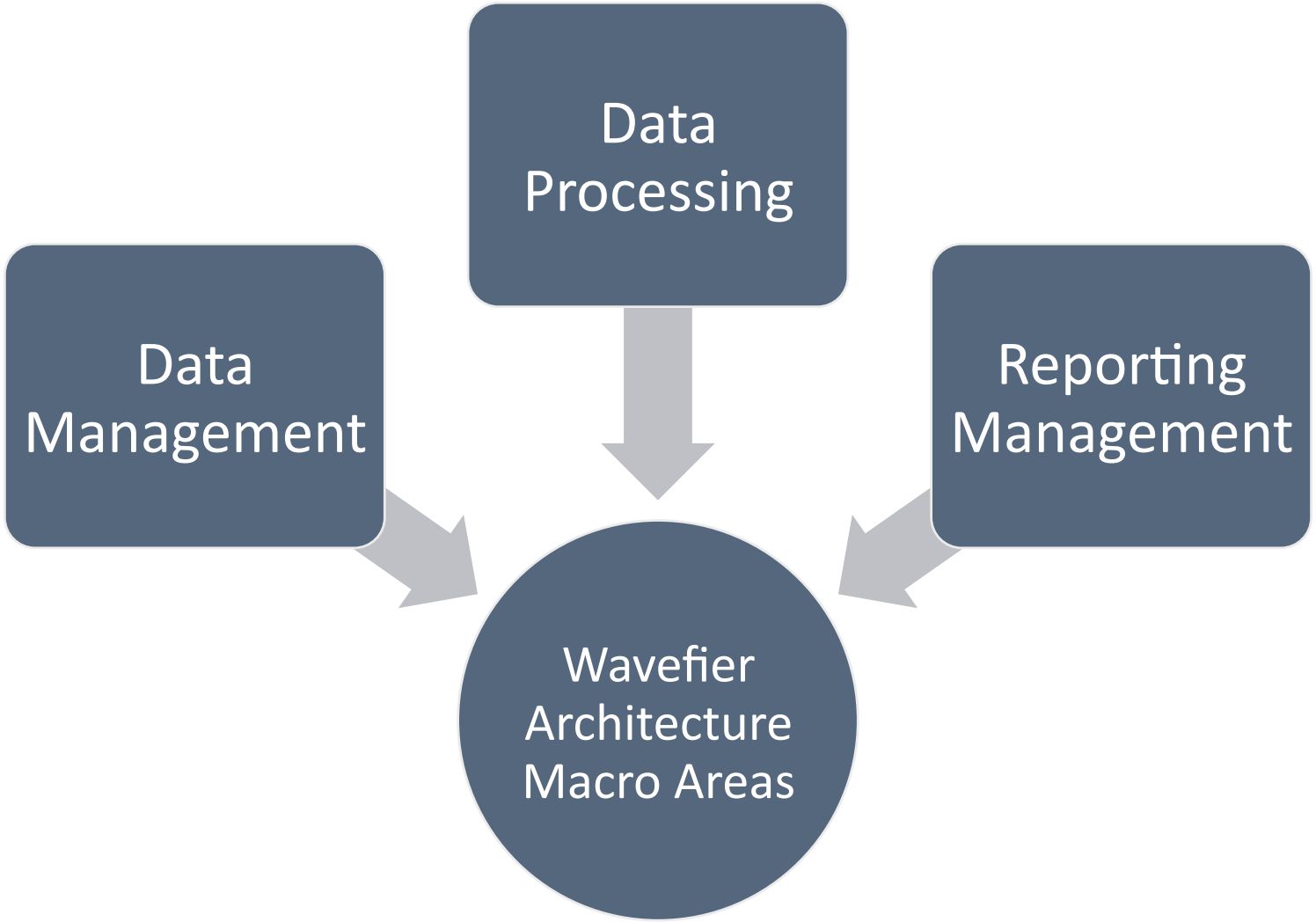
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

IV

The technical and implementation aspects

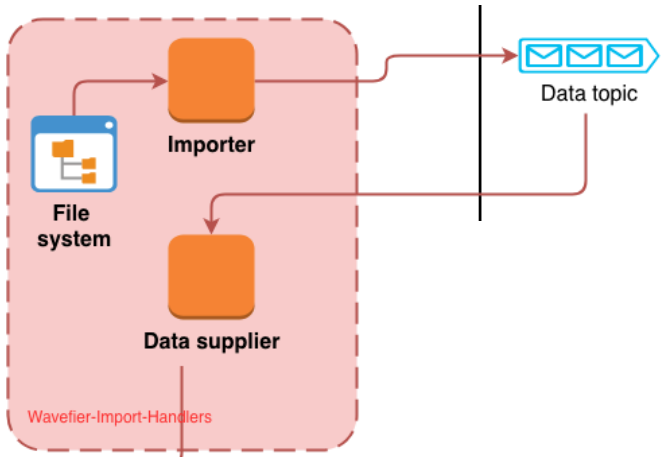
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Architecture Overview



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

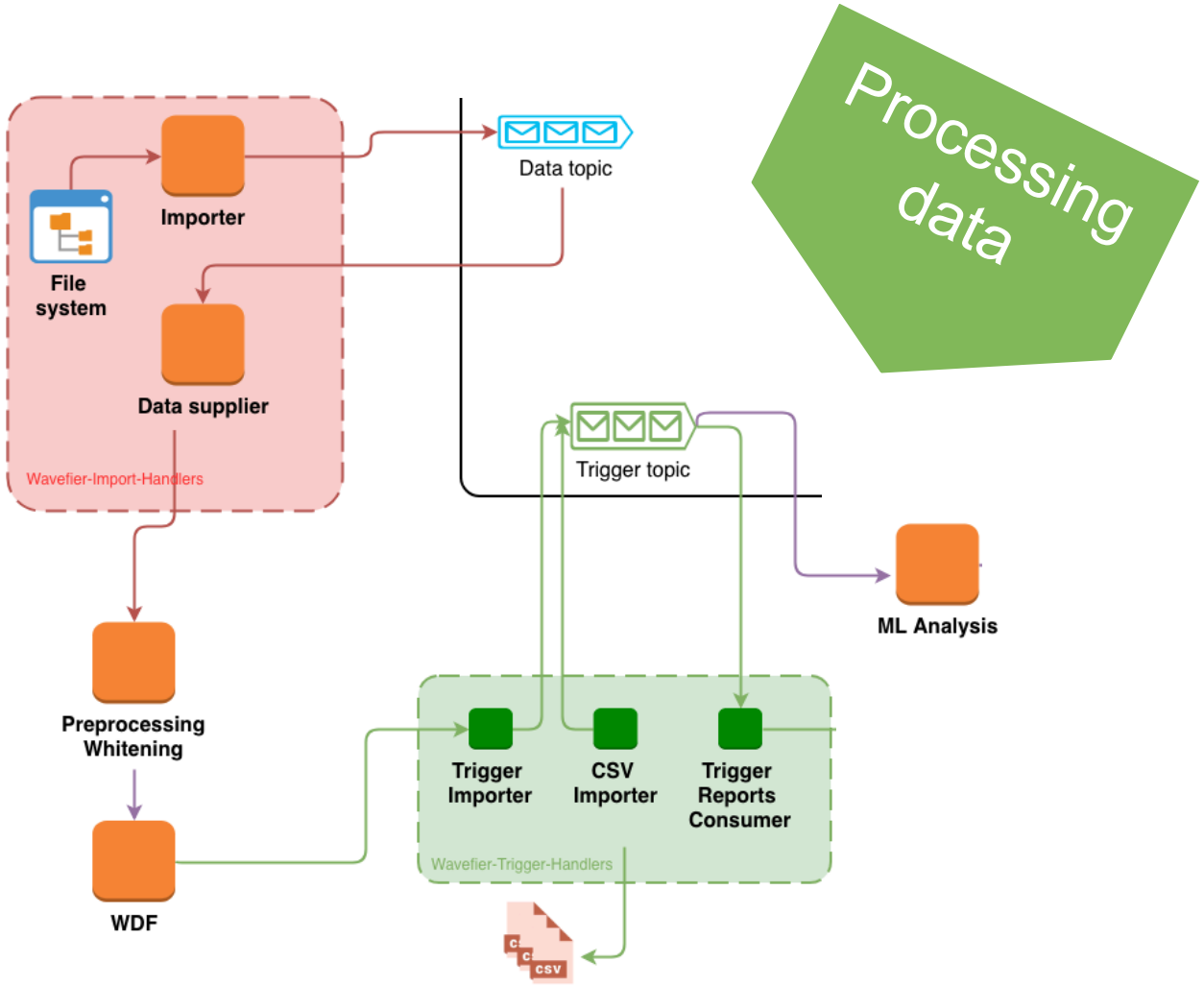
Architecture Overview – Data Injection



Data management

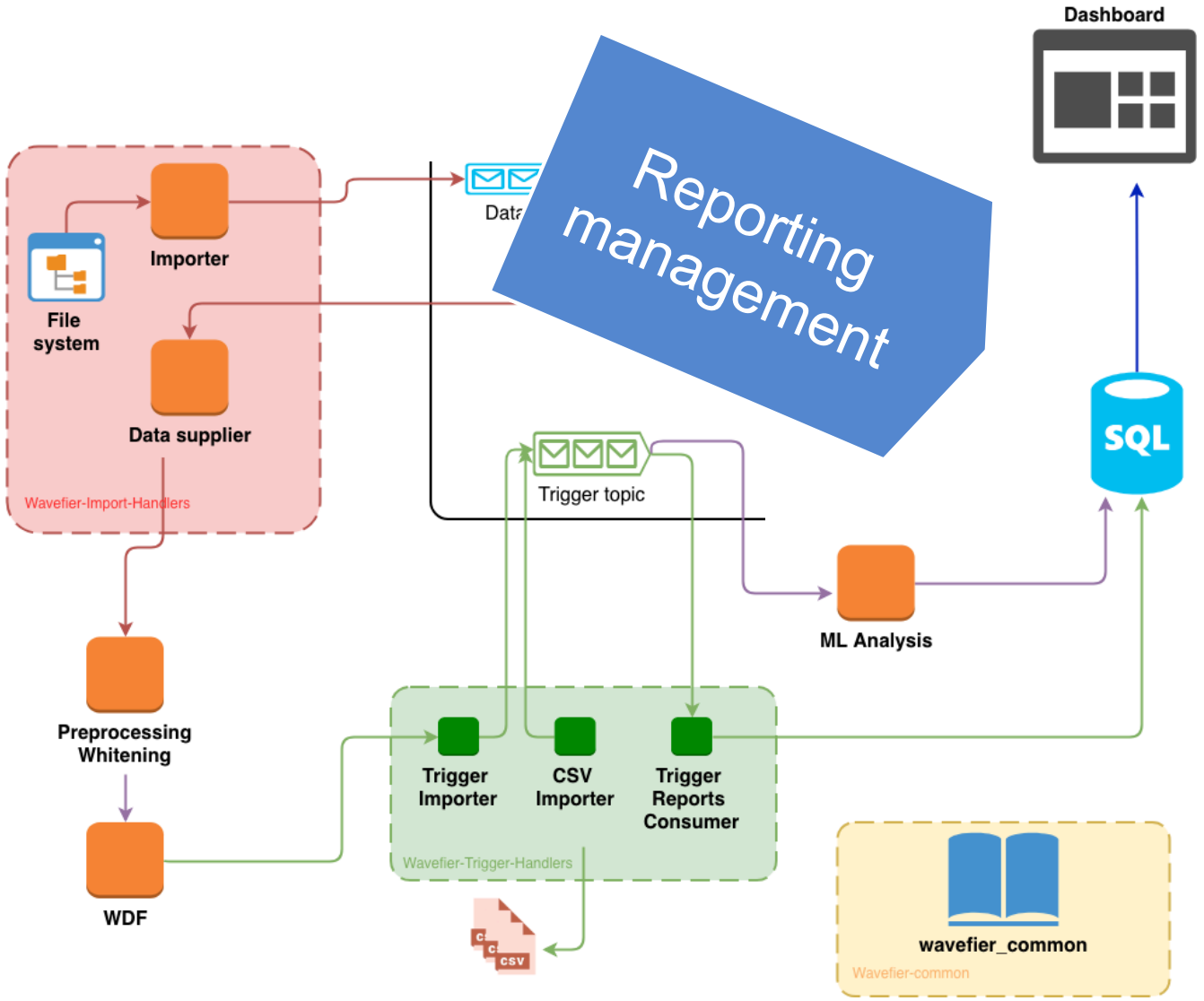
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Architecture Overview – Processing Data



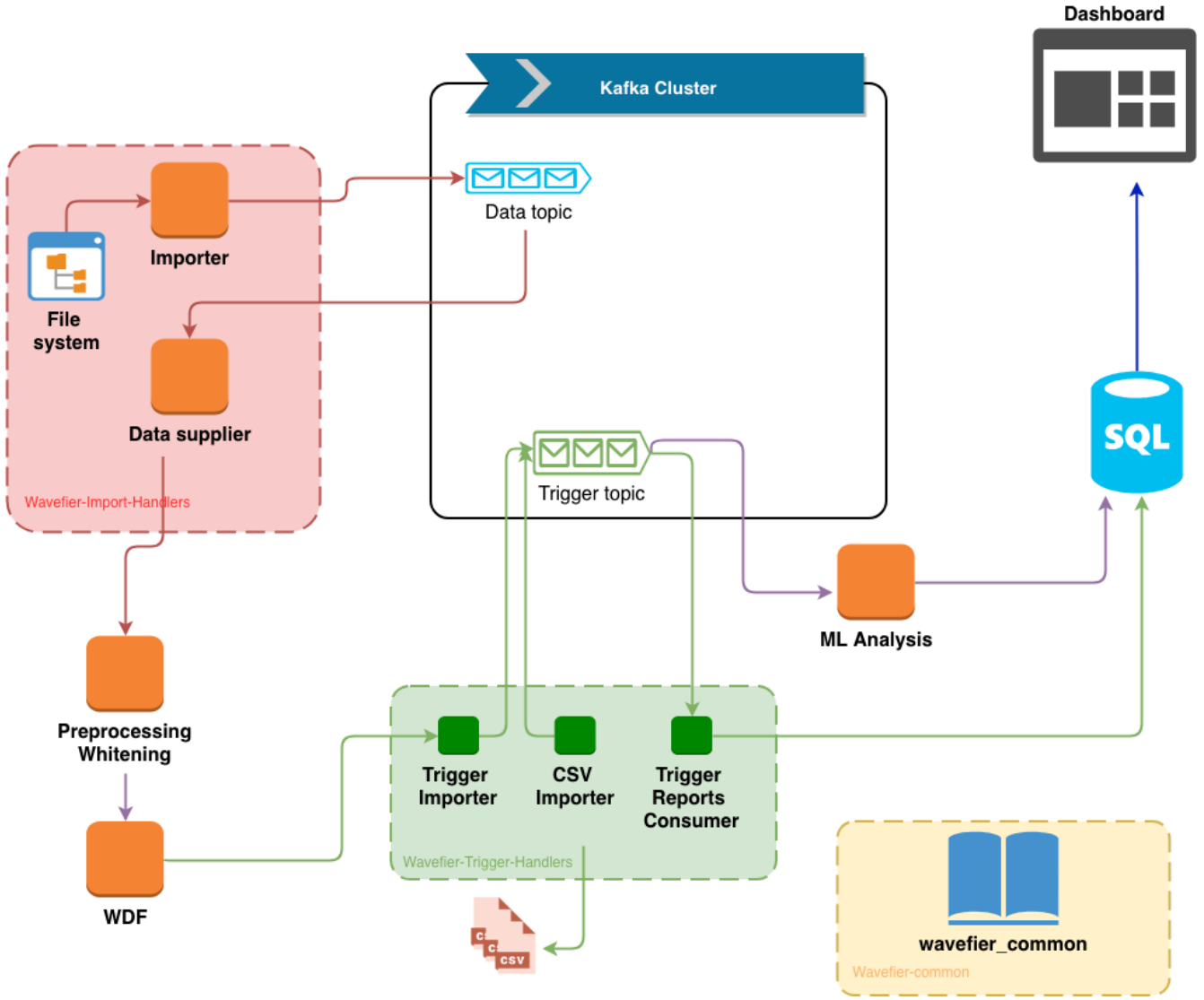
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Architecture Overview – Reporting



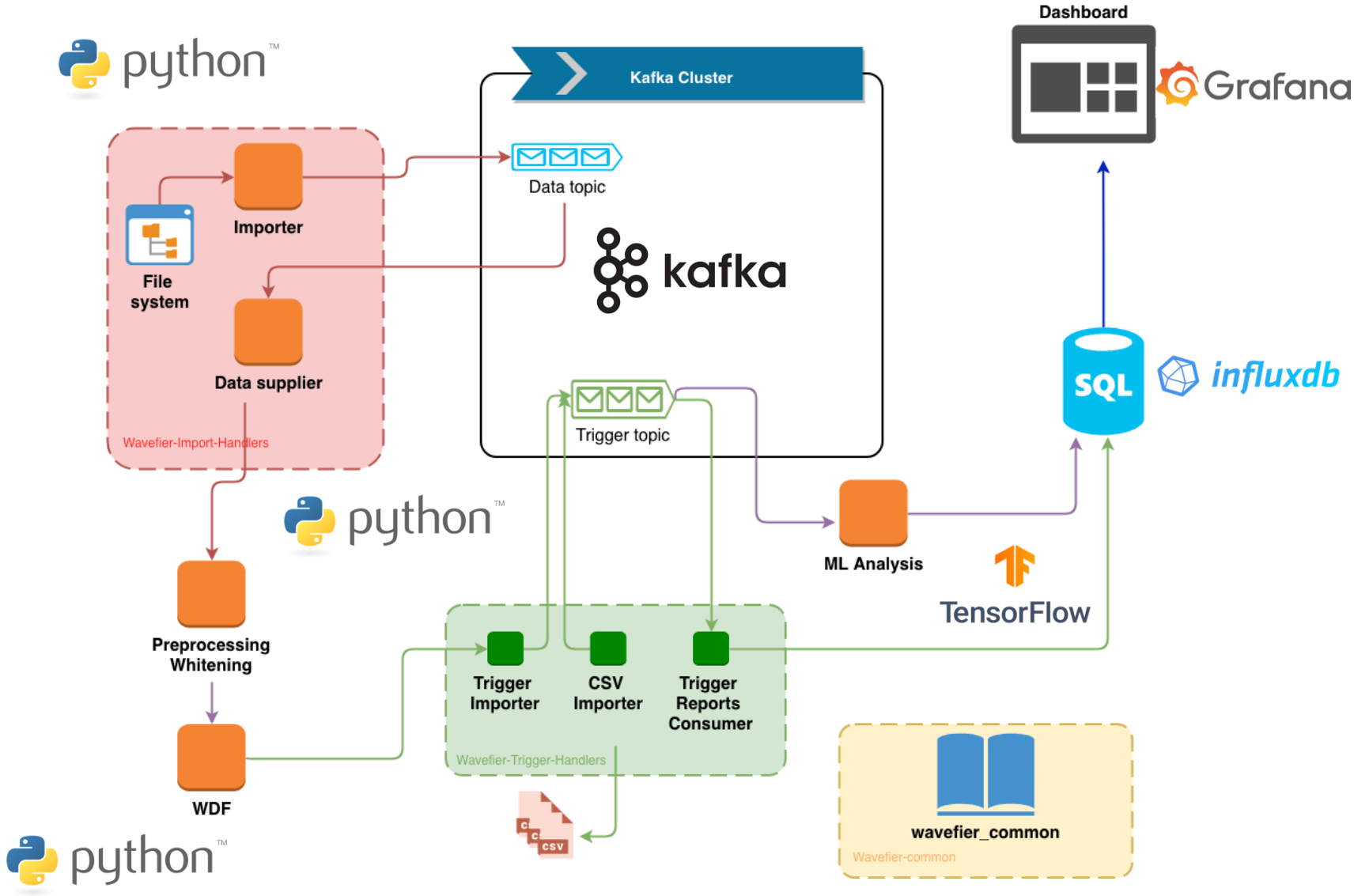
H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Architecture Overview



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Software Deployment



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

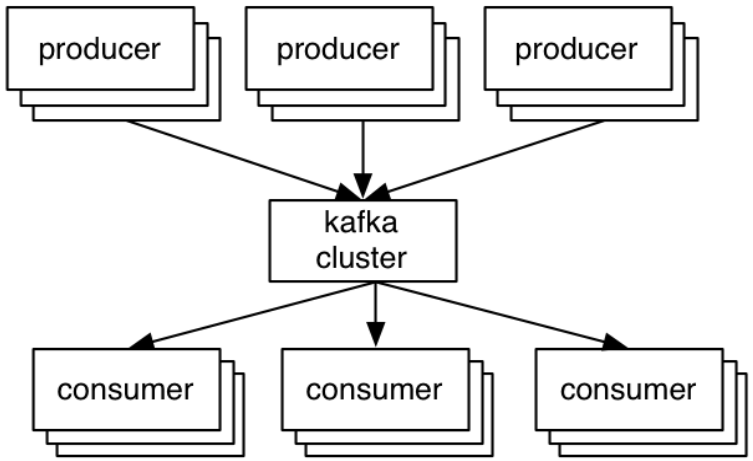
Apache Kafka

Open-source stream-processing software platform developed by LinkedIn and donated to the Apache Software Foundation

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Apache Kafka® is a distributed streaming platform.

What exactly does that mean?



More info on:
https://kafka.apache.org

A streaming platform has three key capabilities :

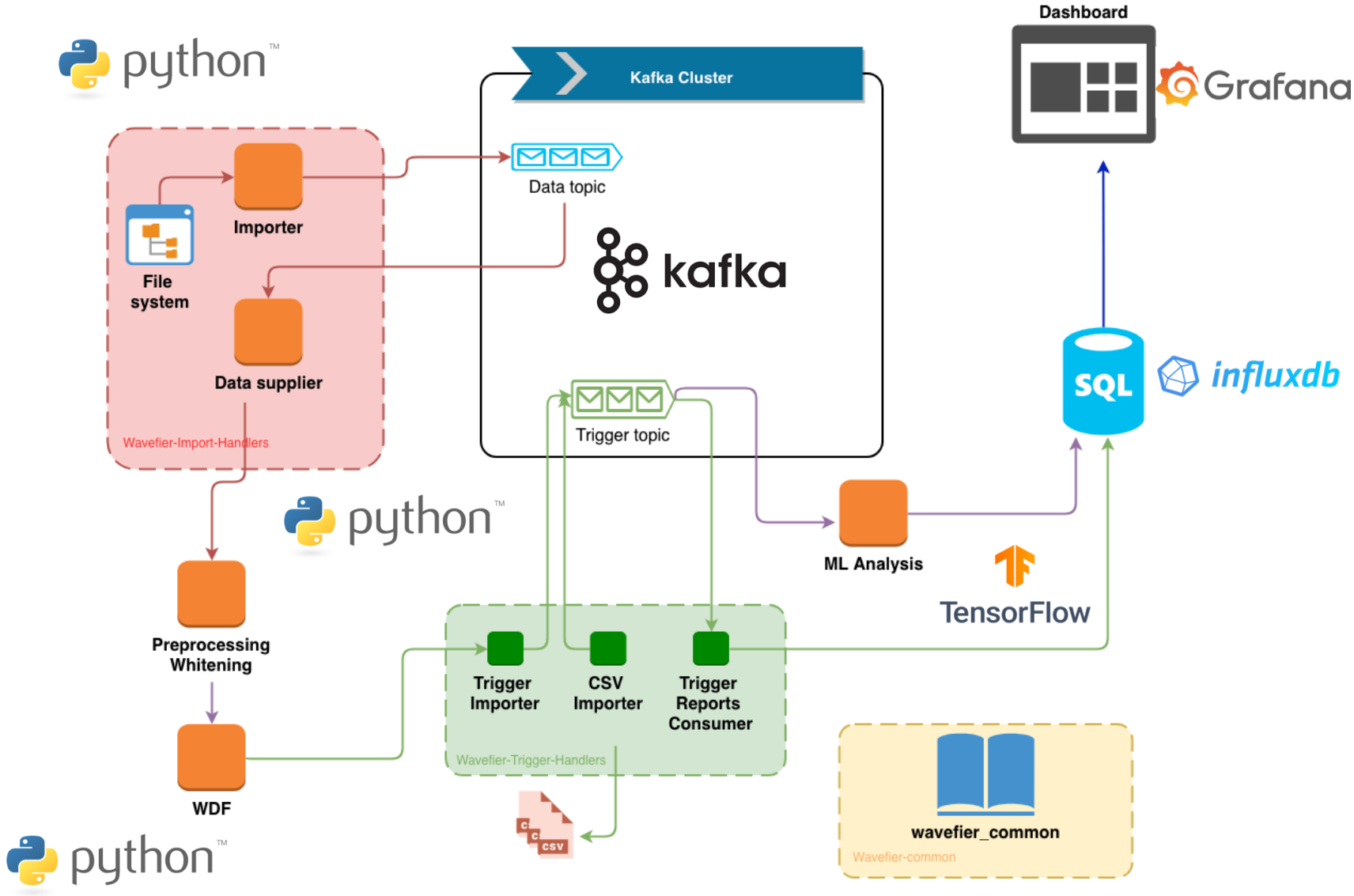
- Publish and subscribe to streams of records
- similar to a message queue
- Store streams of records in a fault-tolerant durable way.
- Process streams of records as they occur.

Kafka is generally used for two broad classes of applications:

- Building real-time streaming data pipelines that reliably get data
- Building real-time streaming applications that transform or react to the streams of data

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Software Deployment



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Grafana

"Grafana is the open source analytics & monitoring solution for every database "

Why grafana?

Useful build-in features

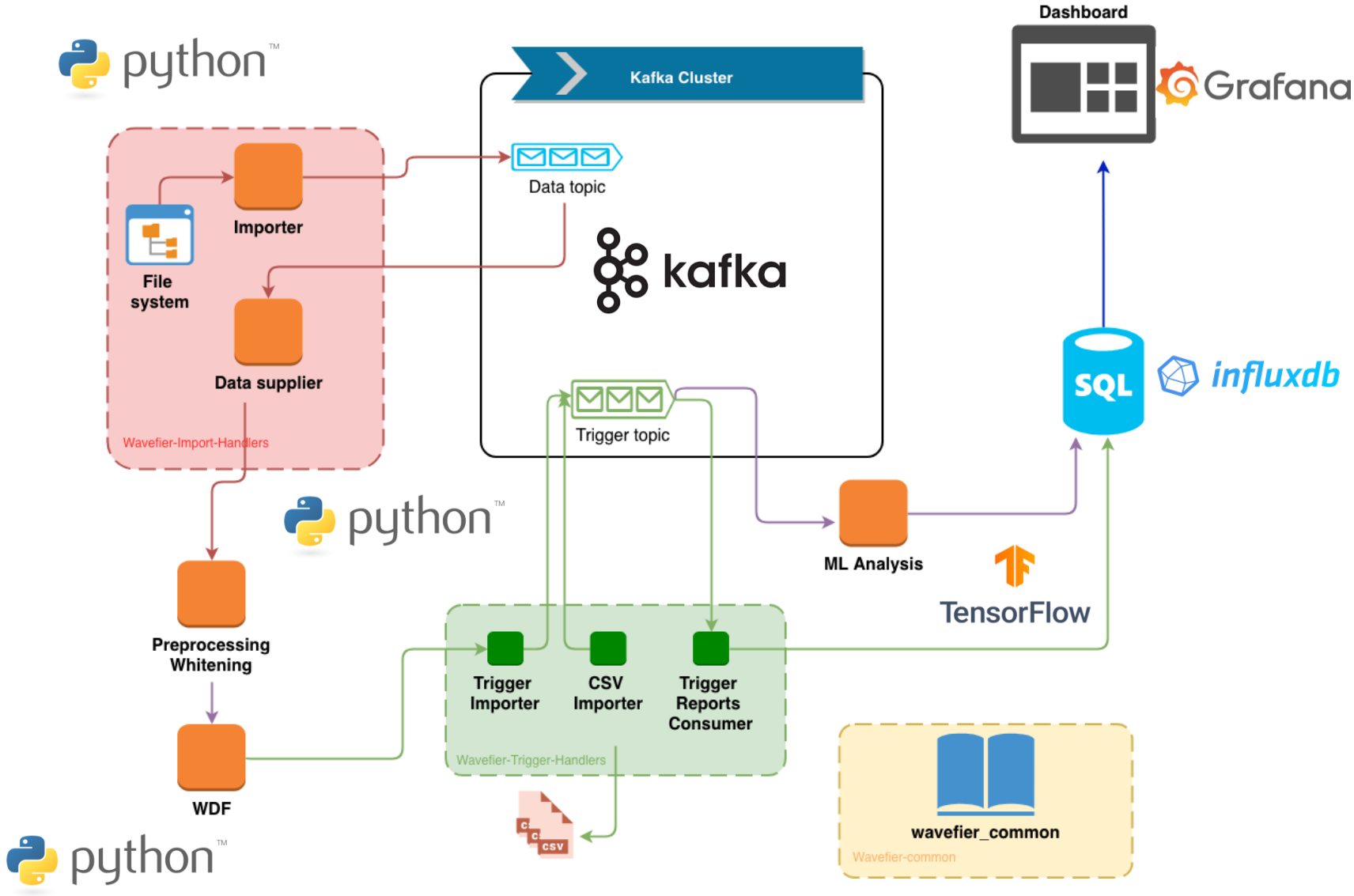
- Authentication, Organization and user settings

Mixed Datasource, Mix different data sources in the same graph

- Grafana supports dozens of databases, natively. Mix them together in the same Dashboard.

Native Notification and Alerting

Software Deployment



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

InfluxDB

"InfluxDB is a time series database designed to handle high write and query loads."

Why InfluxDB?

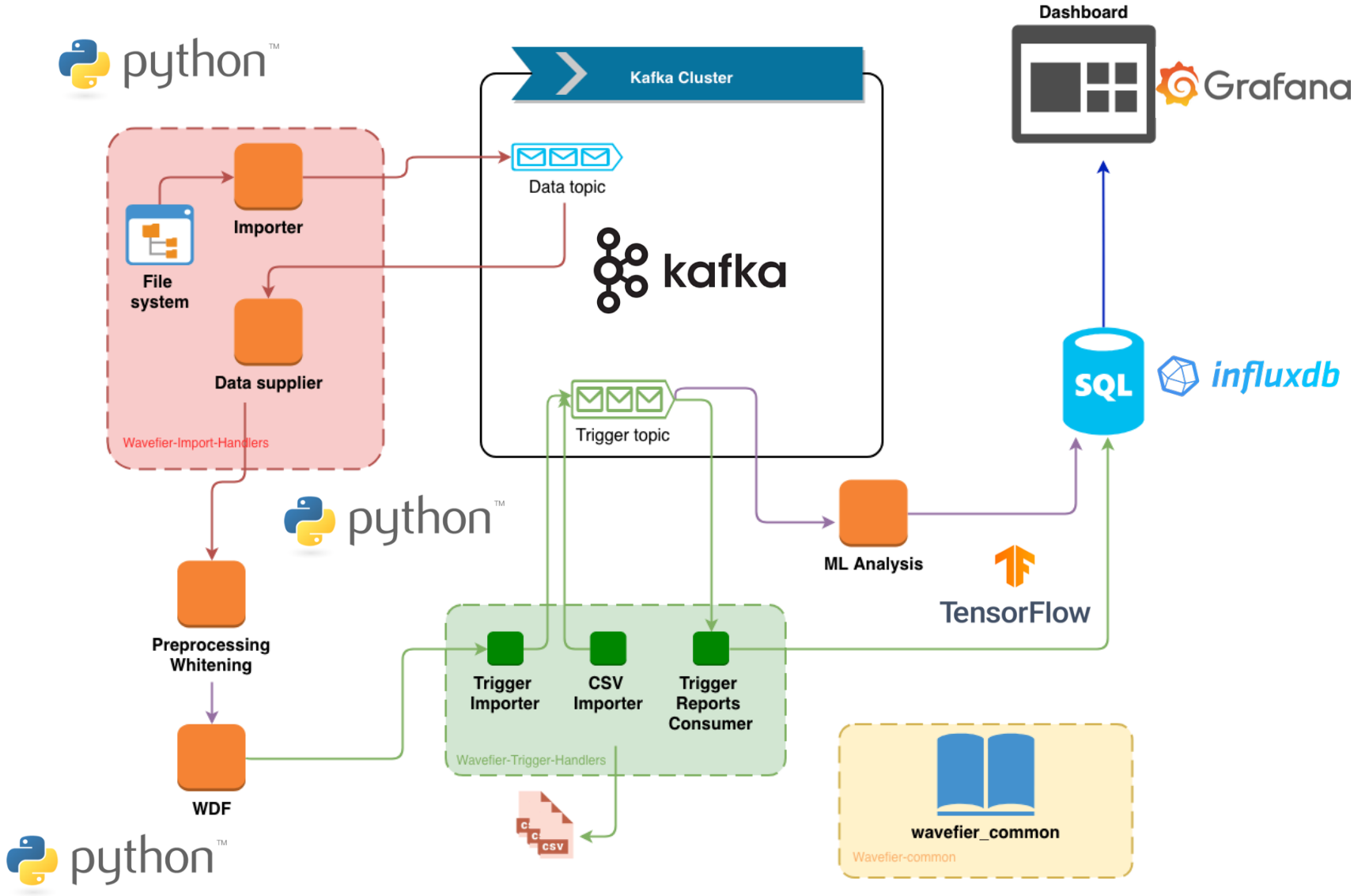
Specific for time series database (TSDB)

All is designed as time series

Friendly because InfluxDB have a SQL-like query language for interacting with it

Grafana has Native support for InfluxDB

Software Deployment



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Software Management

- Distribuite Team: Trust-it, EGO
- Different expertises and point of view: Physics, Software Engineer and Computer science
- Different Environment
- One unique target… the prototype

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

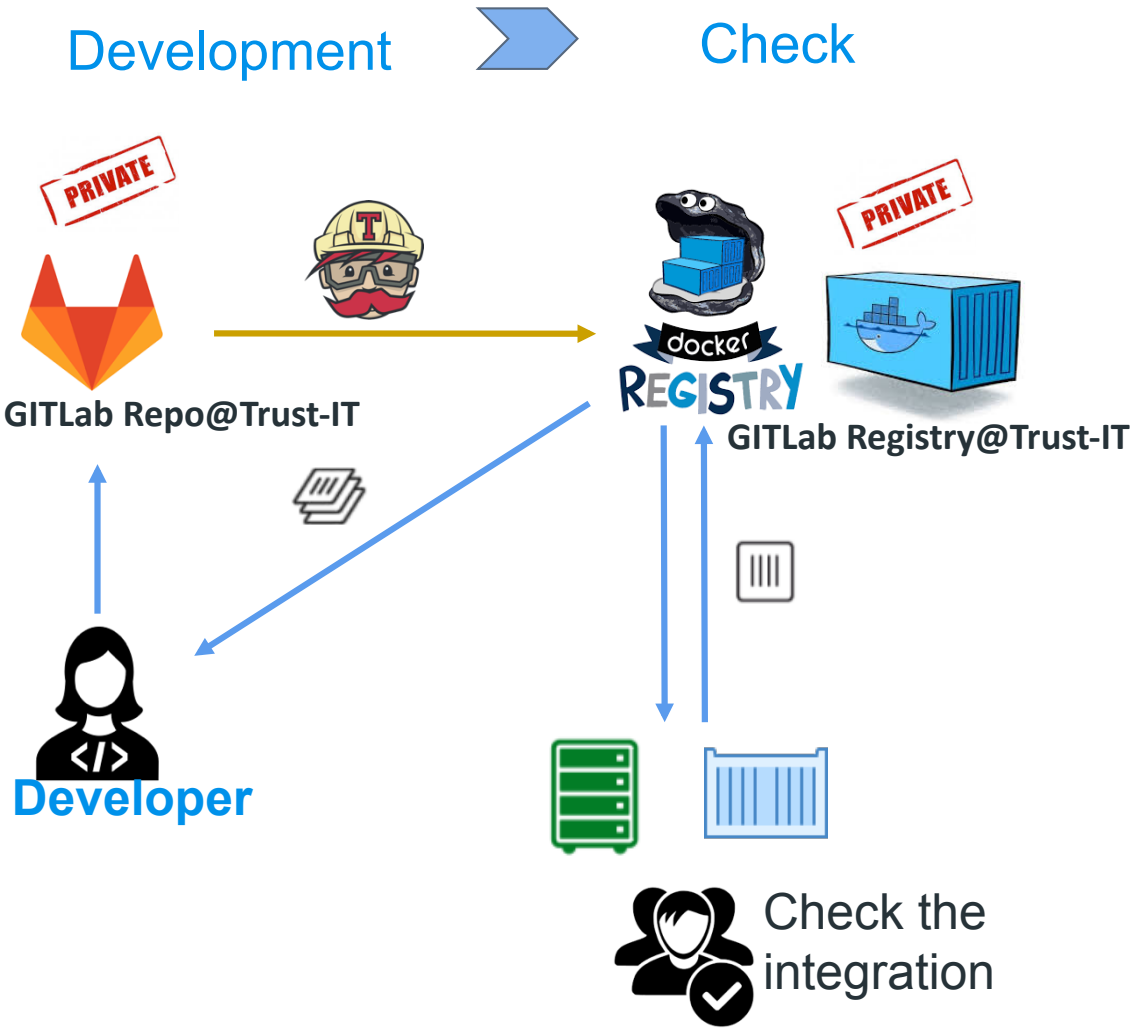
We use Docker!

Docker is an open platform for developers and sysadmins to build, ship, and run distributed applications.

Docker take the concept of container and build an ecosystem around it that would simplify its use

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

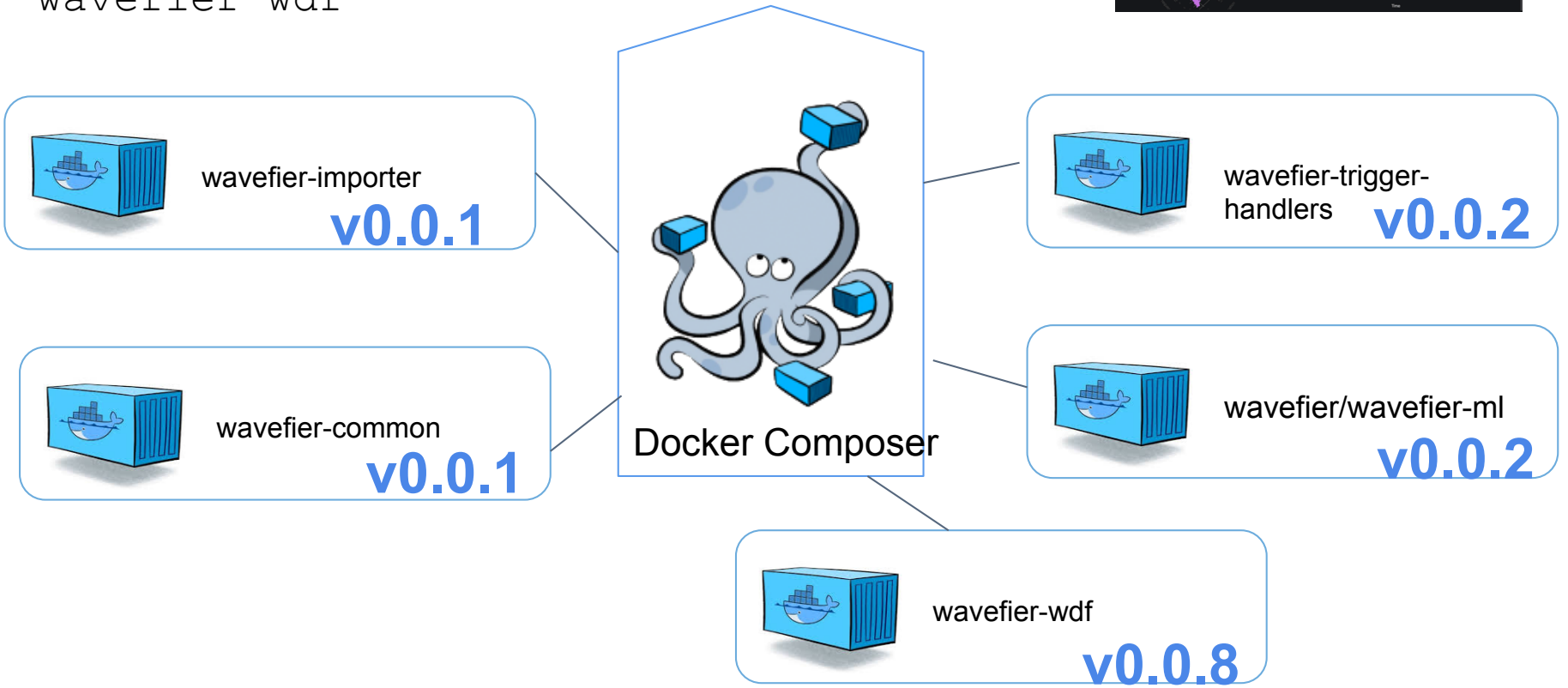
Automation and Continuous Integration



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

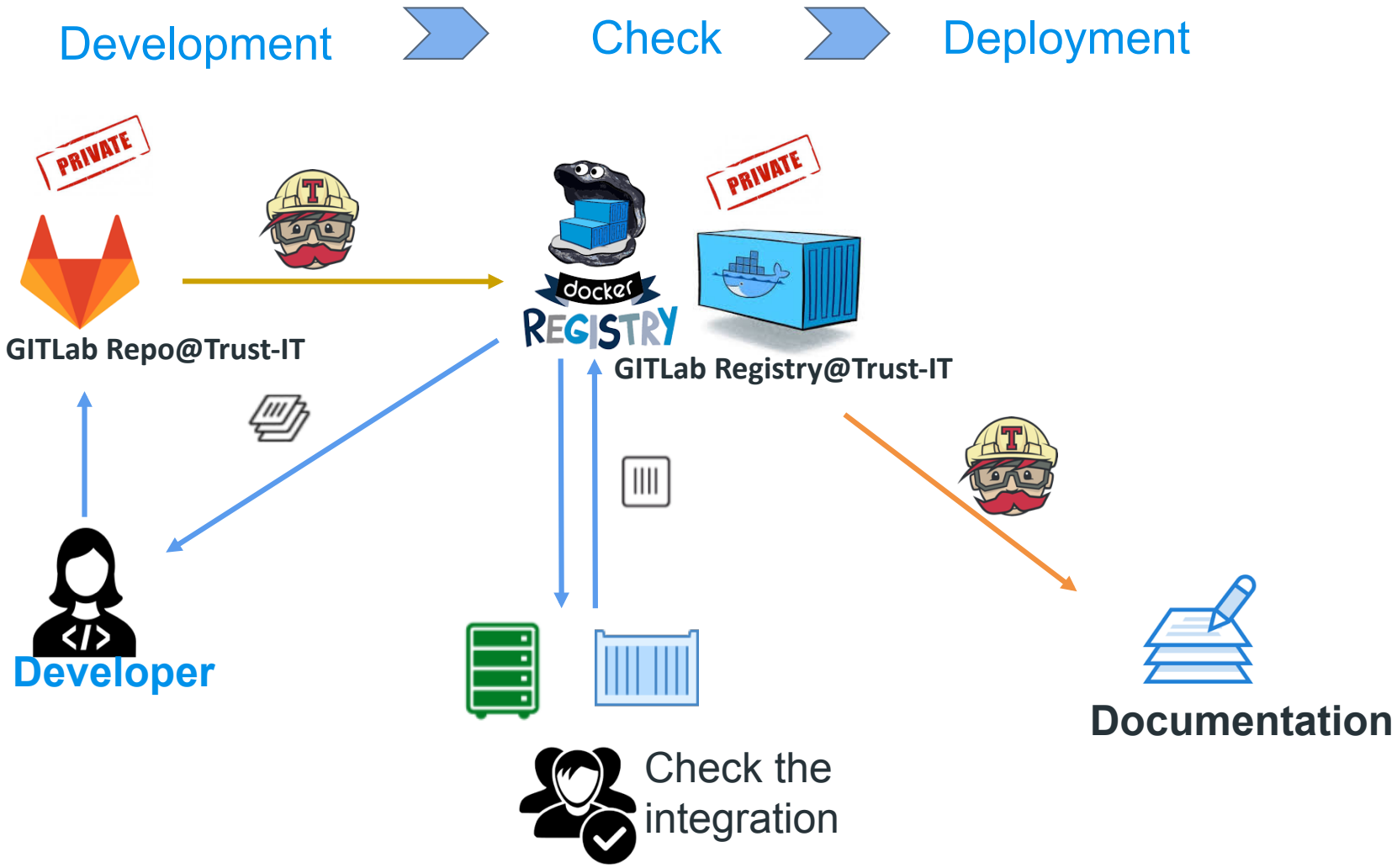
Check the integration - Local Deployment

```
#> docker-compose up
wavefier-wdf
```



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

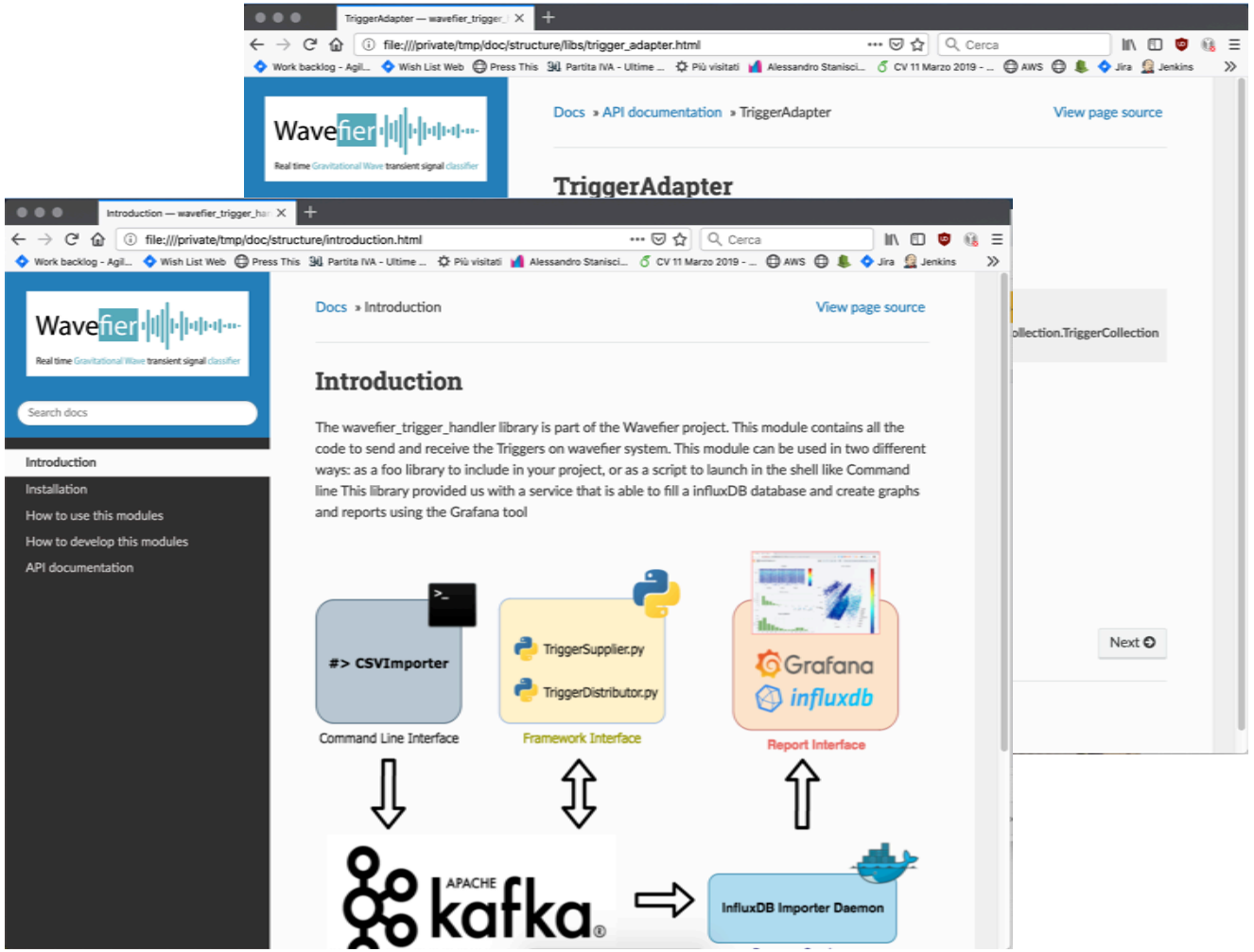
Automation and Continuous Integration



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

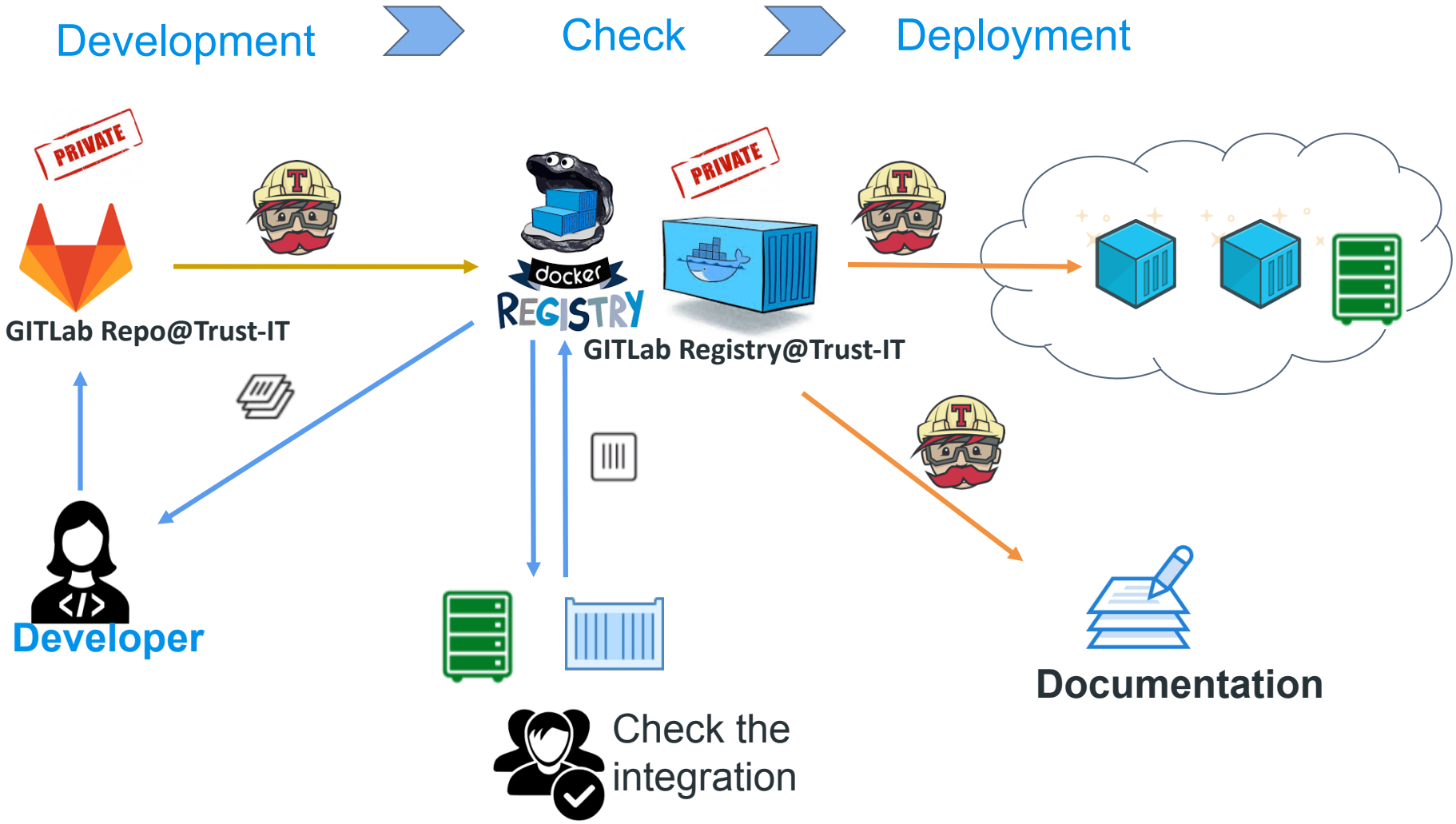
Automation on Documentation

The Documentation is also generated foreach git Commit



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Automation and Continuous Integration



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Kubernetes (K8s)

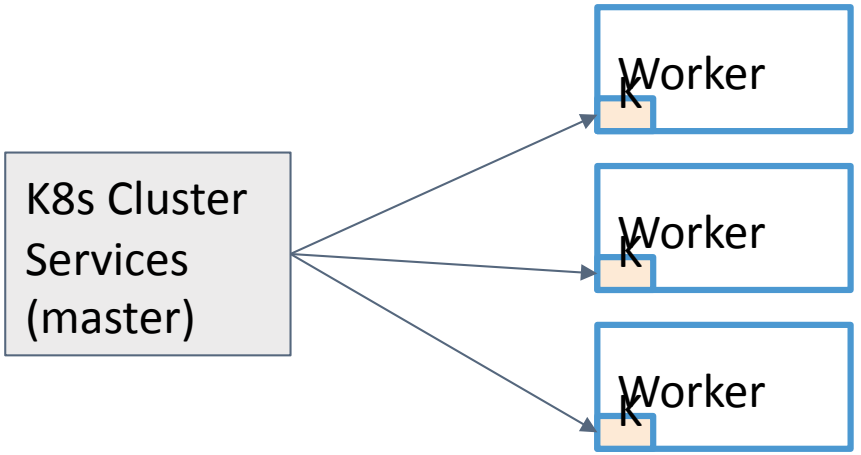
was a project spun out of Google as a open source next-gen container scheduler, designed as a loosely coupled collection of components centered around deploying, maintaining, and scaling applications.

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Architecture overview

Kubernets abstracts away the underlying hardware of the nodes and provides a uniform interface for applications to be both deployed and consume the shared pool of resources.

Masters
are responsible at a minimum for running the API Server, scheduler, and cluster controller. They commonly also manage storing cluster state, cloud-provider specific components and other cluster essential services.

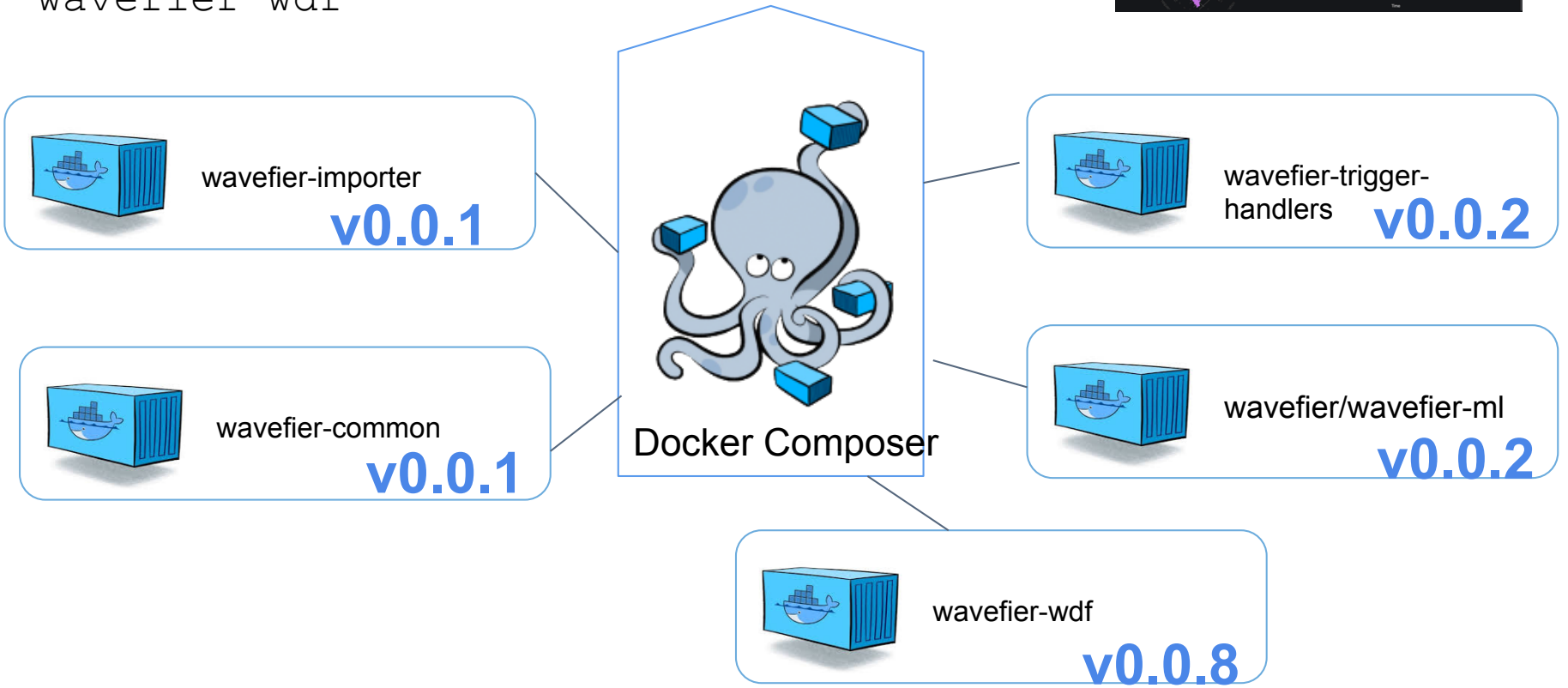


Nodes
Are the 'workers' of a Kubernetes cluster. They run a minimal agent that manages the node itself, and are tasked with executing workloads as designated by the master.

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

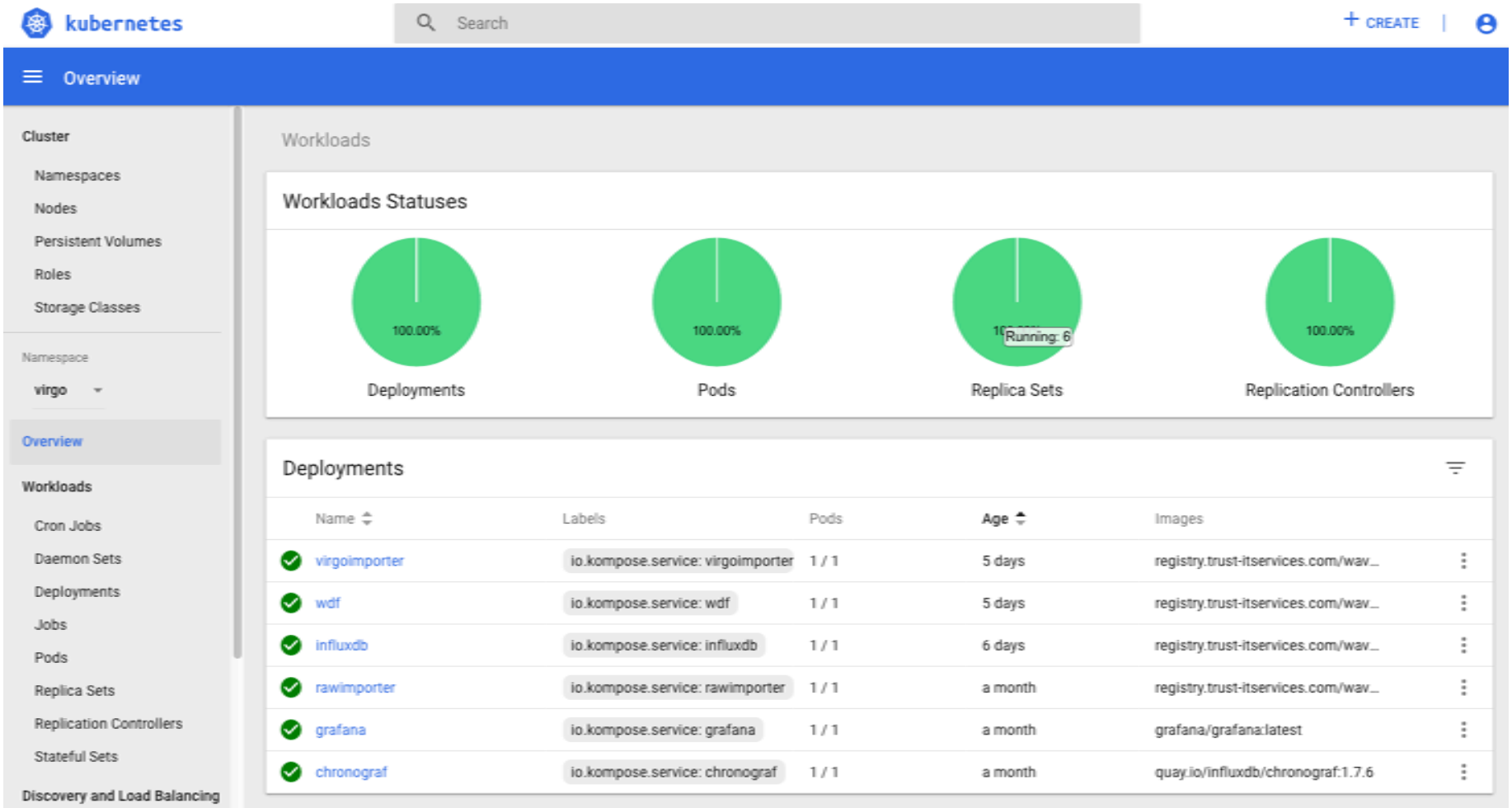
Do you remeber local deployment?

```
#> docker-compose up
wavefier-wdf
```



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

WaveFier running on Kubernetes



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

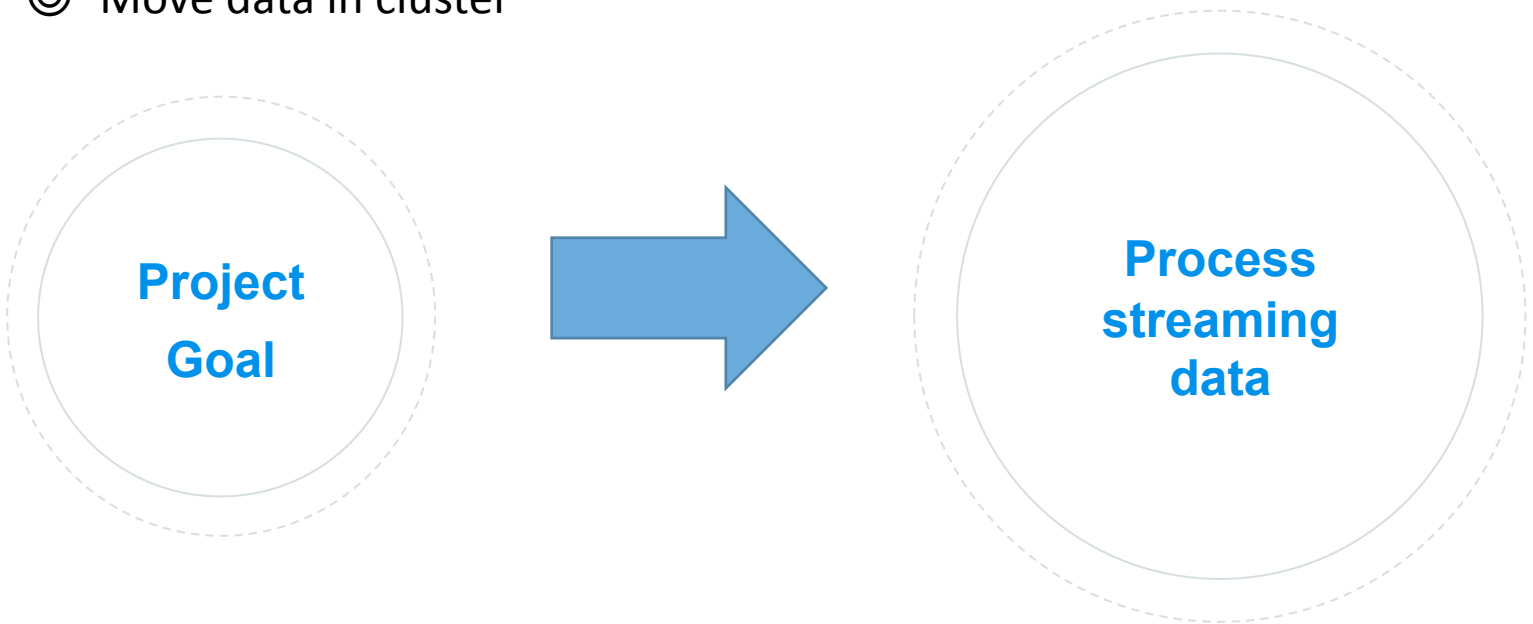
From Offline data to Online data

Offline data

- Pick-up interferometer data
- Store data in files
- Access to cluster
- Move data in cluster

Online data

- Receiving data from different sources
- Receive data from interferometer

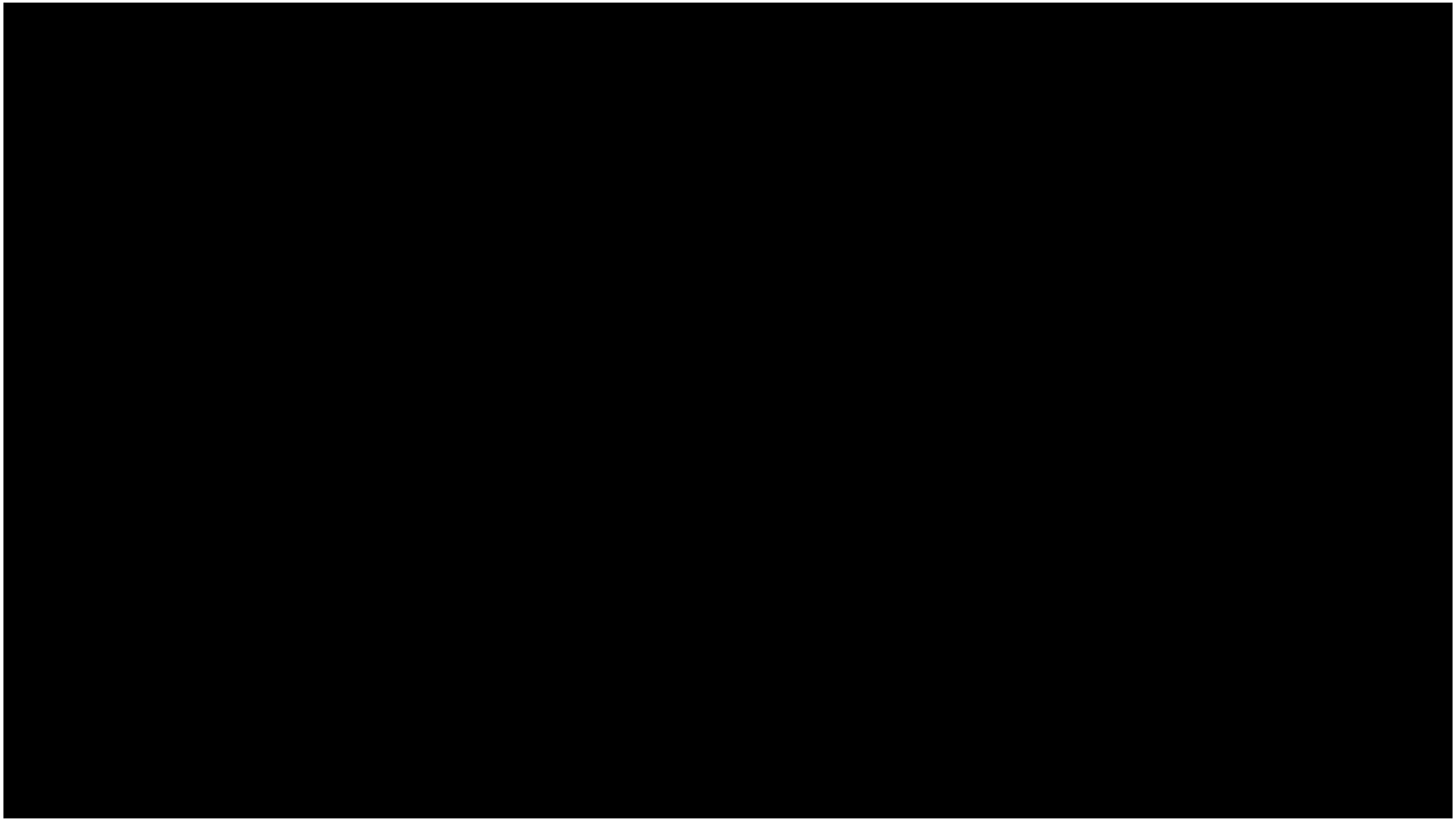


H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Online data - Deployment

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Current Grafana Dashboard with Classification Results



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

V

Next step?

Move from prototype/development to production

Investigate the use of much more sophisticated Machine/Deep learning algorithm, using GPU

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

V. Next step?

Wavefier' repos

DOI: 10.5281/zenodo.3356656

https://zenodo.org/record/3356656
https://repository.asterics2020.eu/content/wavefier
https://gitlab.in2p3.fr/escape2020/wavefier/dockers

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

CREDITS

LAPP and CNRS: Giovanni Lamanna, Jayesh Wagh

Trust-IT: Silvana Muscella, Emanuel Marzini, Filip Morawski, Alessandro Petrocelli, Alessandro Staniscia

And Many thanks to **GARR CSD support**

H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).