

# COMPUTING

Stefano.Bagnasco@to.infn.it

# THE MANDATORY SLIDE WITH BOXES AND ARROWS

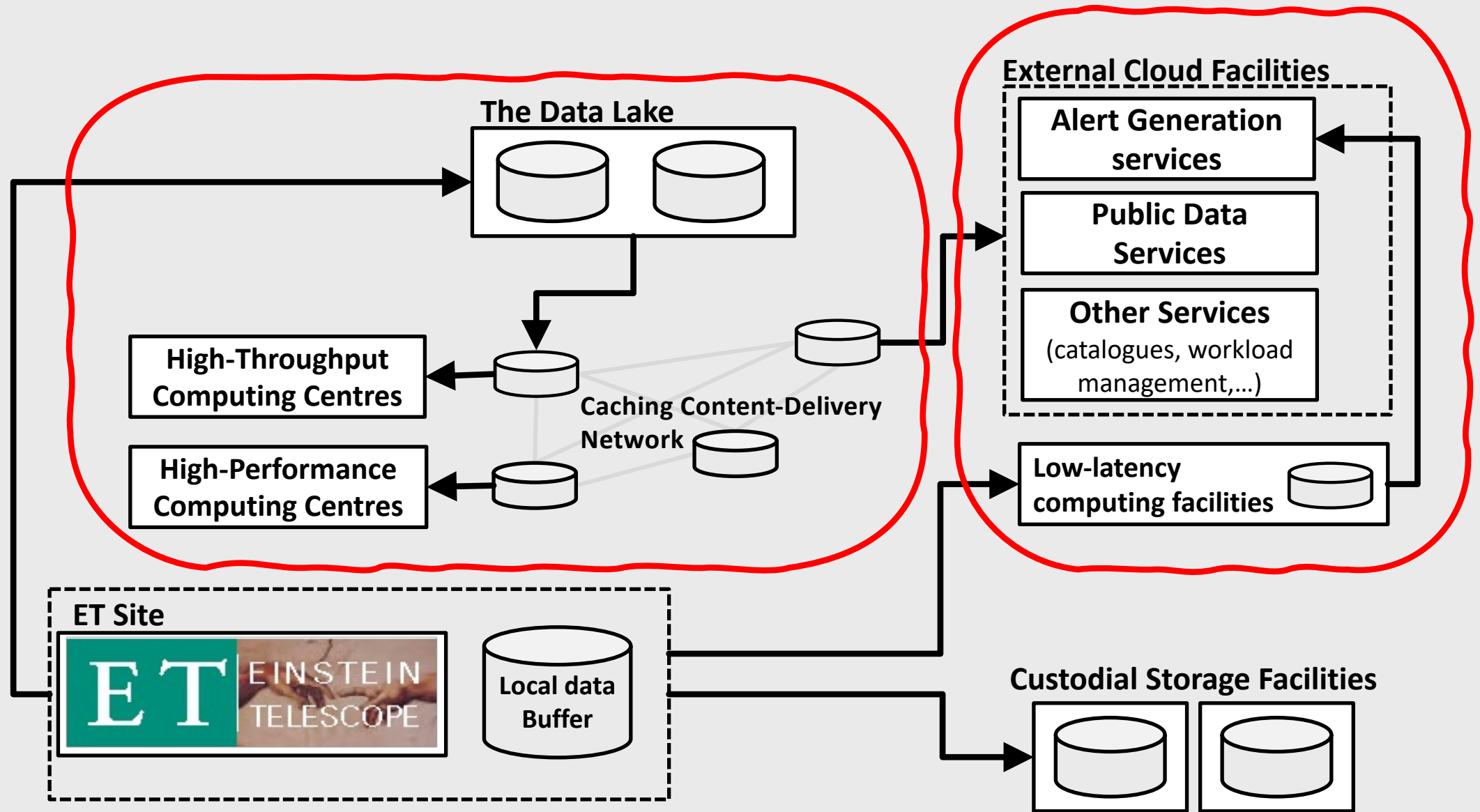


Figure from the ET ESFR-I proposal

# THE MANDATORY SLIDE WITH BOXES AND ARROWS

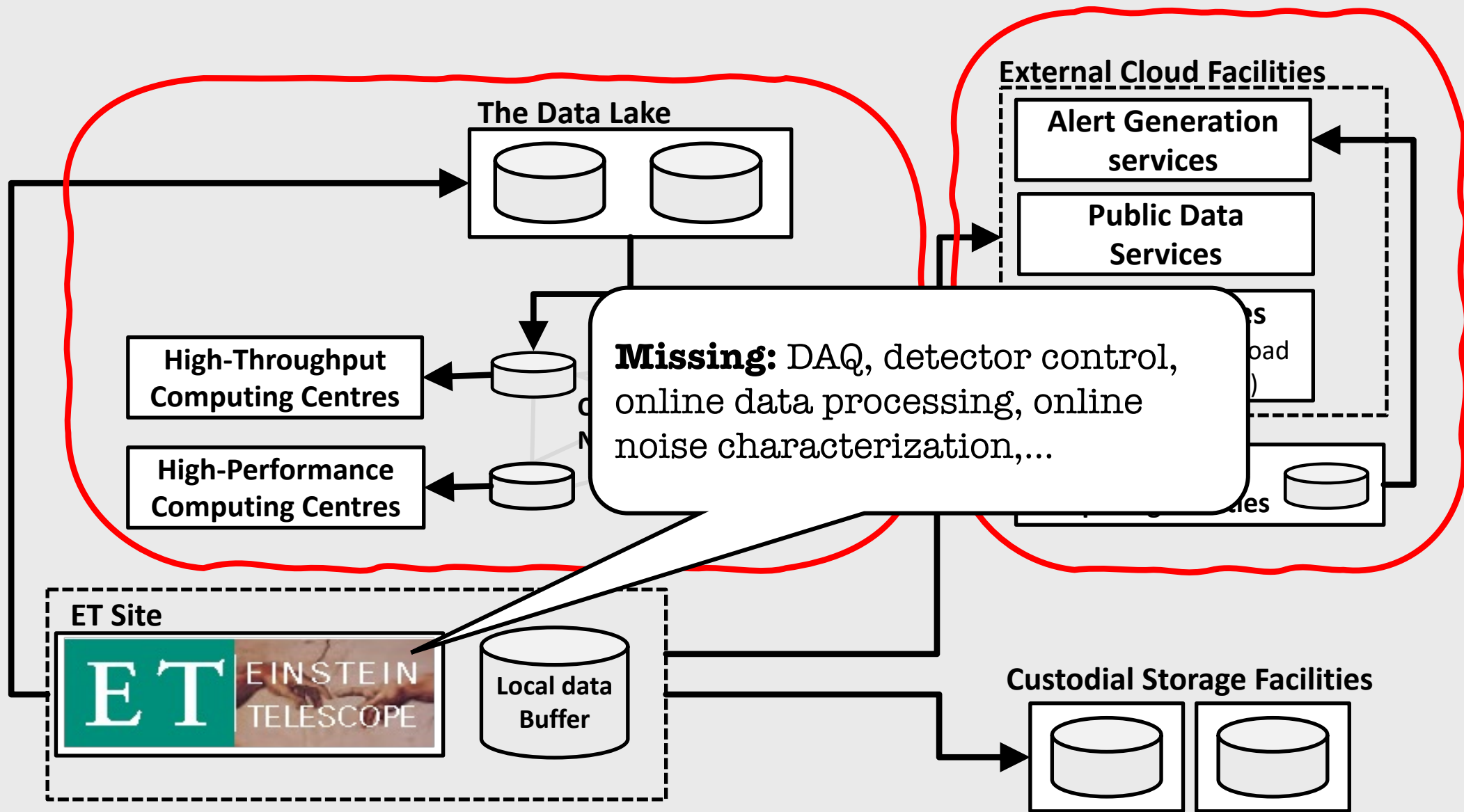


Figure from the ET ESFR1 proposal

# SHOPPING LIST

**Data transfer and storage:** safely and efficiently transfer all data to custodial storage and processing centres, including low-latency transfers;

**Software packaging and distribution:** manage software lifecycle, and make packages available ubiquitously;

**Computing power:** provide and manage computing resources (HTC and HPC) for the processing of data, in all computing domains;

**Data distribution:** make data available to worker nodes in computing centres anywhere, and possibly also to single workstations, including support to public releases of data;

**High-availability service management:** provide a platform for running the collaboration's services (e.g. alert generation services, event databases,...)

**Data cataloguing and bookkeeping:** organise all data and metadata and provide querying and discovering capabilities;

**Job lifecycle management:** provide a uniform job submission and runtime environment to research groups;

**High-level workload management:** keep a database of all jobs and allow the enforcement of priorities and scheduling strategies; provide support for organized large-scale data processing campaigns;

**Monitoring and accounting:** monitor local and distributed computing, checking performance and looking for issues, and provide reliable accounting both at the user/job and site level;

**Authentication, Authorisation and Identity management:** provide consistent AAI across all domains and activities.

**Collaboration services:** provide tools for efficient collaboration management, coordination, and outreach (e.g. document repositories, collaborative tools, administrative databases, communications,...)

# TWO CHALLENGES

- Current data analysis strategies (e.g., straightforward matched filtering) do not scale to the 3G expected rates
  - We don't even know exactly how to estimate the computing needs
  - And of course we will have to devise new strategies (and develop new algorithms and code)
  - «10% of an LHC experiment»
  - Mock Data Challenges will serve several purposes: besides working on analysis code, iteratively study a way of assessing computing requirements and develop a first distributed computing infrastructure
- Multimessenger alert exchange and analyses with many data sources and very high rates
  - Everything needs to be fully automatic (including selection of interesting alerts by consumers)
  - E.g., time-evolving alerts database, “native” multimessenger low-latency analyses,...

# THE ET E-INFRASTRUCTURE BOARD

**EIB:** e-Infrastructure Board (i.e., Computing Board)

**Chairs:** Patrice Verdier (BSC), SB (INFN Torino)

**Division 1:** Software, frameworks, and data challenge support

**Chair:** Andres Tanasijczuk (UCLouvain)

**Division 2:** Services and Collaboration Support

**Chair:** Antonella Bozzi (EGO)

**Division 3:** Computing and data model, Resource Estimation

**Chair:** Gonzalo Merino (PIC)

**Division 4:** Multimessenger alerts infrastructure

**Chair:** Steven Schramm (UGeneva)

**TTG:** Technology Tracking working Group

**Chair:** Sara Vallero (INFN Torino)

**ET-PP WP8:** Computing and Data Model

**Chairs:** Nadia Tonello (BSC), Achim Stahl (UAachen)



# SHORT TERM (I.E., WE NEED IT **NOW**)

- IAM
  - Ongoing (Cyfronet), but room for collaboration
- Service hosting & management: e.g., GitLab
  - Including integration and exploitation model definition
  - Definition of our code development model and best practices
  - Generic “cloud” facility for service deployment?
- Mock Data Challenges and their support
  - Including iterative requirement gathering and model/service evolution
  - Starting from IGWN-like data distribution, evolve it through subsequent MDC rounds collecting feedback by users

# LONGER TERM (VARIABLE URGENCY)

**Timescale:** requirement-gathering workshop organized by ET-PP WP8 in Geneva in October

- Definition of the Computing Model
  - This is a deliverable of ET-PP WP8
- Definition of the data model
  - And data-management model
  - Part of the Computing Model
- Computing resource need assessment
  - Both for offline and low-latency
- Frameworks for “managed” workflows and data
  - I.e., what is often done with DIRAC and RUCIO or their equivalent



# LONGER TERM (VARIABLE URGENCY), CONT'D

- ESCAPE-related activities
  - Data lake (possibly for upcoming MDC rounds). INFN (Torino group, possibly CNAF) might be interested
  - Virtual Research Environment assessment (same)
  - Virtual Observatory standards (e.g., TS-MOC tools by Giuseppe Greco, INFN Perugia)
  - Wavefier low-latency signal classification pipeline evolution (Elena Cuoco group, SNS Pisa)
- Low-latency analysis and alert infrastructure
  - Alert system work now mostly a liaison activity with other communities
  - Developing a modern deployment system for existing LL suite on K8S (Sara Vallero, INFN Torino)
  - M2Tech proposal for next INFRA-TECH call (Steven Schramm, UGeneva but expected INFN participation)
- HPC access and interfaces
  - For “traditional” data processing and ML technologies
  - INFN ICSC project interest to be investigated
- ML applications in low-latency
  - See e.g., InterTwin noise glitch study using generative AI: development of both model and infrastructure (INFN Torino and INFN Pisa)
  - As a “pilot” exploration of challenges: continuous retraining, low- (or even fixed-) latency inference, vetoing or de-noising

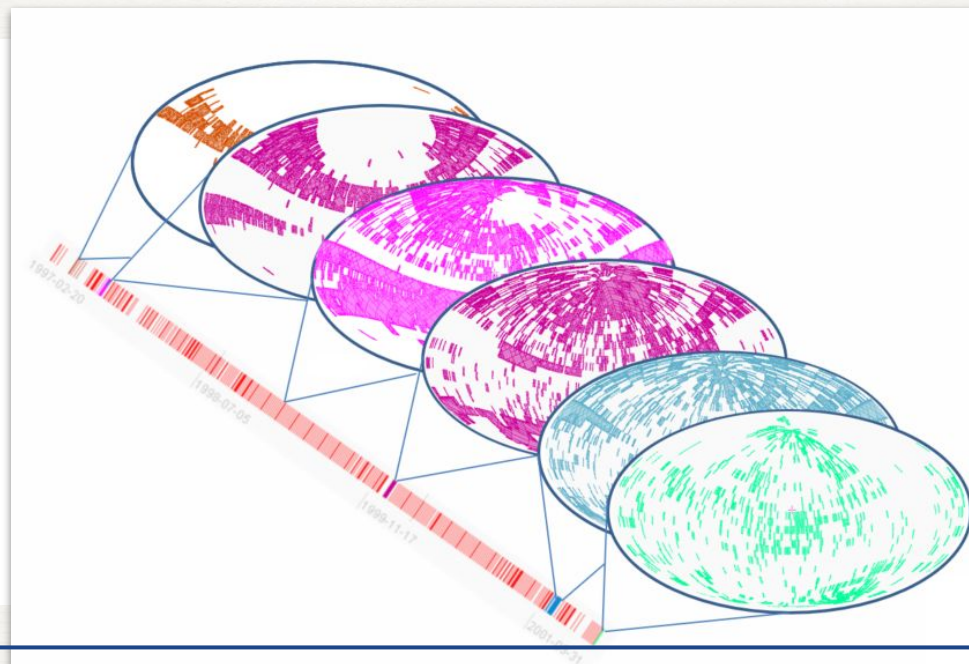
# Spatial and Temporal MOC: ST-MOC



*International  
Virtual  
Observatory  
Alliance*

MOC: Multi-Order Coverage map  
Version 2.0

IVOA Working Draft 2020-10-30

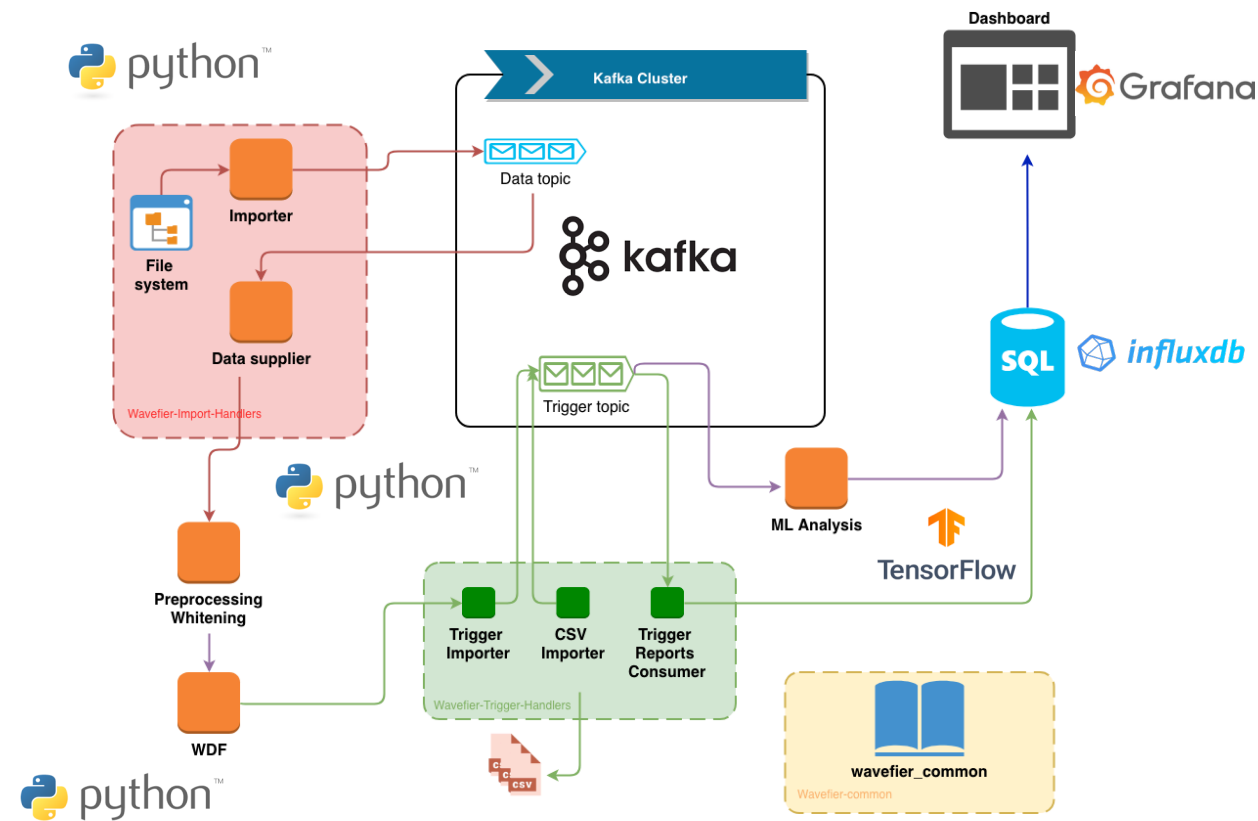


At a given Time range we obtain the corresponding Spatial coverage.

giuseppe.greco@pg.infn.it

## Software Deployment

IV. The technical and implementation aspects

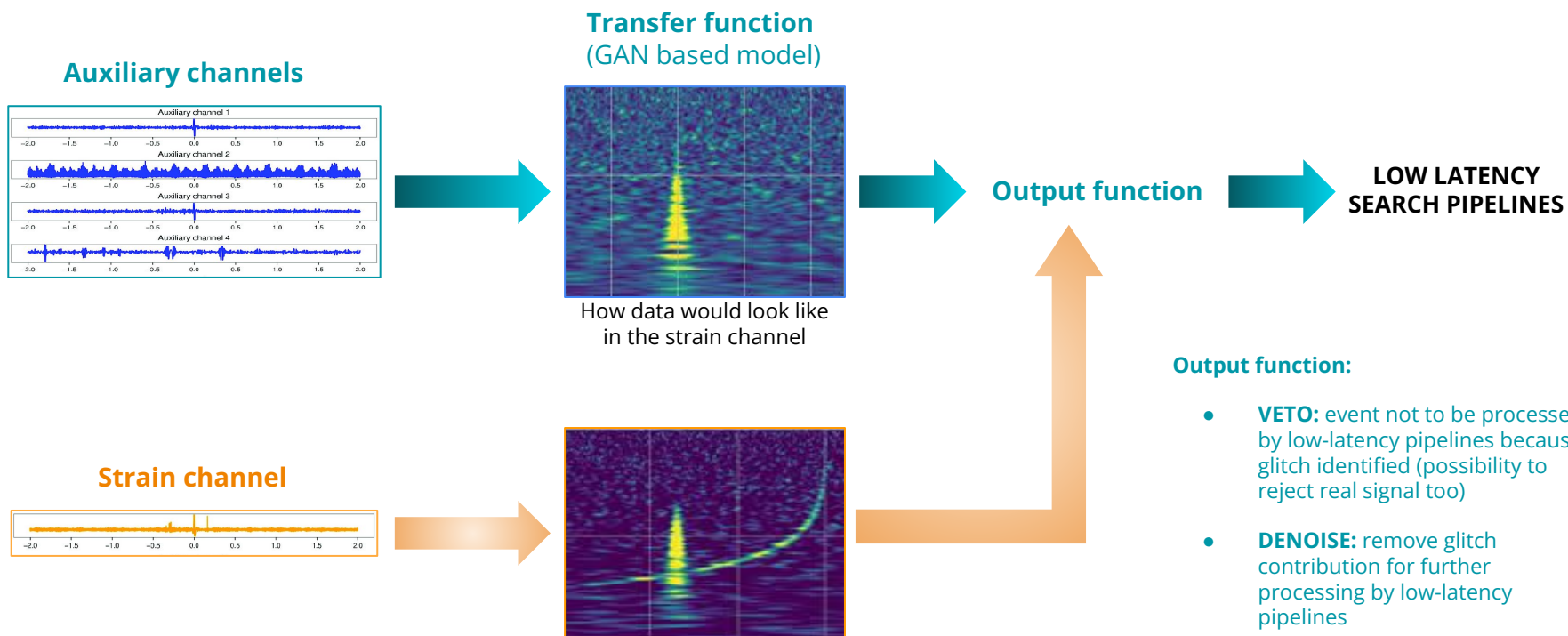


H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

elena.cuoco@sns.it



## WP 4.4: The Digital Twin



13

sara.vallero@to.infn.it

# ETIC-FUNDED COMPUTING FACILITIES

- BETIF in Bologna

- Mostly DAQ and Online Data Processing-related
- Use of GPUs, FPGAs etc.
- WhiteRabbit
- Also: ML technologies
- Contact: [riccardo.travaglini@bo.infn.it](mailto:riccardo.travaglini@bo.infn.it)

- CTLab in Torino

- Small heterogeneous HPC cluster as a platform for HW and SW testing, technological R&D and Technology Tracking
- Scalable thanks to synergy with larger INFN computing projects (TeRABIT and ICSC)
- Contact: [stefano.bagnasco@to.infn.it](mailto:stefano.bagnasco@to.infn.it)

# THANKS!