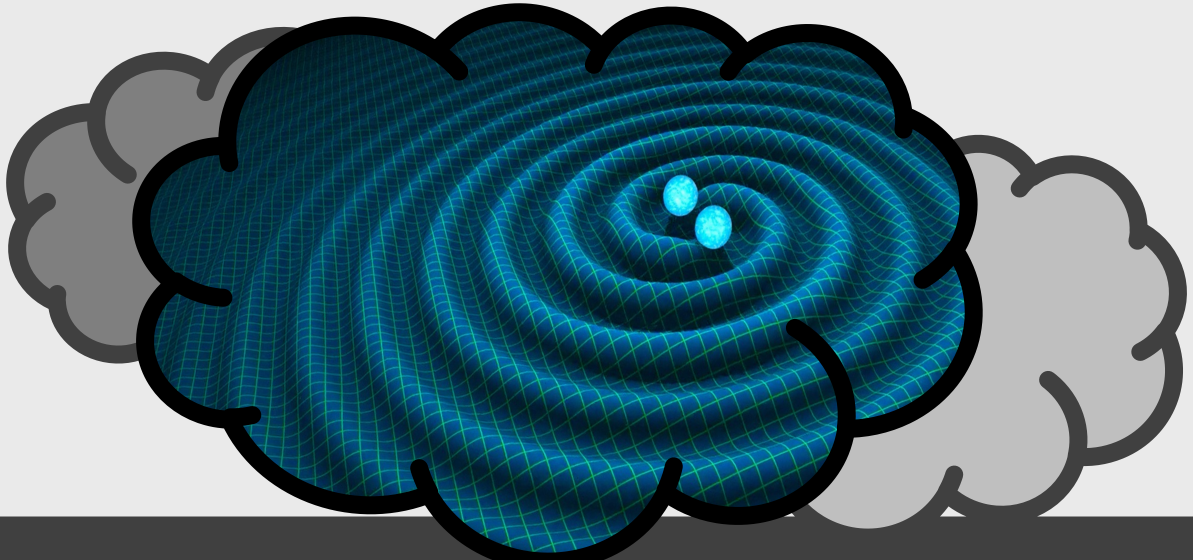


Computing for the Einstein Telescope

Stefano Bagnasco

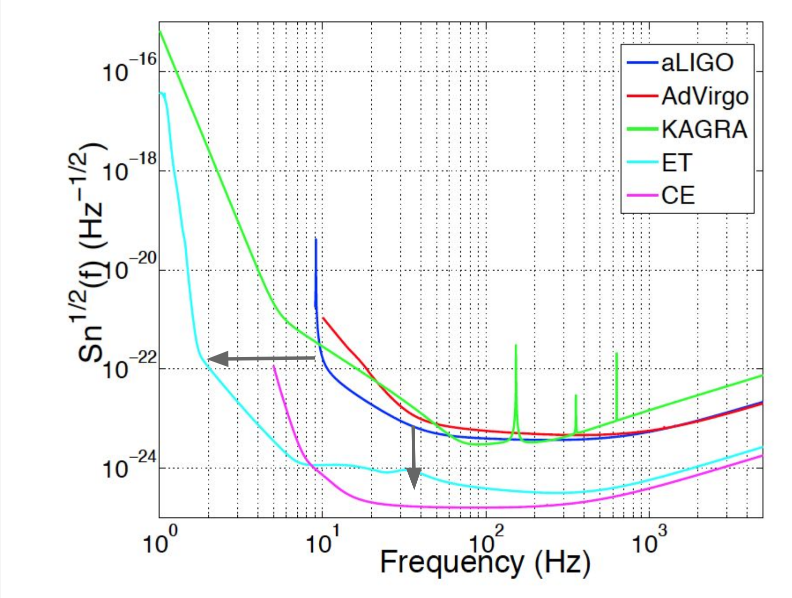
on behalf of the ET-EIB and INFRA-DEV WP8

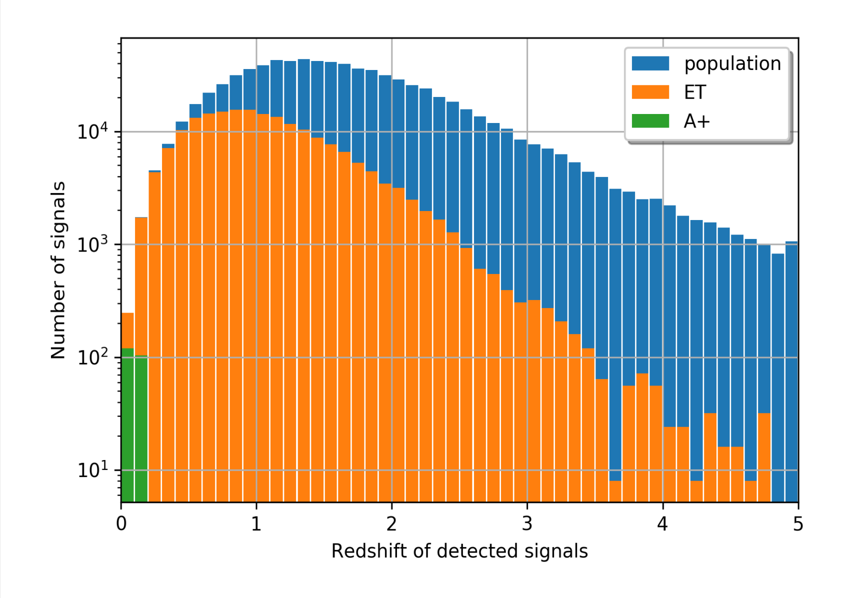
ET-EIB kickoff Workshop | EGO Nov 30, 2021

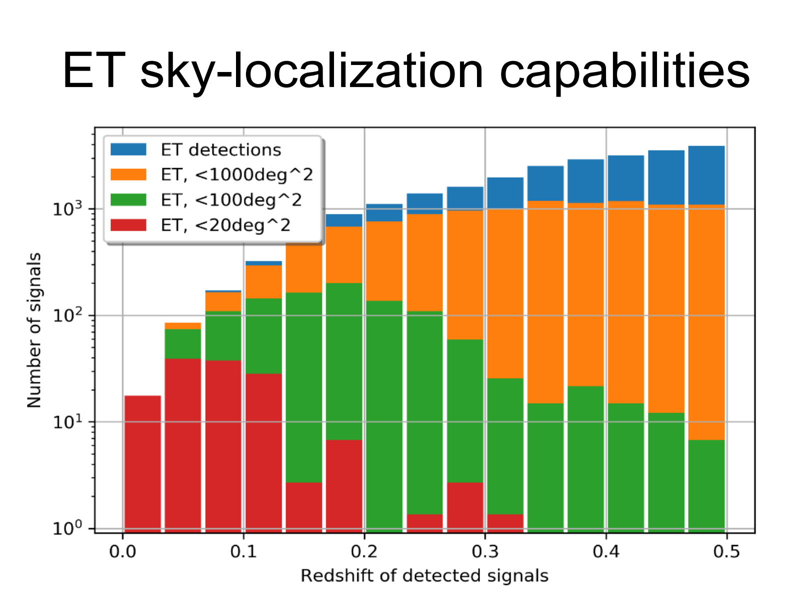
Towards 3G GW computing

- For past runs of 2G detectors, the focus has been on the detection of GW signals
- Focus is already shifting towards observation and MM astronomy triggers
  - In-depth parameter estimation of large numbers of events
  - Early warning alerts for a large number of events
  - High level of automation
- High sensitivity and low frequency cutoff are not your friends
  - See talks from OSB!

Enhanced sensitivity







marica.branchesi@gssi.it

- Lower frequencies (down to 1-5Hz)
- 10^5 BBH detections per year
- o(100) detections per year with <20 deg^2
- Much higher sensitivity
- 10^5 BNS detections per year
- Early warning by minutes (hours)

Some challenges

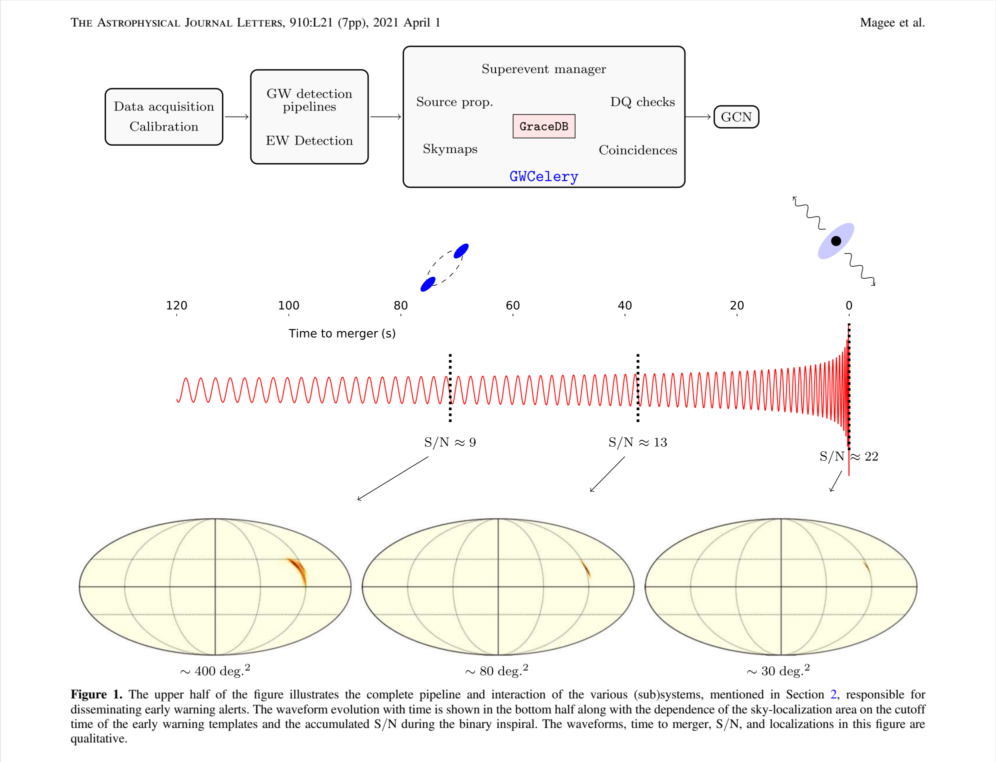
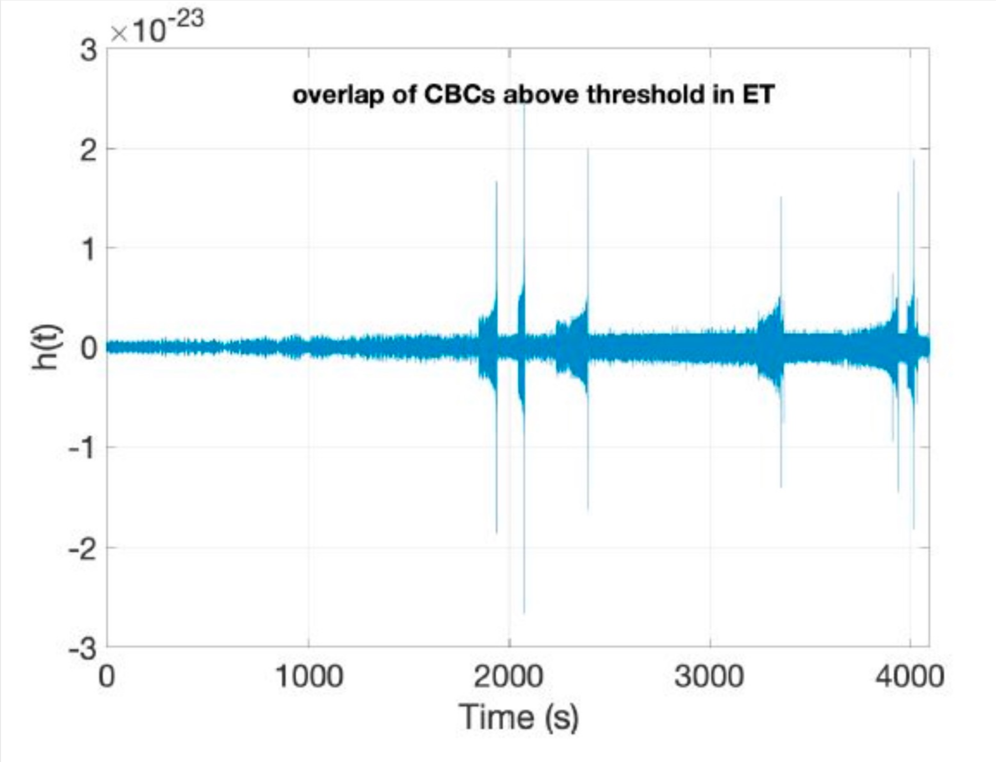


Figure 1. The upper half of the figure illustrates the complete pipeline and interaction of the various (sub)systems, mentioned in Section 2, responsible for disseminating early warning alerts. The waveform evolution with time is shown in the bottom half along with the dependence of the sky-localization area on the cutoff time of the early warning templates and the accumulated S/N during the binary inspiral. The waveforms, time to merger, S/N, and localizations in this figure are qualitative.



- Overlapping signals
- Long duration waveform for CBCs (and moving detector)
- FAR estimate in the presence of a strong foreground
- Environmental correlated noise

Data and more data

- Raw interferometer data don't grow much with increasing instrument sensitivity
  - Current GW detectors are writing o(1PB) per year of raw data per detector
  - Pre-processed data for user analysis is more than 1 order of magnitude smaller
  - In ET we expect about few tens of PB of raw data per year (baseline 6-IFO design, more control channels,...)
  - No big deal today, piece of cake by 2035
- What grows is the amount of useful scientific information embedded in the data
  - And the computing power needed to wring it out
  - It's a task in itself to precisely estimate the computing power needs

ET EINSTEIN TELESCOPE

$1/10^{th}$ of an LHC experiment

- Current computing needs of the entire GW network roughly o(10%) of an LHC experiment
- In ET the event rate will be 10^3 -10^4 times the current one
  - Analysis of the “golden” events (EM counterparts, high SNR or “special” events) would already be within reach using current technologies
    - O(500) events per year = 12.5MHS06-y per year, the same order of magnitude of a LHC experiment in Run 4
    - Target: $1/10^{th}$ of an LHC experiment in Run 4
- But: to be done (mostly) in low-latency!
- Need to start early Mock Data Challenges to develop and validate everything

Three computing domains

On-site infrastructure

Online

- Data acquisition and pre-processing
- Instrument control
- Environmental monitoring
- …

(Mostly) plain old HTC and HPC

Offline

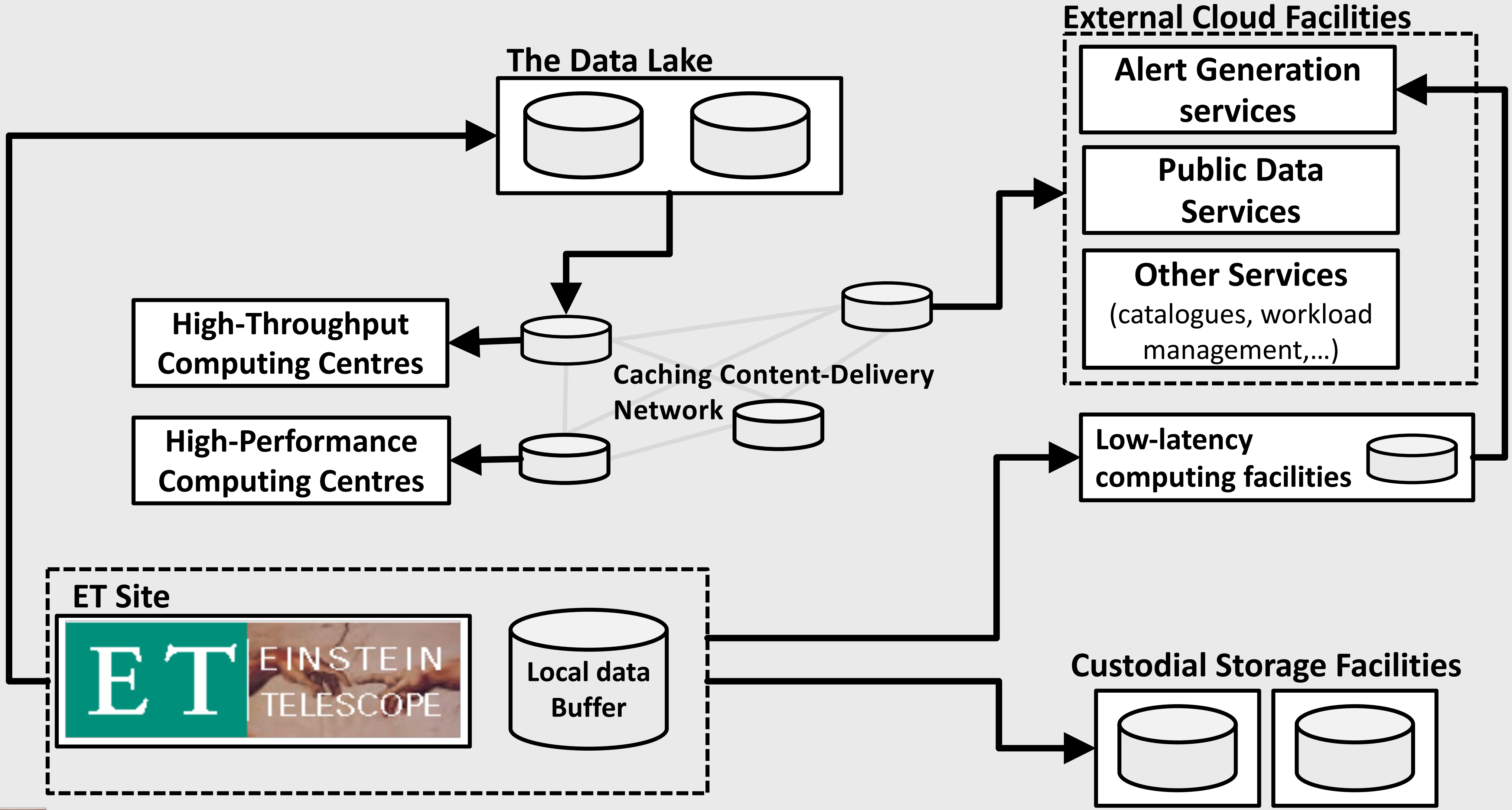
- Deep searches
- Offline parameter estimation
- (Template bank generation)
- …

Here's the fun

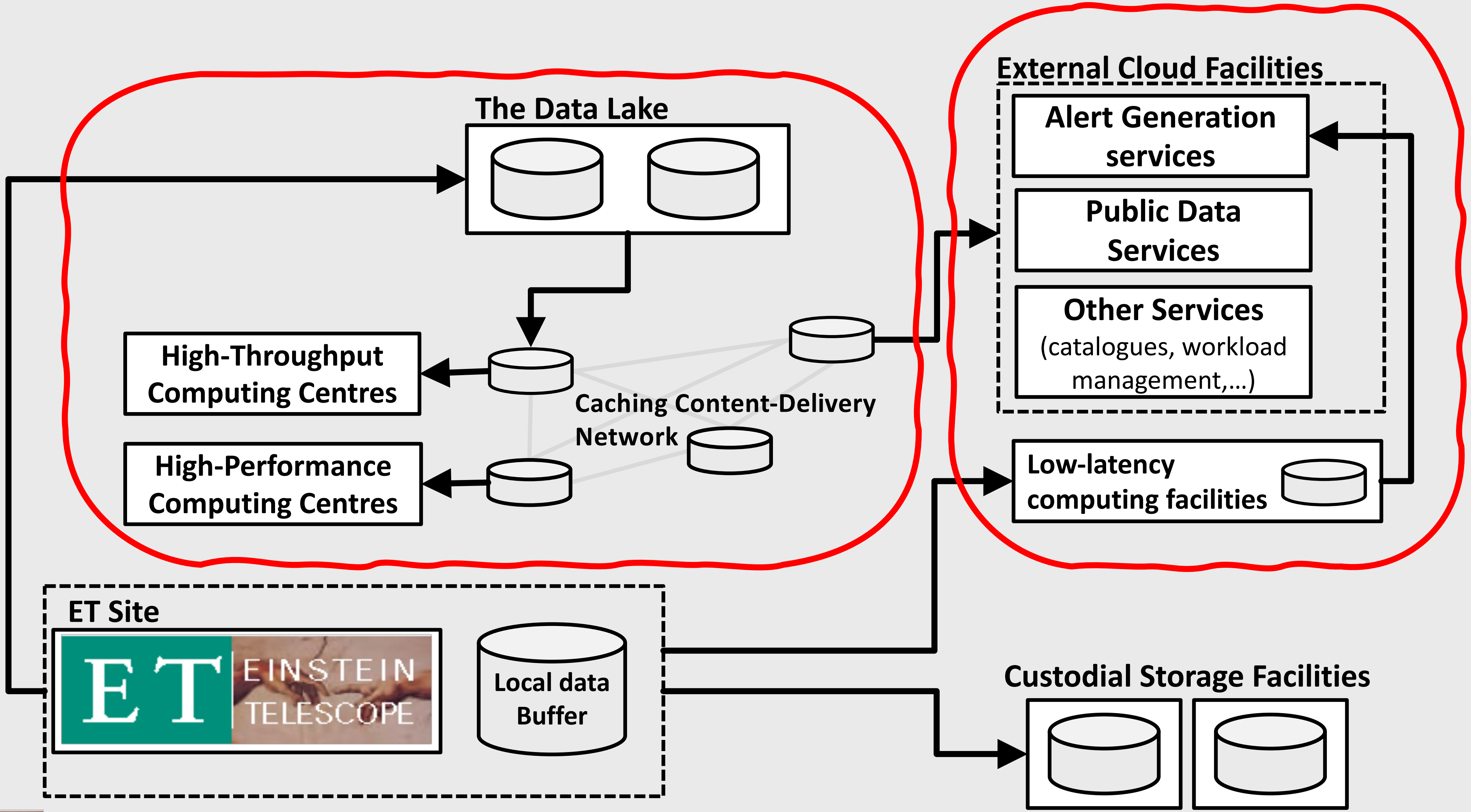
Low-latency

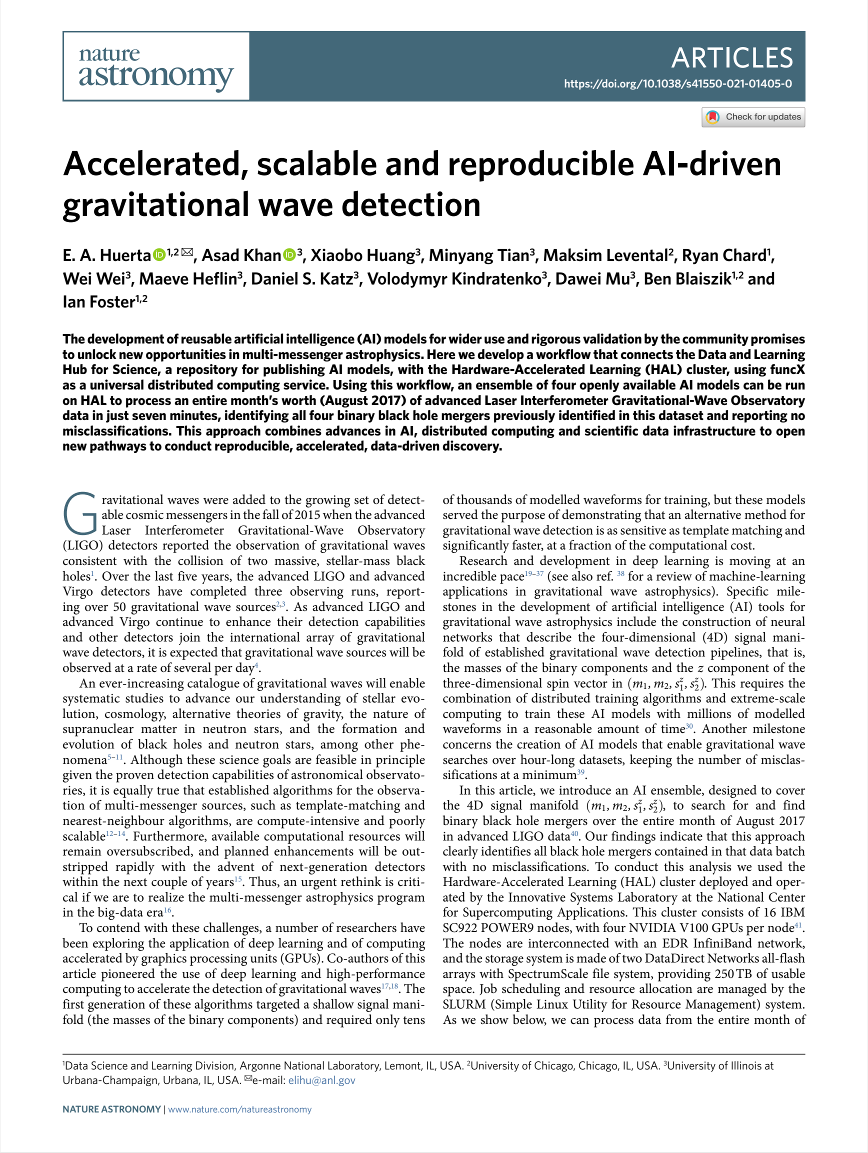
- Candidate search
- Sky localization
- Parameter estimation
- Alert generation and distribution

The mandatory slide with boxes and arrows



The mandatory slide with boxes and arrows





nature astronomy

ARTICLES

https://doi.org/10.1038/s41550-021-01405-0

Check for updates

Accelerated, scalable and reproducible AI-driven gravitational wave detection

E. A. Huerta[1,2]✉, Asad Khan[3], Xiaobo Huang[3], Minyang Tian[3], Maksim Levental[2], Ryan Chard[1], Wei Wei[3], Maeve Heflin[3], Daniel S. Katz[3], Volodymyr Kindratenko[3], Dawei Mu[3], Ben Blaiszik[1,2] and Ian Foster[1,2]

The development of reusable artificial intelligence (AI) models for wider use and rigorous validation by the community promises to unlock new opportunities in multi-messenger astrophysics. Here we develop a workflow that connects the Data and Learning Hub for Science, a repository for publishing AI models, with the Hardware-Accelerated Learning (HAL) cluster, using funcX as a universal distributed computing service. Using this workflow, an ensemble of four openly available AI models can be run on HAL to process an entire month's worth (August 2017) of advanced Laser Interferometer Gravitational-Wave Observatory data in just seven minutes, identifying all four binary black hole mergers previously identified in this dataset and reporting no misclassifications. This approach combines advances in AI, distributed computing and scientific data infrastructure to open new pathways to conduct reproducible, accelerated, data-driven discovery.

Gravitational waves were added to the growing set of detectable cosmic messengers in the fall of 2015 when the advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors reported the observation of gravitational waves consistent with the collision of two massive, stellar-mass black holes[1]. Over the last five years, the advanced LIGO and advanced Virgo detectors have completed three observing runs, reporting over 50 gravitational wave sources[2,3]. As advanced LIGO and advanced Virgo continue to enhance their detection capabilities and other detectors join the international array of gravitational wave detectors, it is expected that gravitational wave sources will be observed at a rate of several per day[4].

An ever-increasing catalogue of gravitational waves will enable systematic studies to advance our understanding of stellar evolution, cosmology, alternative theories of gravity, the nature of supranuclear matter in neutron stars, and the formation and evolution of black holes and neutron stars, among other phenomena[5–11]. Although these science goals are feasible in principle given the proven detection capabilities of astronomical observatories, it is equally true that established algorithms for the observation of multi-messenger sources, such as template-matching and nearest-neighbour algorithms, are compute-intensive and poorly scalable[12–14]. Furthermore, available computational resources will remain oversubscribed, and planned enhancements will be outstripped rapidly with the advent of next-generation detectors within the next couple of years[15]. Thus, an urgent rethink is critical if we are to realize the multi-messenger astrophysics program in the big-data era[16].

To contend with these challenges, a number of researchers have been exploring the application of deep learning and of computing accelerated by graphics processing units (GPUs). Co-authors of this article pioneered the use of deep learning and high-performance computing to accelerate the detection of gravitational waves[17,18]. The first generation of these algorithms targeted a shallow signal manifold (the masses of the binary components) and required only tens of thousands of modelled waveforms for training, but these models served the purpose of demonstrating that an alternative method for gravitational wave detection is as sensitive as template matching and significantly faster, at a fraction of the computational cost.

Research and development in deep learning is moving at an incredible pace[19–37] (see also ref. [38] for a review of machine-learning applications in gravitational wave astrophysics). Specific milestones in the development of artificial intelligence (AI) tools for gravitational wave astrophysics include the construction of neural networks that describe the four-dimensional (4D) signal manifold of established gravitational wave detection pipelines, that is, the masses of the binary components and the z component of the three-dimensional spin vector in (m_1, m_2, s_1^z, s_2^z). This requires the combination of distributed training algorithms and extreme-scale computing to train these AI models with millions of modelled waveforms in a reasonable amount of time[30]. Another milestone concerns the creation of AI models that enable gravitational wave searches over hour-long datasets, keeping the number of misclassifications at a minimum[39].

In this article, we introduce an AI ensemble, designed to cover the 4D signal manifold (m_1, m_2, s_1^z, s_2^z), to search for and find binary black hole mergers over the entire month of August 2017 in advanced LIGO data[40]. Our findings indicate that this approach clearly identifies all black hole mergers contained in that data batch with no misclassifications. To conduct this analysis we used the Hardware-Accelerated Learning (HAL) cluster deployed and operated by the Innovative Systems Laboratory at the National Center for Supercomputing Applications. This cluster consists of 16 IBM SC922 POWER9 nodes, with four NVIDIA V100 GPUs per node[41]. The nodes are interconnected with an EDR InfiniBand network, and the storage system is made of two DataDirect Networks all-flash arrays with SpectrumScale file system, providing 250 TB of usable space. Job scheduling and resource allocation are managed by the SLURM (Simple Linux Utility for Resource Management) system. As we show below, we can process data from the entire month of

[1]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL, USA. [2]University of Chicago, Chicago, IL, USA. [3]University of Illinois at Urbana-Champaign, Urbana, IL, USA. ✉e-mail: elihu@anl.gov

NATURE ASTRONOMY | www.nature.com/natureastronomy

ML is not yet a mainstream “tool of the trade”, but a huge lot of R&D is already ongoing

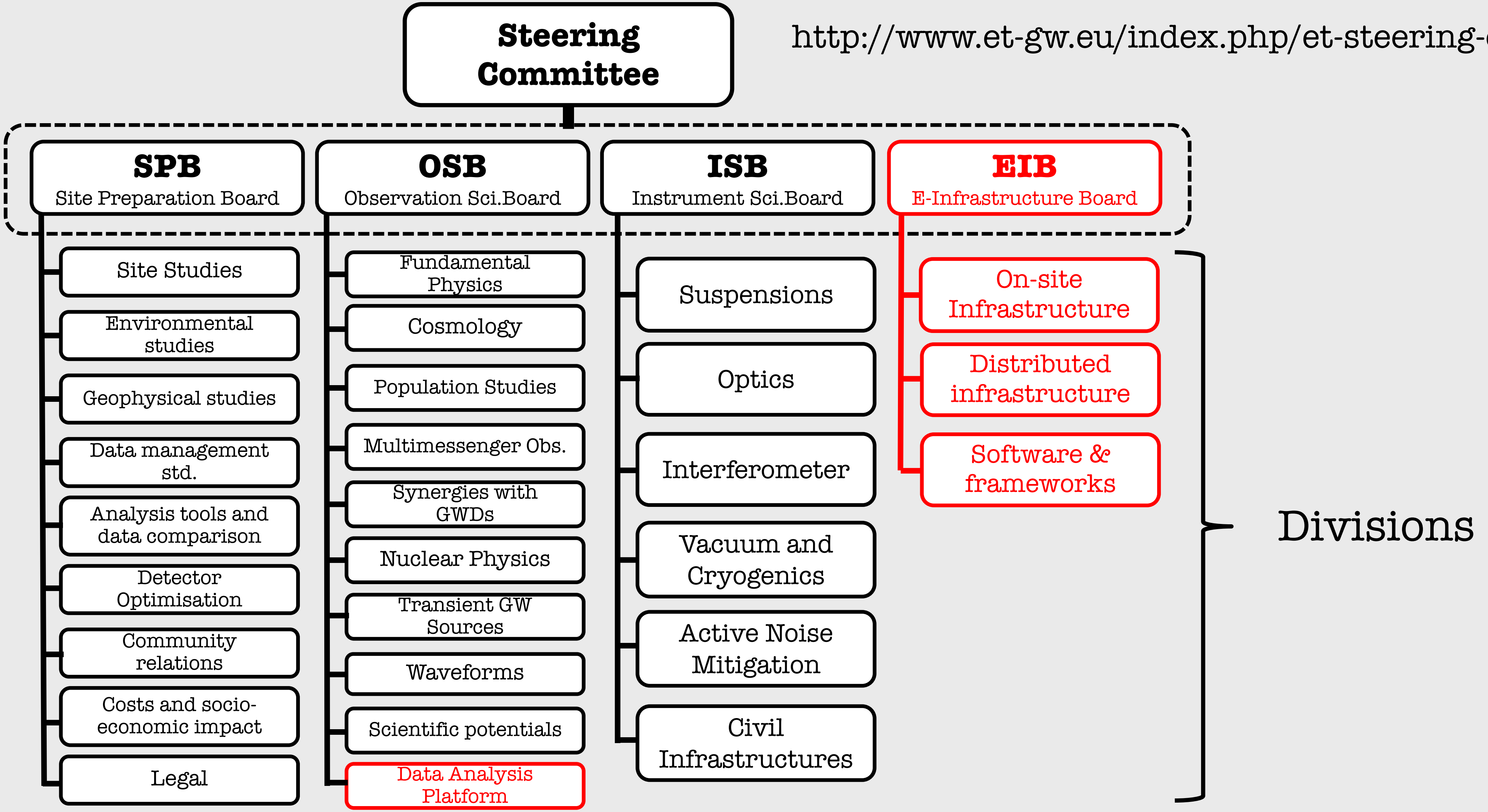
- Efficiency & speed
  - Signal Classification
  - Parameter estimation
  - Noise glitch hunting
  - (Template bank generation)
- Technology exploitation
  - Use advanced hardware (GPU, TPU…)
  - FPGAs / custom hardware
- Automatization
  - Automatize standard procedure for Data Quality
  - Automated de-noising with synthetic noise from GANs?

The ecosystem

- First and foremost, other 3G facilities
  - CE (and LISA!)
  - Will there be an equivalent of the IGWN?
- Several EM and astroparticle initiatives coming of age in the same time frame
  - CTA, SKA, KM3Net, Vera Rubin Observatory, Hyperkamiokande…
  - Will there be a MM-specific (virtual) shared infrastructure like the WLCG?
  - How will the 2030's heir to today's NASA GCN work?
  - The architecture of the next LL alert distribution system is being defined now!
  - We need to be involved since the beginning
- The EU is building the European Opens Science Cloud
  - Scientific Computing in the Digital Continuum
  - How concrete will it be in 2035?

ET Proto–collaboration

http://www.et-gw.eu/index.php/et-steering-committee



EIB mandate

«...to design, create and operate an evolving, efficient and functional e-infrastructure environment at a reasonable cost for the collaboration. Initially the focus will be the development of a Computing Model for the ET».

- Prepare a plan of the studies and activities that need to be undertaken for the development of the ET computing.
- Propose a computing model and its updates to the collaboration.
  - Current chairs: S. B. (INFN), Achim Stahl (Uni Aachen), Patrice Verdier (IN2P3)
  - https://apps.et-gw.eu/tds/ql/?c=16044

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,...)

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

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

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,...)

The Computing Model

- The overall architecture of the e-Infrastructure, either as a single integrated system or as a few separate systems (e.g. instrument control and DAQ, low-latency, and offline)
- A documented way of evaluating the required computing power and storage space from the evolving scientific program of the collaboration
- Estimates of the involved costs and growth timelines
- A description of the data flows, with estimates for the needed network performances
- A description of the User Experience and workflows for relevant activities
- A description of the tools chosen or to be developed to provide all the required functionalities (foundation libraries, frameworks, middleware,...)
- Separate "Work Breakdown Structure" and "Implementation Plan" documents

INFRADEV WP8: Computing and data model

Chairs

- **Achim Streit** - Aachen University (Germany)
- **Sergi Girona** – Barcelona Supercomputing Center (Spain)

Objectives

- **Definition of the computing and data model** of the Einstein Telescope, including the definition of the workflow, estimate of the resources.
- **Data Access** – technical guidelines and principles for implementing the data access policies

INFRADEV WP8: Computing and data model

Tasks

T8.1 T0 data center

Conceptual design of the center in close collaboration with the instrument science board. Definition of the services provided by the center, delimitation against services realized with distributed computing.

T8.2 Computing and Data Model

Development of the computing and data model in close cooperation with the instrument science board and observational science board of ET. Definition of the workflow from the instrument to the publication.

T8.3 Resources

Estimate of the computing resources (computing power and data storage), the personnel, and the operational cost required for all aspects of ET computing. The potential for mitigation must be addressed.

T8.4 Data Access Implementation

Guidelines for the data policy compliance, relevant to the data storage, access, process and distribution, on all relevant time scales, respecting the EU policies on open data.

INFRADEV WP8: Computing and data model

PRELIMINARY Milestones + deliverables

	Workshops	WPs		Deliverable	Due date
M8.1	Workflows Requirements collection and constraints: computing and data	WP8			M12
			D8.1	Computing and Data Requirements	M18
M8.2	Computing Infrastructures availability for ET workflows, characteristics	WP8, WP9			M24
M8.3	On site infrastructure, computing and data model	WP8, WP6			M36
M8.4	Low latency and offline workflows and computing model	WP8, WP6			M40
			D8.2	Computing and Data Model	M42
M8.5	Data management, access, policy and implementation	WP8, WP2, WP6			M46
			D8.3	Data Access Implementation Guidelines	M48

ET EINSTEIN TELESCOPE

Under discussion

- **Budget (~1 FTE)**
  - Role: Coordination, system engineer knowledge
  - Risk: No easy to find the right person
  - Mitigation: Split the role among more persons, complement with in-kind contribution.

- **Participants roles**
  - Workshops participation and contribution. The workshops output will be the input for the WP deliverables.
  - Collect possibility to give in-kind contributions (personnel)

A word of warning

- We need special professional profiles
  - Something between physical science and computer engineering
  - Not exactly “pipeline developers”, not exactly “system architects”
- Such personpower is difficult to find
  - Skilled personpower for computing activities is scarce
  - Hard to train and keep, hard to hire
- This is not a problem for the GW community only
  - And neither limited to the EU
- For example, HSF had some recommendations for that
  - Training, career incentives,…
  - We should plan also for that