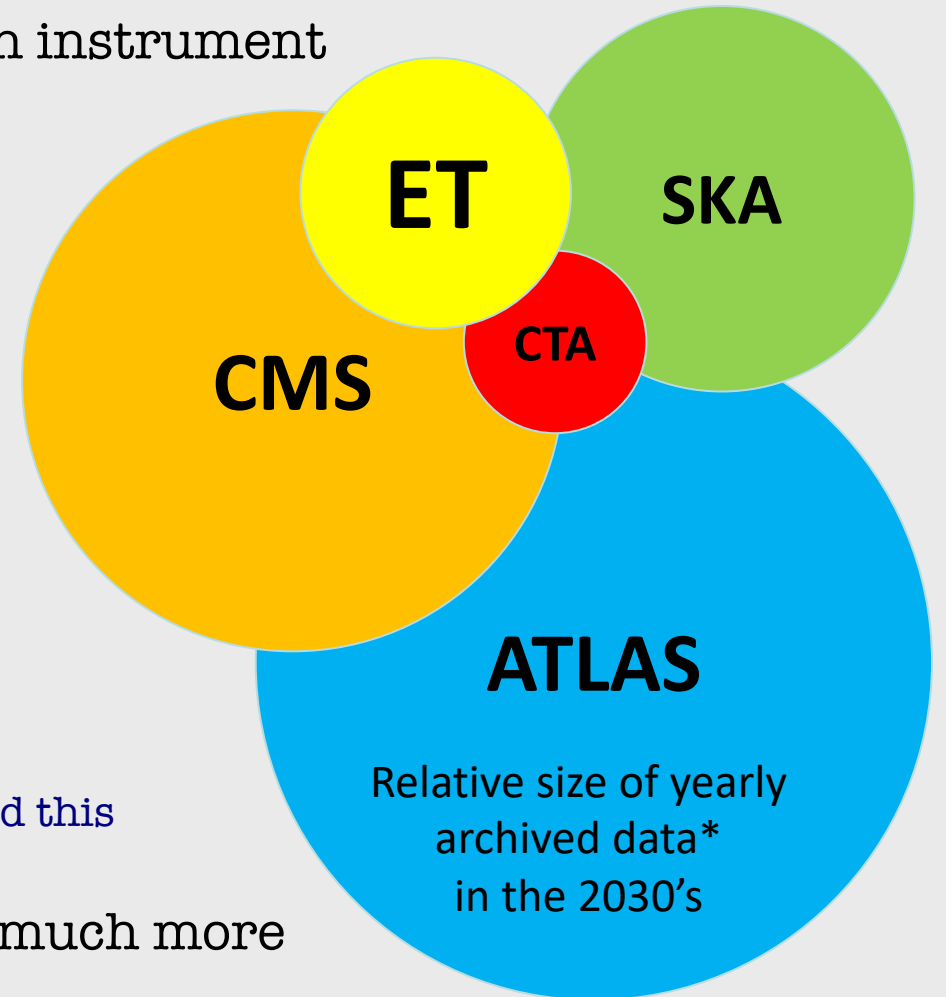


EIB SHORT INTRODUCTION

Stefano Bagnasco, Patrice Verdier,
Nadia Tonello, Achim Stahl

CHALLENGES SUMMARY

- Luckily, raw interferometer data don't grow much with instrument sensitivity
 - We're not exploding like HL-LHC!
 - We expect about few tens of PB of raw data per year
 - No big deal today, piece of cake by 2035
- The amount of scientific information encoded in the data does grow (a lot)
 - And the computing power needed to wring it out
 - It's already a task to estimate the computing power needs!
- Current computing algorithms and strategies won't trivially scale
 - High rates, long signals
 - Need to rethink some of them
 - We need to support and run Mock Data Challenges to understand this
 - Target: keep being $o(10\%)$ of an LHC experiment at any time
- Low-latency alert generation and management will be much more complex
 - Design, planning and implementation already started outside of GW
 - Need to have our say now (see also dedicated session on Wednesday)



*to say nothing of weather forecast, genomics, Earth observation, oil industry, GAFAM and everybody else

THE E-INFRASTRUCTURE BOARD

Chairs: SB, Patrice Verdier (IP2I-IN2P3)

Division 1: Software, frameworks, and data challenge support

- Chair: Andres Tanasijczuk (U. Catholique de Louvain)
- ISB Div10 liaison: John Veitch (U. Glasgow)

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 alert infrastructure

- Chair: Steven Schramm (U. Geneva)

TTG: Technology Tracking Working Group

- Chair: Sara Vallero (INFN Torino)

ET-PP WP8: COMPUTING AND DATA ACCESS

Leaders: Nadia Tonello (BSC), Achim Stahl (U. Aachen)

Task 8.1: T0 data center

- Leader: Patrice Verdier (IP2I-IN2P3)

Task 8.2: Computing and Data Model

- Leader: Anastasios Fragkos (U. Geneva)

Task 8.3: Resources

- Leader: Silvio Pardi (INFN Napoli)

Task 8.4: Data Access Implementation

- Leader: Nadia Tonello (BSC)

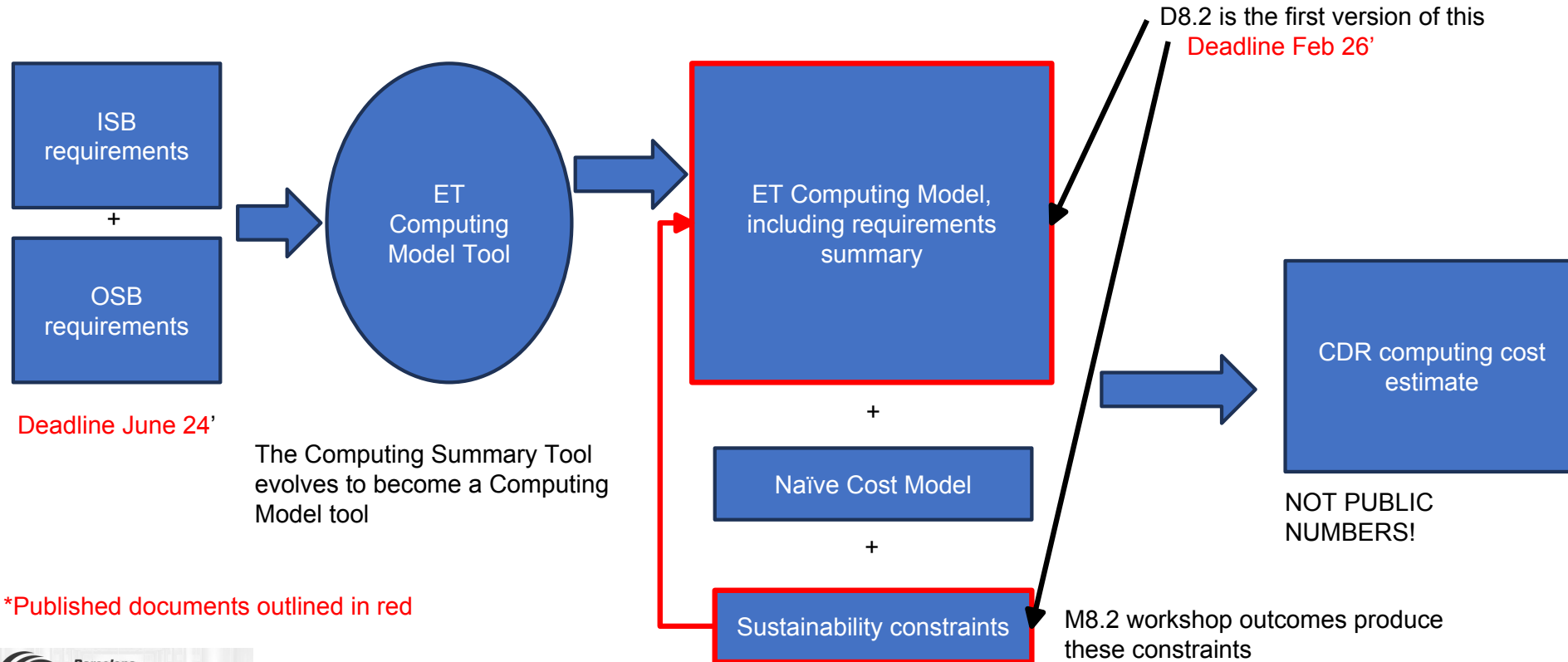
COMPUTING MODEL REQUIREMENTS WORKSHOP

- Organized by ET-PP WP8, hosted by University of Geneva dept. of Astronomy in Versoix, Oct 26th-27th
 - <https://indico.ego-gw.it/event/590/>
- Preparation of the first ET-PP WP8 deliverable
 - First (early) draft available on b2drop
 - Possibly a writing F2F session in Barcelona
- Collecting input from ISB and OSB, and from other stakeholders (WLCG, Virgo, supercomputing centres, ESCAPE, HSF,...)
- Good participation also from outside the EIB
- Very successful also for establishing communication channels
 - See Paul Laycock's presentation tomorrow
- Baseline message: we need work just to estimate the needs
 - Cannot come out with numbers yet

REQUIREMENTS COLLECTION



From computing requirements to computing model



Deadline June 24'

The Computing Summary Tool evolves to become a Computing Model tool

*Published documents outlined in red

D8.2 is the first version of this
Deadline Feb 26'

M8.2 workshop outcomes produce these constraints
Deadline Sep 24'



4 September 2023

8

alba.gonzalvez@bsc.es



MOCK DATA CHALLENGES

- MDC as multipurpose tools
 - Develop and exercise analysis code and strategies
 - Collect data about analyses' requirements and performance
 - Build the data analysis community and bootstrap new groups
 - Educate the community in the use of common distributed computing tools and best practices for sustainability and manageability
 - Iteratively test the distributed computing infrastructure
- Mock Data Challenge support plans
 - MDC1: provide data distribution layer (OSDF: CVMFS + cache) and survey the activities (status: see Tania Regimbaud's talk tomorrow)
 - MDC2: provide (possibly a set of) prototype tools for workload management etc.
 - MDC3-n: iterate

SYNERGIES WITH VIRGO

Bottom line: computing needs grow more or less as a continuum from O4 to O5 to post-O5 to ET, and technologies keep evolving; not something specific to post-O5, but it extends the scope for synergies.

- Distributed computing infrastructure
 - CPU power needs grow continuously with sensitivity (CBC PE)
 - ET already needs a working and evolving computing infrastructure (for MDCs, simulations,...)
- Low-latency alert distribution network
 - High rates imply high automation, long signals imply new features (e.g., continuous alert updates)
 - In the coming years, the developments may be driven by running experiments (e.g., Vera Rubin Observatory): the GW community already needs to be present
- Carbon-efficient computing
 - And, in general, technology tracking: heterogeneous computing, efficient algorithms, ML,...
 - Same message: development is a continuum

THE WIDER PHYSICS COMPUTING COMMUNITY

- The IGWN effort
 - International Gravitational-Wave observatories' Network, i.e., the LIGO-Virgo-KAGRA common computing infrastructure
 - Architecture, deployment and coordination/management effort
 - Already existing, GW community in Europe slowly getting familiar with using it
- The WLCG Collaboration
 - Runs the LHC Computing Grid, the largest in Europe (not sure about the world)
 - Virgo is already “Observer experiment”
- The ESCAPE Collaboration
 - See next slides (and Giovanni Lamanna’s talk tomorrow)

- Started as a large EU project
 - Development of “next generation” tools for computing and open science...
 - ...serving the HEP, Astrophysics and Astroparticle ESFRI clusters
- Now a collaboration between ESFRIs
 - ET MoU for participation signed in July...
 - ...so we are represented in the Strategy and Technical boards
 - For the time being by the Collaboration, pending the creation of the legal entity
- Will also fund computing R&D projects through the OSCARS program
 - Competitive selection by a committee
 - Small projects (1-2 years, 100-200 KEUR)



The new ESCAPE Collaboration work programme

ESCAPE CC

Operating the community-based “Competence Center” for EOSC-alignment, train and support, extended outreach, financial model for services and networking with other SCL-CCs

ESCAPE EVSI

R&I for an “European Virtual Institute for Research Software” for advanced technologies



ESCAPE DIOS

Data Infrastructure for Open Science

Access physical & e-infrastructures
Processing & Analysis
Security & Operations



ESCAPE OSSR

Open-source Scientific Software and Service Repository

Aggregator & Integrators
Sharing and Discover
Training & Support



ESCAPE ESAP

ESFRI Science Analysis Platform

Processing & Analysis
Sharing and Discover
Training & Support



ESCAPE CS

Citizen Science

Sharing and Discover



ESCAPE VO

Virtual Observatory

Processing & Analysis
Sharing and Discover
Training & Support



Entities

VRE services

Programmes

ESCAPE COSO

Challenging “Open Science Objectives” by RI commitments in Open Science Projects (OSP) as well as Cross-Cluster Open Science Projects (COSP)

ESCAPE TECH

Bring the FAIRness within technology, R&D and innovation projects as well as explore new “close-to-sensors” low-latency open-data science

ESCAPE CARS

Career development and rewarding for researcher committing in Open Science. Planning, tracking, and assessing scientific knowledge production

ESCAPE SDSS

Building synergies on “Sector Data Spaces” for Society: Green deal, Health, Manufacturing, Education and Skills

PUTTING ALL TOGETHER

- Use IGWN infrastructure as baseline
 - IGWN uses the European computing centres as an extension of the OSG (which is nonoptimal...)
 - However, the functionality is there (OSDF + HTCondor)
- Use ESCAPE as the first toolbox
 - First the “Data Lake” (DIOS), then the Virtual Research Environment
 - Also, Virtual Observatory (INFN Perugia)
- Develop a common initial R&D program
 - Data Lake (Rucio) for data distribution
 - VRE for data access and job management
 - ...

...SO, FOR EXAMPLE:

- The current IGWN data distribution infrastructure is based on OSG-managed Open Science Data Federation
 - Essentially extending the US system to European centres
 - Less than optimal: little experience with the tools, time zone delay in getting support, disuniformity with HEP experiments,...
 - The next generation WLCG data distribution infrastructure will be based on the ESCAPE Data Lake
 - Prototype deployment available
- R&D activity to provide the same functionality using the ESCAPE Data Lake
 - Synergy between EIB Div1, ETIC and Virgo
 - Target is using the same shared infrastructures (as we promised in the ESFRI proposal!)
 - Some MDC data already copied to the DL, studies ongoing
 - Also, dedicated deployment under study

- “Technologies for MultiMessenger Astronomy” proposal for INFRA-TECH call that was not funded in 2022
- Partners are ET+Virgo, CTA and KM3Net
- WP6 is “” (WP leader Steven Schramm)
 - T6.1 “Fast data processing and edge computing”
 - T6.2 “Sustainable large-scale computing”
 - T6.3 “Multimessenger alert tools”
- Several important topics
 - E.g., CC involvement, multimessenger alert system development, ML technologies,...

- Collaboration support:
 - Mailing list server migrated to Sympa (currently consolidating all mailing lists in same server)
 - ETMD now source for et-all mailing list
 - AAI development still stuck (but working on that)
- Technology Tracking Working Group:
 - re-starting computing Technology Tracking seminars
 - Aimed at the whole collaboration (announcements will be distributed)
- Wide interest in sustainability:
 - Computing will likely be the most energy-greedy activity
 - Will we be limited not by available computing resources but by acceptable carbon footprint?
- Still need to attract people!
 - Building a real computing team will be a sustainability requirement

- Development of Computing Model
 - Requirements, architecture, resources,...
 - Computing TDR, PBS, WBS as tools to implement the Computing Model
- Focusing on Mock Data Challenges
 - As a tool for developing analysis techniques, gathering requirements and exercising infrastructure
 - Also, start promoting good practices for sustainability and manageability
- Liaising with the wider physics computing community
 - ESCAPE, IGWN, WLCG,...
- Low-latency alerts also a core item
 - Need not to be left behind