E-INFRASTRUCTURE BOARD

Stefano Bagnasco, INFN

for the ET-EIB

ET Monthly Meeting | Feb 23, 2024

EIB Chairs: Stefano Bagnasco (INFN), Patrice Verdier (IP2I Lyon - IN2P3)

ET-PP WP8 leaders: Achim Stahl (U. Aachen), Nadia Tonello (BSC)

Division 1: Software, frameworks, and data challenge support
Andres Tanasijczuk (UC Louvain)

Division 2: Services and Collaboration Support
Antonella Bozzi (EGO)

Division 3: Computing and data model, Resource Estimation
Gonzalo Merino (PIC)

Division 4: Multimessenger alerts infrastructure
Steven Schramm (Université de Genève)

TTG: Technology Tracking working Group
Sara Vallero (INFN Torino)

Task 8.1: TO data center
Leader: Patrice Verdier (IP2I-IN2P3)

Task 8.2: Computing and Data Model
Leader: Anastasios Fragkos (Geneva)

Task 8.3: Resources
Leader: Silvio Pardi (INFN Napoli)

Task 8.4: Data Access Implementation
Leader: Nadia Tonello (BSC)

Liaison with OSB Div. 10: John Veitch (University of Glasgow), Elena Cuoco (EGO)

Joint ET-PP WP8 & ETC-EIB management (e.g., weekly call for coordination)

EIB Resp: Verdier Bagnasco 👤 46 (FRTE: 11.985)

○ DIV 01 - Software, frameworks and data-challenge support Resp: Tanasijczuk 👤 5 (FRTE: 0.8)

○ DIV 02 - Services and collaboration support Resp: Bozzi 👤 3 (FRTE: 0.485)

○ DIV 03 - Computing and data model, resource estimation Resp: Merino 👤 2 (FRTE: 0.1)

○ DIV 04 - Multimessenger-alert infrastructure Resp: Schramm 👤 5 (FRTE: 0.55)

○ TTG - Technology-tracking working group Resp: Vallero 👤 5 (FRTE: 1.4)
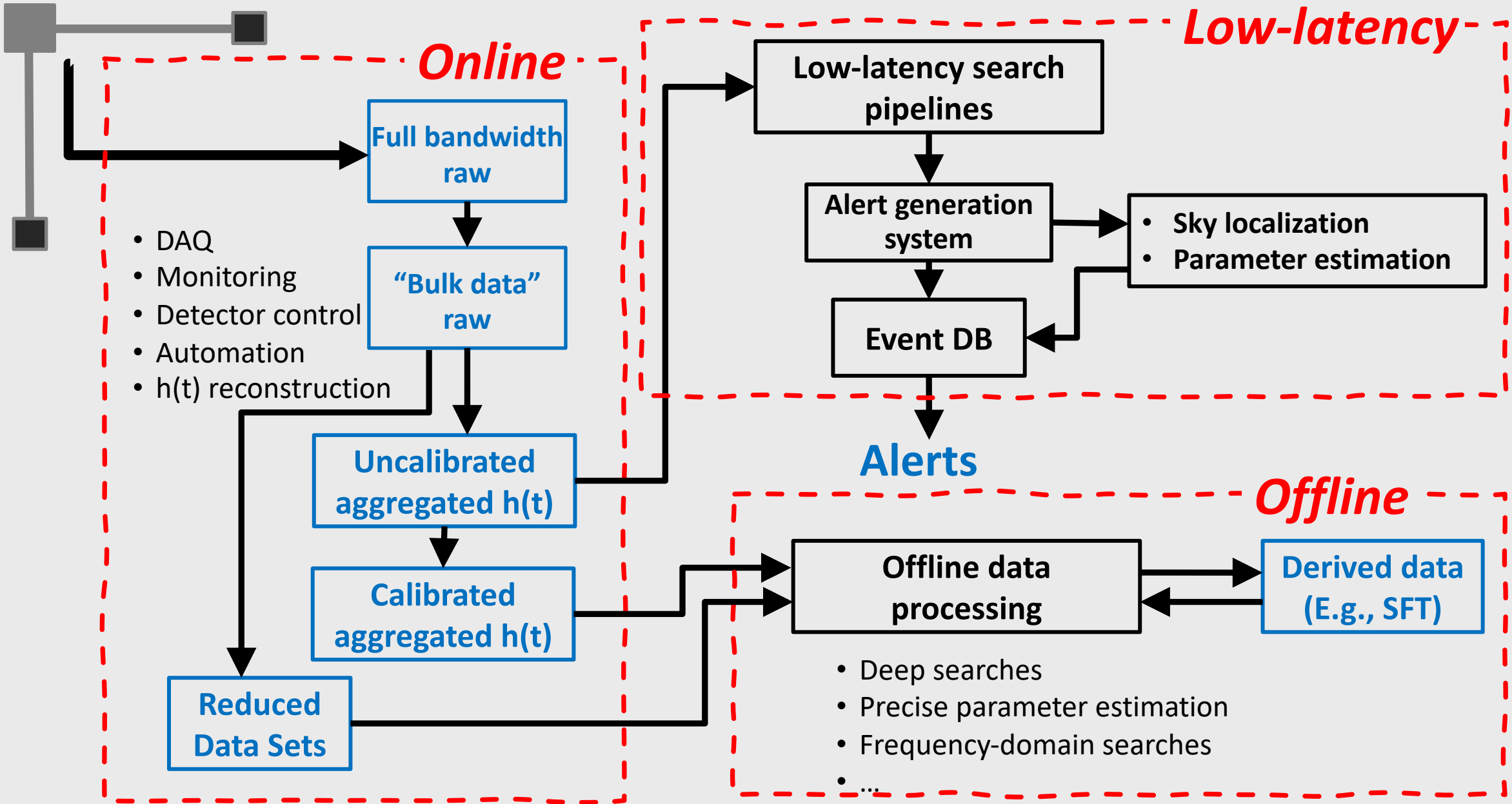
Plan: use MDCs to evaluate community expectations and test prototypes

- First MDC: come as you are, we just provide data distribution (through IGWN)
- Second MDC: provide more tools and collect feedback
- Iterate

- Discussions at the Maastricht symposium
  - But not much really happened since then… will start soon!

- Independent packaged parts of the final architecture
  - Providing limited functionalities, possibly some as mere demonstrators
  - But actually to be released to users (i.e., they MUST be functional)
  - Different implementations may exist, with different tools/technologies used to provide same functionality
  - Integration of existing tools, no excessive bespoke developments, to map "kits" onto smallish projects
- Examples:
  - ESCAPE Datalake + RucioFS for data distribution
  - IAM-based AAI
  - ESCAPE Datalake + VRE interactive data analysis
  - OSDF + INFNCloud interactive data analysis
  - "Packaged" and quality-tested MDC data generation tool
  - Paul's rich metadata tool

HORIZON EUROPE

The OSCARS – *O*pen *S*cience *C*lusters' *A*ction for *R*esearch & *S*ociety proposal was submitted in March in response to the call HORIZON-INFRA-2023-EOSC-01-01 for a budget of EUR 25.00 million.

Two activities:
1) Consolidation of the Science Cluster approaches between the RI communities, where ESCAPE focuses on Data Lake, VRE and training. As part of this activity, we were also asked to study and implement an ESCAPE "community-based competence centre".
2) Demonstrate and pilot the use of EOSC resources by multiple research communities through cross-RI and/or cross-domain open science projects and services, by running 'open calls with cascading grants' with a budget of approximately EUR 16.00 million.

OSCARS is based on the co-participation of the five science cluster collaborations in Europe ENVRI-FAIR (Environment and Earth Sciences), EOSC-LIFE (Biomedical Sciences), ESCAPE, PANOSC (Light and Neutron Facilities), SSHOC (Social Sciences and Humanities).

5   Giovanni Lamanna   07/07/2023

- ETAP (Université de Genève)
  - Access to multiple ESCAPE Data Lakes.
  - Rich metadata service integration
  - Access to multiple rich metadata instances
  - A lightweight CRM service monitoring the VRE

- MADDEN (INFN-TO & Université Catholique de Louvain)
  - Multi-RI Data Lake managed with Rucio, simulating data exchange with CE
  - Development and test of RucioFS (POSIX FUSE mount of the Rucio catalogue)
  - Extend RucioFS to support advanced metadata
  - Strong synergy with Virgo O5

- Second OSCARS call (November?)
  - Streaming data for LL?
  - Something IVOA-related?

SUBMITTED!

- System based on widely-adopted tools and standards, plus our own flavour

  - INDIGO IAM as core component

  - ETMD as authoritative membership source

  - SAML and OIDC services as identity providers (home institutions, ORCID, EGO Active Directory,…)

- Michel Jouvin (IJCLab) leading the integration work

  - Involving ETMD (Gary Hemming), IAM developers (Francesco Giacomini et al.), CNAF support

- Next step: deploy!

  - CNAF accepted to host the service

  - Exact agreement details being defined (who does what?)

- CNAF will take responsibility for deploying the initial instance and later to upgrade it from one version to another one
- ET (persons to be identified, MJ +...) will take the responsibility to manage the VO (IAM instance configuration and usage)
  - Will not require access to the instance other than through the INDIGO IAM API/dashboard (i.e, no direct access to hardware).
- The vision/plan is that ET IAM accounts will be created/managed through ETMD, the ET member database, in a way similar to what is done for LHC experiments.
  - Users will not be able to apply for an account if they are not in ETMD.
  - During the initial phase, until the synchronisation scripts (based on INDIGO IAM API, not on direct access to the database) are put in place, some accounts for test/early adopter users will be created through the IAM dashboard as an exception but it will be a limited number of users.
- As ET is in an early phase we don't have strong requirements for 24x7 availability
  - Can live with the usual best-effort approach used by CNAF for these kind of services?

Michel.Jouvin@ijclab.in2p3.fr

- ETIC TT platform being built in Torino ("TechZoo")
  - Heterogeneous and expandable HPC platform
  - Interoperable with the TeRABIT "HPC Bubble"
  - Access layer via INFNCLOUD, common with similar facility at INFN-BO
  - Usable for code porting, testing, special architectures, accelerators evaluation etc.
  - ...and for regular computing (e.g., numerical relativity)

- Hardware being configured, possibly more coming
  - Very first timid activities starting (e.g., porting RIFT code to nVidia GH)

- Expect a call for ideas/applications in summer

What is SPECTRUM?

A project granted under the call HORIZON-INFRA-2023-DEV-01-05, which aims to prepare a Computing Strategy for Data-intensive Science Infrastructures in Europe for the High Energy Physics (HEP) and Radio Astronomy (RA) domains

Expected outcomes

The realisation of a **Community of Practice (SPECTRUMCoP)** to gather and inform about future directions and needs in data-intensive research on the one side, and about future e-infrastructures on the other

A **Strategic Research, Innovation and Deployment Agenda (SRIDA)** and a Technical Blueprint about agreed processing models and solutions, to provide feedback on investment to funding agencies and policy makers

Who is part of SPECTRUM?

SPECTRUM gathers selected stakeholders in the HEP and RA research domains, and at the same time experts from the e-infrastructures (HPC, Clouds, Quantum Computing). The former group brings **directions and future needs**, the latter bring **expectations** for new e-infrastructures about technical and policy aspects.

Why is SPECTRUM different from previous attempts?

Previous interactions between the research and e-infrastructure communities have been **a posteriori**, attempting to adapt scientific workflows to already operational facilities. This has been only partially successful due to technical (non-compliant system architectures, …) and policy (user access, …) incompatibilities.

SPECTRUM wants to move the handshaking process **a priori**, before the e-infrastructures are designed and deployed

Tommaso.Boccali@pi.infn.it

JENA - Joint ECFA, NUPECC, APPEC Activities

The JENA Symposium in May 2022 in Madrid (https://indico.cern.ch/event/1040535/) revealed an increased need for discussions on the strategy of EU federated computing at future large-scale research facilities.

Focused workshop on the strategy of computing in Bologna June 2023 aimed to define computing requirements in the next decade and to try and find synergies.

Outcome: creation of 5 WGs to generate input (whitepapers) for JENA Symposium in 2025:

- WG1: HTC, WLCG and HPC
- WG2: Software and Heterogeneous Architectures
- WG3: Federate Data Management, Virtual Research Environments and FAIR/Open Data
- WG4: Machine Learning and Artificial Intelligence
- WG5: Training, Dissemination, Education

JENAA — Joint ECFA-NuPECC-APPEC Activities

12

SPECTRUM project

- WG1: Data Management and Access
- WG2: Workflow management and organization
- WG3: Compute Environment
- WG4: SW Tools
- WG5: Scientific Use cases
- WG6: Facilities

JENA Computing Initiative

- WG1: HTC, WLCG and HPC
- WG2: Software and Heterogeneous Architectures
- WG3: Federate Data Management, Virtual Research Environments and FAIR/Open Data
- WG4: Machine Learning and Artificial Intelligence
- WG5: Training, Dissemination, Education

ET-PP WP8 Computing and Data models

Workshops

		WPs	Due date	
M8.1	Workflows Requirements collection and constraints: computing and data		Sep 2023	Oct 2023
M8.2	Computing Infrastructures availability for ET workflows, characteristics	WP9	Aug 2024	July 2024
M8.3	On site infrastructure, computing and data model	WP6	Aug 2025	
M8.4	Low latency and offline workflows and computing model	WP6	Dec 2025	
M8.5	Data management, access, policy and implementation	WP2, WP6	July 2026	

WP8 milestones

Deliverables

		Lead	Due date	
D8.1	Computing and Data Requirements	UniGe	Feb 2024	Feb 2024
D8.2	Computing and Data Model	UniGe	Feb 2026	
D8.3	Data Access Implementation Guidelines	IFAE	July 2026	

WP8 Documents

ET Symposium - May 2024

2

Nadia.Tonello@bsc.es