

**EINSTEIN
TELESCOPE**

ET-PP Annual Meet 22-24 July 2025

WP8 Computing and data model
Parallel Session: ETO-WP6-WP8
Computing and Data Access Policy
& ET DMP (D6.6, D8.1, D8.2,D8.3)

Nadia Tonello (BSC), Achim Stahl
(DESY), Michele Punturo (INFN)

23th July 2025



EINSTEIN
TELESCOPE

EiB-WP8 organization Chairs: Stefano Bagnasco (INFN Torino), Patrice Verdier (IP2I-IN2P3)

ET-PP WP8 Task 8.1

T0 data center

Patrice Verdier (IP2I-IN2P3)

ET-PP

Task 8.2

Computing and Data Model

Paul Laycock (Geneva)

Task 8.3

Resources

Silvio Pardi (INFN Napoli)

Task 8.4

Data Access Implementation

Nadia Tonello, Oscar Reina (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)

ET-EiB



EiB - ETPP WP8, our 2025 in a nutshell

EINSTEIN
TELESCOPE

- Feb 18-20th - Einstein Telescope data analysis workshop
<https://agenda.infn.it/event/44693/overview>
- **March 5th - D8.1 Computing and Data Reqs. Resubmission**
- **ET-PP Midterm review, May 15th 2025**
- May 21&22 - Pre-Merger & Low Latency workshop
<https://indico.ego-gw.it/event/852/overview>
- May 26-30 - XV Einstein Telescope Symposium in Bologna
<https://indico.ego-gw.it/event/819/>
- Workshop on Online Infrastructure, Computing and Data Model
30th June, 1st July, Lyon, France <https://indico.ego-gw.it/event/820/>
- **ET-PP RP2 Midterm review REPORT, 3rd July 2025**
- **ET-PP Annual Meeting**
22-24 July 2025 Barcelona. <https://indico.ifae.es/event/2183/overview>



ET-PP WP8

Computing and data model



EINSTEIN
TELESCOPE

23th July 2025



EINSTEIN
TELESCOPE

ET-PP WP8 Milestones & Deliverables

ET-PP WP8

Computing and data model

WP8 milestones	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 and sustainability	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	

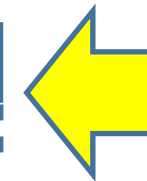
WP8 Deliverables			Lead	Due date	
	D8.1	Computing and Data Requirements	UniGe	Feb 2024	Feb 2024
	D8.2	Computing and Data Model	UniGe	Feb 2026	
	D8.3	Data Access Implementation Guidelines	IFAE	July 2026	

EC RP1 received

Deliverable 8.1 (Computing and Data requirements)
Changes requested
(submitted 05-03-2025 [1])

EC RP2 received [2]
Further review required

Deliverable 8.1
IN REWRITING
Deliverable 8.2
IN PROGRESS



[1] Available at TDS at <https://apps.et-gw.eu/tds/?r=19444>

[2] WP8 report and review materials available at <https://indico.ifae.es/event/2128/>



EINSTEIN
TELESCOPE

ET-PP WP8 D8.2 Computing and Data Model.

Computing and data model

D8.2 preparation: Timeline (suggested in ET Symposium WG)

- **July 22nd ET-PP** - we have a full template, no missing pieces, editors assigned
 - **Mid-September** we have an editors writing meeting
 - **End of October** we release a draft to external reviewers
- They have 2 weeks

Release to ET collaboration by the end of the year

- **Editor assignment & Progress tracking (B2DROP)**
- Collaborative document edition set up (Overleaf-CERN)
- TBD at Lyon Workshop on Onsite Infrastructure, Computing and Data Model. <https://indico.ego-gw.it/event/820/>

AFTER D8.1 RE SUBMISSION

UID	LEVEL	RUNNING TITLE	UID	TITLE	COMMENTS	EDITOR	REVIEWER	STATUS
0	0	ET Computing Model	0	ET Computing Model		?	?	
0.1	1	Scope of the Computing Model	0.1	Scope of the Computing Model		AS	SB	
1	1	Part I: Introduction	1	Part I: Introduction		?	?	
1.1	2	Science Drivers	1.1	Science Drivers		AS	PL	STARTED
1.2	2	Core Mission	1.2	Core Mission		PL	AS	
1.3	2	Collaboration Science Goals	1.3	Collaboration Science Goals		?	?	
2	2	Part II: Core Mission	2	Part II: Core Mission		?	?	
2.1	3	Computing Model Overview	2.1	Computing Model Overview		SB	?	
2.1.1	4	Onsite Computing	2.1.1	Onsite Computing		PV	FC	
2.1.2	4	Offsite Computing	2.1.2	Offsite Computing	Includes Low Latency / that's not obvious to	PL	SB	
2.2	3	Core Services	2.2	Core Services		?	?	
2.2.1	4	Data Format	2.2.1	Data Format		?	?	
2.2.1.1	5	FAIR principles	2.2.1.1	FAIR principles		OR	NT	
2.2.1.2	5	GW detector data format	2.2.1.2	GW detector data format		XY	NT	
2.2.1.3	5	GW event/alert format	2.2.1.3	GW event/alert format		SS	OR	
2.2.2	4	Databases	2.2.2	Databases		?	?	
2.2.2.1	5	EventDB	2.2.2.1	EventDB		DM	PL	
2.2.2.2	5	Event metadata capture	2.2.2.2	Event metadata capture		?	NT	
2.2.2.3	5	Conditions metadata	2.2.2.3	Conditions metadata		PV	PL	
2.2.3	4	Data management	2.2.3	Data management		PL	NT	
2.2.3.1	5	Data archival	2.2.3.1	Data archival		GM	?	
2.2.3.2	5	Low latency data distribution	2.2.3.2	Low latency data distribution		SV	?	
2.2.3.3	5	Distributed data management	2.2.3.3	Distributed data management		FL	?	
2.2.3.4	5	Data for Open Science	2.2.3.4	Data for Open Science		NT	?	
2.2.4	4	Workload management	2.2.4	Workload management		SV	?	
2.2.4.1	5	Workload requirements definition	2.2.4.1	Workload requirements definition		SV	?	
2.2.4.2	5	Monitoring and accounting	2.2.4.2	Monitoring and accounting		GM	DM	
2.2.5	4	Software management	2.2.5	Software management		DM	?	
2.2.5.1	5	Software framework	2.2.5.1	Software framework		SB	AT	
2.2.5.2	5	End-to-end testing and Q/CD	2.2.5.2	End-to-end testing and Q/CD		PL	SV	
2.2.5.3	5	Release management	2.2.5.3	Release management		DM	?	
2.2.5.4	5	ML training	2.2.5.4	ML training		JV	?	
2.2.6	4	Identity and Access Management	2.2.6	Identity and Access Management		MJ	GH	
2.2.7	3	Network	2.2.7	Network		XY	?	
2.3	3	Onsite Computing	2.3	Onsite Computing		PV	FC	
2.3.1	4	Detector controls	2.3.1	Detector controls	Including refinement interface	?	?	
2.3.2	4	DAQ	2.3.2	DAQ		?	?	
2.3.3	4	Calibration	2.3.3	Calibration		?	?	
2.3.4	4	Detector characterization	2.3.4	Detector characterization		NA	?	
2.4	3	GW Events	2.4	GW Events		JV	?	
2.4.1	4	Search pipelines	2.4.1	Search pipelines		JV	?	
2.4.2	4	Parameter estimation	2.4.2	Parameter estimation		VB	?	
2.4.3	4	Alert system	2.4.3	Alert system		DM	?	
2.4.3.1	5	Alert generation	2.4.3.1	Alert generation		DM	?	
2.4.3.2	5	Event generation	2.4.3.2	Event generation		DM	?	
2.4.3.3	5	Event enrichment	2.4.3.3	Event enrichment		DM	?	
2.4.3.4	5	External alert handling	2.4.3.4	External alert handling		SS	?	
2.5	3	Resource Estimates	2.5	Resource Estimates		GM	SB/PV	
2.5.1	4	Computing and data resource estimates	2.5.1	Computing and data resource estimates		GM	SB/PV	
2.5.2	4	Computing and software personnel estimates	2.5.2	Computing and software personnel estimates		DM	PL	
3	3	Part III: Analysis	3	Part III: Analysis		?	?	
3.1	4	Collaboration	3.1	Collaboration		?	?	
3.1.1	5	Training and documentation	3.1.1	Training and documentation		?	?	
3.1.2	5	User support	3.1.2	User support		?	?	
3.1.3	5	Collaborative services	3.1.3	Collaborative services		GH	?	
3.1.4	5	Interface to ET0-IT	3.1.4	Interface to ET0-IT		PL	ETO	
3.2	4	Open Science	3.2	Open Science		OR	?	
3.2.1	5	Open Science support	3.2.1	Open Science support		?	?	
3.2.1.1	6	Data Management	3.2.1.1	Data Management		NT	?	
3.2.1.2	6	FAIR software for GW science	3.2.1.2	FAIR software for GW science		?	?	
3.2.1.3	6	User support	3.2.1.3	User support		?	?	
3.2.2	5	Resource Estimates	3.2.2	Resource Estimates		?	?	
3.2.2.1	6	Computing and data estimates	3.2.2.1	Computing and data estimates		?	?	
3.2.2.2	6	Personnel estimates	3.2.2.2	Personnel estimates		?	?	
4	4	Part IV: Strategic Considerations	4	Part IV: Strategic Considerations		?	?	
4.1	5	Sustainability	4.1	Sustainability		?	?	
4.1.1	6	Training the new generation	4.1.1	Training the new generation		?	?	
4.1.2	6	Archiving the new generation	4.1.2	Archiving the new generation		?	?	
4.1.3	6	Tracking the new generation	4.1.3	Tracking the new generation		?	?	
4.1.4	6	Monitoring the new generation	4.1.4	Monitoring the new generation		?	?	
4.1.5	6	Reporting the new generation	4.1.5	Reporting the new generation		?	?	
4.1.6	6	Summarizing the new generation	4.1.6	Summarizing the new generation		?	?	
5	5	Part V: Executive Summary	5	Part V: Executive Summary		?	?	
5.1	6	Summary tables	5.1	Summary tables		?	?	



EINSTEIN
TELESCOPE

ET-PP WP8 D8.2 Computing and Data Model.

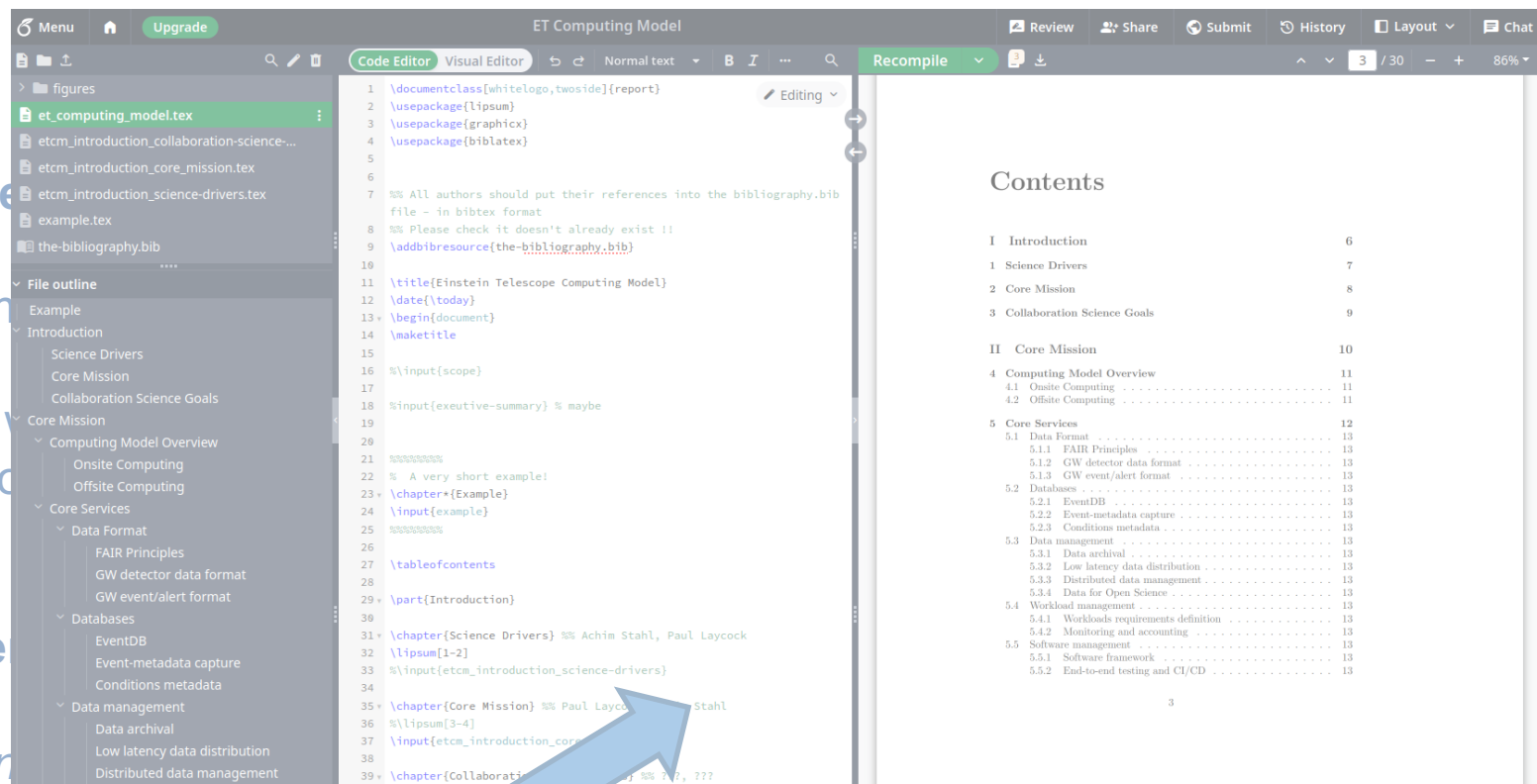
Computing and data model

D8.2 preparation: Timeline (suggested)

- **July 22nd ET-PP** - we have a full term of editors assigned
- **Mid-September** we have an editors meeting
- **End of October** we release a draft to the community
They have 2 weeks

Release to ET collaboration by the end of October

- *Editor assignment & Progress tracking*
- **Collaborative document edition set up (Overleaf-CERN)**
- TBD also at Lyon Workshop on Onsite Infrastructure, Computing and Data Model. <https://indico.ego-gw.it/event/820/>



AFTER D8.1 RESUBMISSION

And not to forget: Actions from ET Data Analysis Workshop (Bologna, February 2025)

TODO LIST

Call for Computing
Resources



- Define a (set of) standard software distribution(s) [Elena]
- Provide a CVMFS repository for distribution (and policies for inclusion) [Stefano]
- Call for computing resources requests (also for MDC datasets production) [Stefano]
- ET Computing Cloud MoU [Stefano, Elena]

Data and Metadata
Requirements



- Data/metadata catalogue (annotations) [Paul]
 - or at least (for the time being) clear and permanent association between config file(s) and dataset
 - Define the requirements for metadata [Oscar]

FAIR for Software



- New MDC production code review for modularity [Anuradha]
- Prioritized list of requested datasets [Tania & Gianluca]
 - Resources (CPU, storage, personpower for development & running)
 - Strategies for reuse and "modular" generation
- Set up MDC analysis Code working group [Paul]
- Start definition of best practices for code [Andres]
- Organize MDC timeline definition meetings (or whatever) [Stefano]
- Define a plan for user-facing portal (docs, tutorials, data and code repos, à la GWOSC) [Alberto]
- ETAP/MADDEN/ETIC/whatever coordination (workshop?) [Federica]
- "Baseline" technology Q&A-based seminars [Sara]

And not to forget: Actions from ET Data Analysis Workshop (Bologna, February 2025)

1) Catalog of Computing & Storage resources at BSC for ET / ET-PP

- Via ET-PP RES project (Red Española de Supercomputación)
- VMs & HMS (disk/tape) storage plan (100TB 2025, 200TB 2026)

2) Data and Metadata requirements

- Based on data and metadata catalogue and/or association between config files and datasets
- Real and Simulated Data (MDC)
- “[LVK] Gravitational Wave detectors analyze their data together”
- NOW: Requirements for GW strain data release (As requested for D8.1)

3) FAIR best practices for ET / GW

- Taking into account FAIR for both Research Data & Software Management (**DOING**)
- Abstract submitted to XV ET Symposium: *Towards a FAIR Learning Path for ET* (**DONE**)
- NOW: Requirements for GW strain data release (As requested for D8.1)



EINSTEIN
TELESCOPE

And now for something completely different...





EINSTEIN
TELESCOPE

ET-PP WP8 D8.1 CDR EC RP2 Revision summary (1/3) Computing and data model

Review report 'cover letter'.

Remarks to be taken into account (or packages may be declared void (!))

*...the requirement to clarify the intentions of ET-PP on the ET data handling in Deliverable 8.1. It has been promised in the **Description of Action** that the **ET data management** [1] (an ET-PP deliverable) would follow **FAIR principles on open data**. When the question was posed directly during the review meeting, the consortium **refused to give a clear answer on the data retention period, which is simply not acceptable** [2]. Some more detailed arguments are given in the scientific review...*

*...[reviewer's] requests for re-opening some deliverables and milestone documents and revising them to give more substance is expected to be **fully followed**...*

*The **deadline** for this revision to finish is hereby set to **10 September**.*

[1] ET Data Management (Plan). D6.6 WP6. Due by the end of the ET-PP project (month 46)

[2] ET Data release/retention policy. **To be released.**



EINSTEIN
TELESCOPE

ET-PP WP8 D8.1 CDR EC RP2 Revision summary (2/3) Computing and data model

Report recommendations:

The **plans for low-latency release of gravitational wave strain data (or lack thereof) have to be provided**. I repeat that a commitment (or not) to low-latency data releases has huge implications on the rules governing the ET Collaboration, so **the endorsement (or not) of low-latency data releases must be addressed**. Plans are also necessary to **harmonise the data-sharing policies of ET and its partner, the Cosmic Explorer** which read : “released as open data as quickly as possible (i.e., as close to real time as possible)”, and with “no reserved science [...] for any group or collaboration” (Section 9.2 of the Cosmic Explorer Horizon Study, <https://arxiv.org/pdf/2109.09882.pdf>).

In terms of compliance with the DoA this **refusal to address this request is at odds with task 8.4: “Data Access Implementation”** / coordinator BSC: 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.



EINSTEIN
TELESCOPE

ET-PP WP8 D8.1 CDR EC RP2 Revision summary (3/3) Computing and data model

Report review: WP8 Deviations from Objectives and Work Plan

*Deliverable D8.1. Computing and Data requirements was **evaluated in Oct 2024 review**. We **requested some revisions**. Some were taken into account, **some only partially**, such as the arguments for **using a factor 5 in estimating ET data requirements** based on LIGO/Virgo detectors. [3]*

*The description of how to handle low latency **gravitational wave data releases** is **completely lacking**, and this would be a **significant deviation** from the current mode of operation, where data is released years after it is acquired. **So a plan is needed**. [4]*

[3] ...the rationale for using a factor 5 in estimating ET data requirements based on LIGO/Virgo detectors should be clearly presented.

[4] The plans for **low latency [data]** production and **public dissemination** of **calibrated** data, data **quality flags**, and their **revisions**, is lacking. **The report must include this.**

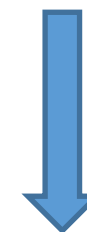
ET-PP WP8 D8.1 Computing and Data Requirements RP2 REVIEW REPORT, ACTION PLAN

Computing and data model

- **We must address the points as stated by the Officer / Referees, specially addressing the Data Release comments but also adhere to our original plan & project timelines ***
 - ET-PP **Data Mananagement Plan** (D6.6, WP6) is due by the end of ET-PP (month 46).
 - ET Open Data release policy is beyond scope of ET-PP, **interactions between ET-PP WPs and ET needed.**
 - ET-PP Annual meeting, Barcelona 22-24 July 2025 (**HERE WE ARE**)

* ET-PP Milestones & Deliverables

- D8.2 Computing and Data Model, WP8, month 42
- D6.6 Data Management Plan and Data Access Policy, WP6, month 46
- D8.3 Data Access Implementation, WP8, month 47





EINSTEIN
TELESCOPE

ET-PP WP8 D8.1 Computing and Data Requirements RP2 REVIEW REPORT, ACTION PLAN

Computing and data model

- **Meanwhile, our plan for D8.1 amendment (thx EiB Stefano, Steven, Patrice et al.)**
 - **Draft a response** to the reviewers in order to transmit them our approach and plans
 - **Prepare an update to D8.1** with further info and justification of the **5-factor** scaling
 - **Prepare an update to D8.1 with plan and associated requirements estimation for public release of GW strain data and metadata** for the Einstein Telescope based on **two potential public data release policies**:
 - **Cosmic Explorer like (aka full public immediate release)**. Proposed model by CE Horizon study, with “as close to real-time as possible” & unrestricted public data release [5]
 - **LIGO/Virgo/Kagra like (aka embargo)**. Current solution in use by LVK using 18-24month embargo period and GWOSC infrastructure [6]

**ET-PP WP8 + EiB
IN PROGRESS
FULL PRIORITY**



EINSTEIN
TELESCOPE

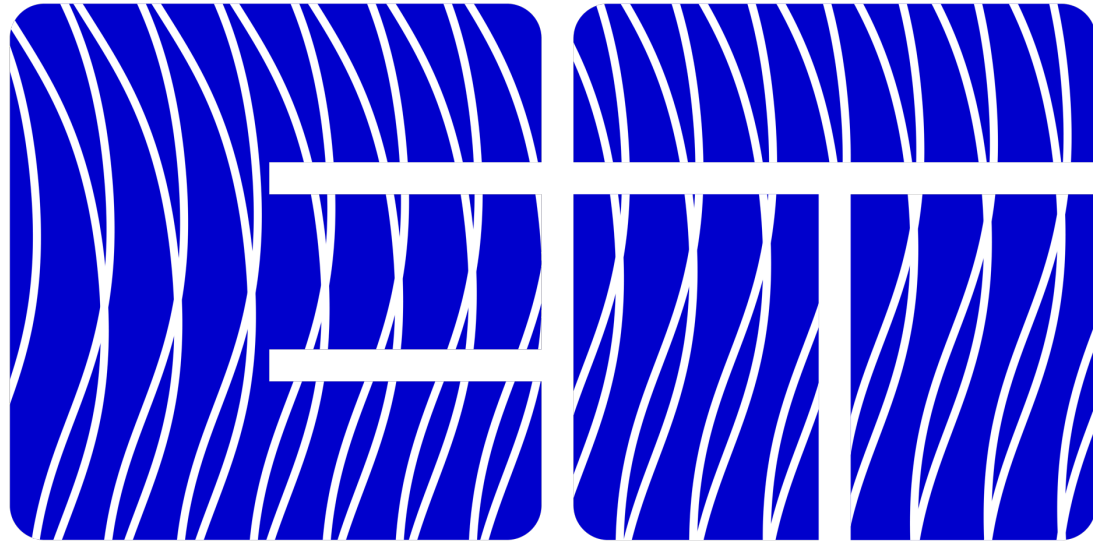
The CE conundrum

ET-PP WP8

Computing and data model

- «as close to real-time as possible»: does this mean also immediate, public LL data release?
 - Probably not, but...
 - “LL data” means streaming, non-persistent data with $o(\text{sec})$ latency + $o(\text{hours})$ buffer
 - “aggregated” $h(t)$ are the same data repacked with $o(\text{minutes})$ latency to origin server
 - reviewer comment is ambiguous (as is CE text)
 - Should we foresee subscribable LL data streams e.g., for multimodal searches?
- «Since the number of Cosmic Explorer control and data channels will be similar to that of Advanced LIGO, we expect data rates of 2 Tb per day of detector operation. Storage and dissemination of data of this scale is a solved problem with current technology».
 - Are they seriously planning to directly distribute raw data?
- «Approximately 10 FTEs will be needed to perform tasks related to data preparation, alert generation, data curation, and user community support during the operations phase of the project».
 - Our take: this is plain wrong (BTW, if this is the case there is no collaboration data analysis to take care of)
 - If we do not mention it they will tell us we did not mention it
 - If we do mention it the number will haunt us... 10? 15-20? 17.5? 59?

Public data only	17,5	4,66
Private analysis only	55,5	37,94
All public data	59	27,92
All private analysis	97	61,2
Both	40,4	21,36
Grand total	178,1	88,3



**EINSTEIN
TELESCOPE**

ET-PP WP8 Report

Computing and data model

Anything else ?

Questions ?

THANK YOU !