Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

EuMocks: Mocking the Universe for Euclid++

Pierluigi Monaco, UniTS (+INAF, INFN, IFPU, ICSC),
M. Lepinzan, T. Castro, L. Tornatore, G. Taffoni, C. Carbone

Spoke 3 Technical Workshop, Bologna December 17/19 2024

Scientific Rationale

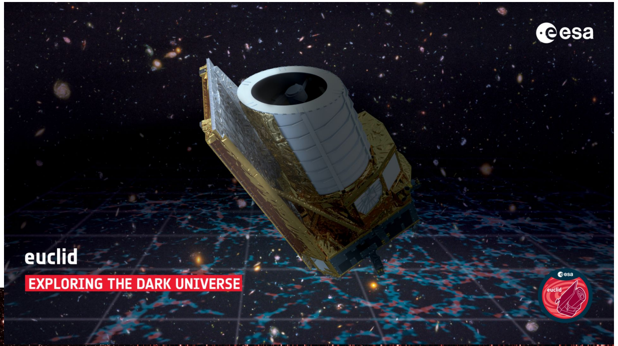
Euclid will survey the universe down to redshift z~2, **mapping the large-scale structure** to measure its geometry and growth rate to shed light on dark sector

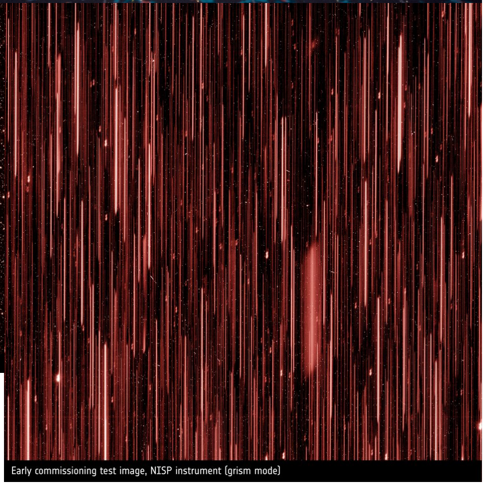
A spectroscopic sample will be based on **slitless spectroscopy** of ~14,000 sq deg of the sky, detecting the Halpha line at 0.9<z<1.8

Control of **systematic errors** will be the issue to tackle to provide convincing and potentially groundbreaking results.

To this aim we need **thousands of simulations** of the Universe observable by Euclid...

and by **SKAO**, DESI, LSST, Roman...





Euclid's first light images

Technical Objectives, Methodologies and Solutions (2023)

We aim at producing **3500 simulations of a volume of ~4Gpc with ~10^{12} particles, resolving halos of ~10^{11} M_{sun}/h**, with output on a past-light-cone covering **half of the sky** and starting at **z=3**.

Standard N-body codes are too slow to produce such a massive set of simulations
Approximate methods: **PINOCCHIO**

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Technical Objectives, Methodologies and Solutions

We already produced **1000 simulations of a volume of ~3Gpc with ~2e10^{11} particles, resolving halos of ~10^{11} M$_{sun}$/h**, with output on a past-light-cone covering **half of the sky** and starting at **z=4**.

These are supported by a larger set of simulations based on smaller boxes or on boxes with lower mass resolution.

Name	L_s	N_{part}	N_{real}	min. M_h	V_s	tot. V_s	θ	area	z_{start}	V_s	tot. V_{lc}
Geppetto	1.2	2160^3	3500	1.52e+11	1.73	6047	30°	2763	2.0	12.72	44520
EuclidLargeMocks	3.38	6144^3	1000	1.48e+11	38.61	38614	70°	13571	4.0	163.86	163855
Minerva-like	1.5	1000^3	10000	2.67e+12	3.38	33750	—	—	—	—	—
NewClusterMocks	3.87	2160^3	1000	4.90e+12	57.96	57960	60°	10313	2.5	69.70	69700

Technical Objectives, Methodologies and Solutions

We already produced **1000 simulations of a volume of ~3Gpc with ~2e10^{11} particles, resolving halos of ~10^{11} M$_{sun}$/h**, with output on a past-light-cone covering **half of the sky** and starting at **z=4**.

These are supported by a larger set of simulations based on smaller boxes or on boxes with lower mass resolution.

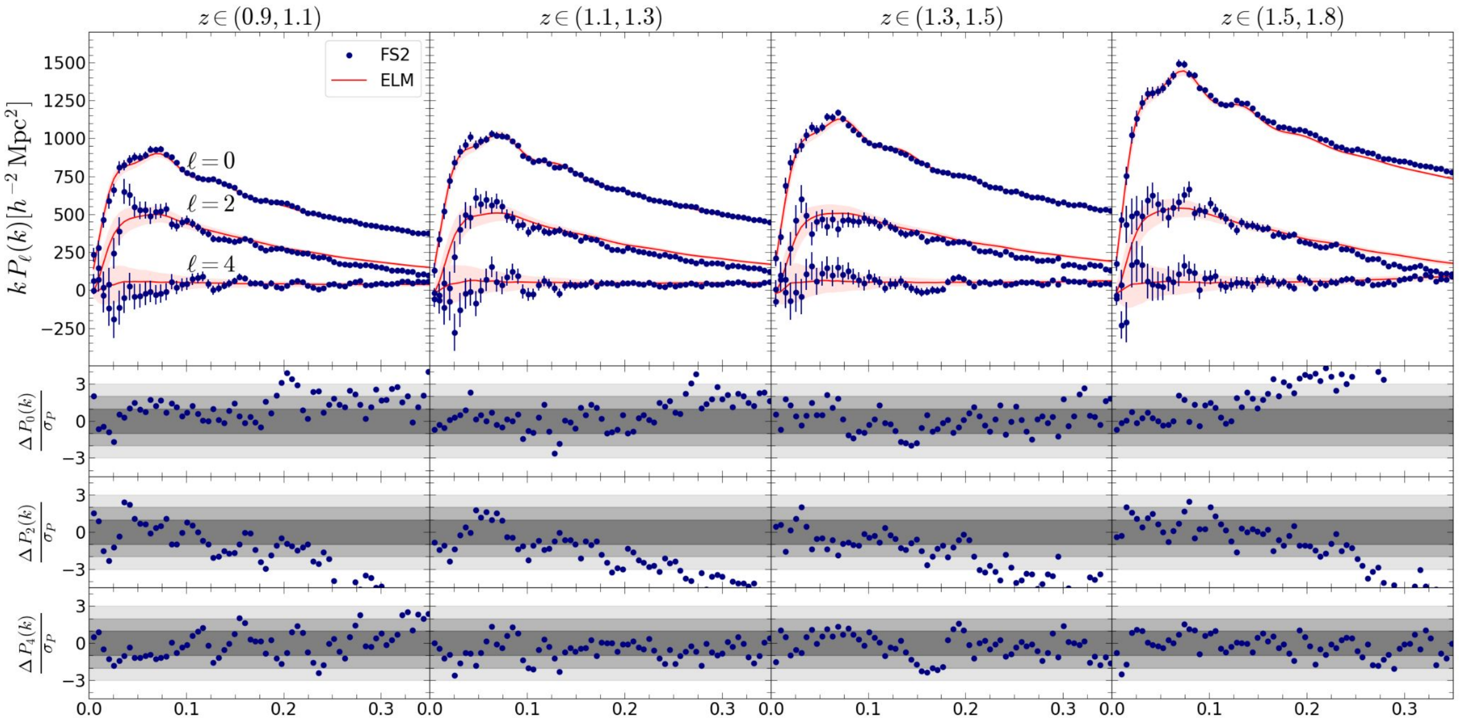
Name	L_s	N_{part}	N_{real}	min. M_h	V_s	tot. V_s	θ	area	z_{start}	V_s	tot. V_{lc}
Geppetto	1.2	2160^3	3500	1.52e+11	1.73	6047	30°	2763	2.0	12.72	44520
EuclidLargeMocks	3.38	6144^3	1000	1.48e+11	38.61	38614	70°	13571	4.0	163.86	163855
Minerva-like	1.5	1000^3	10000	2.67e+12	3.38	33750	—	—	—	—	—
NewClusterMocks	3.87	2160^3	1000	4.90e+12	57.96	57960	60°	10313	2.5	69.70	69700

The largest collection of halos on the lightcone ever simulated!

Spoke 3 + Euclid paper in preparation + press release

Technical Objectives, Methodologies and Solutions

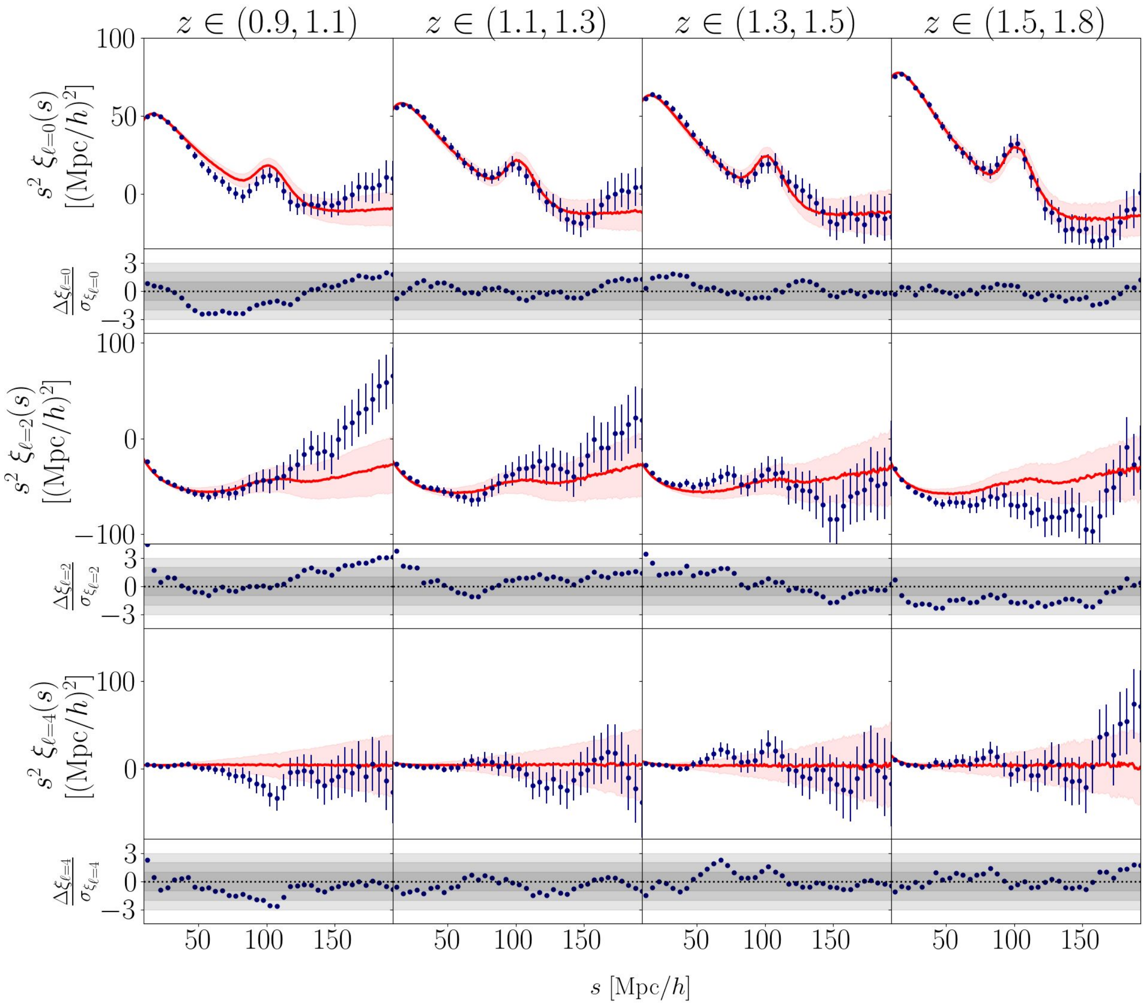
Power spectrum of **Flagship galaxy mock catalog** (based on N-body) compared to **1000 EuclidLargeMocks** calibrated on it



Credit: Yousry Elkhashab

Technical Obje

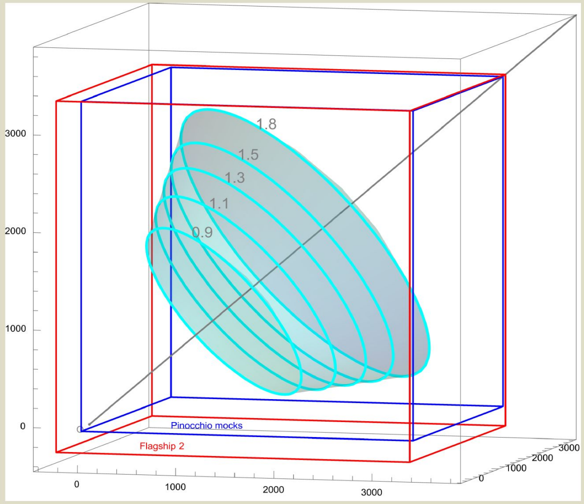
2-point correlation function of **Flagship galaxy mock catalog** (based on N-body) compared to **1000 EuclidLargeMocks** calibrated on it

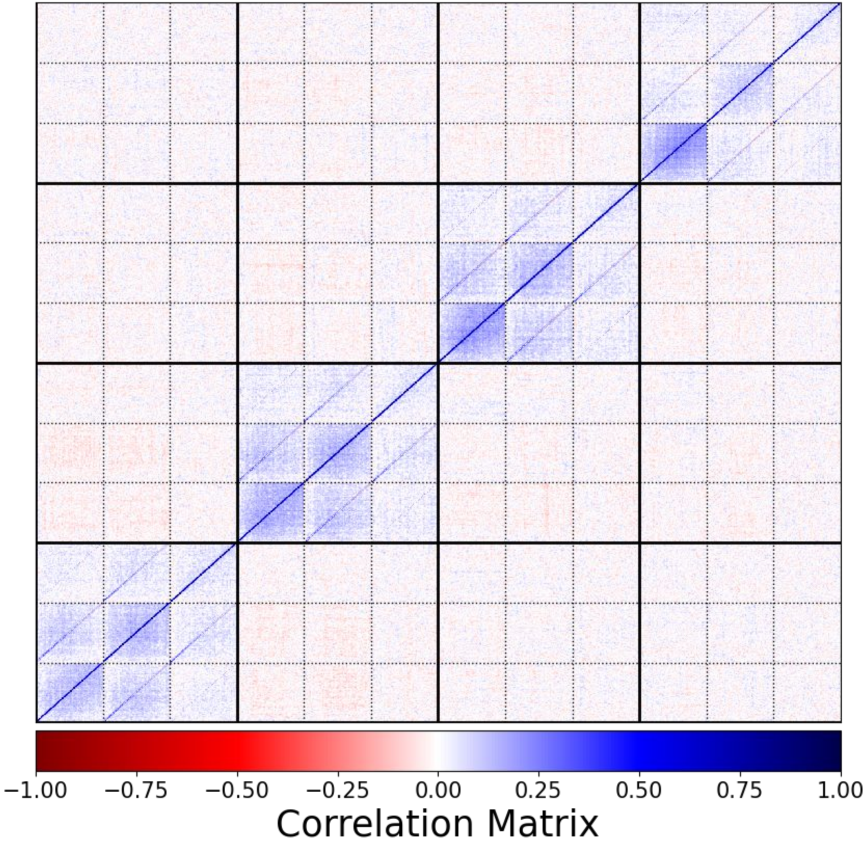


Credit: Gabriele Parimbelli

What we need in Euclid

Correlation matrix of power spectrum measurements on an idealized survey geometry (a cone of 60 degree of aperture, 2763 deg^2





Credit: Emiliano Sefusatti, Jacopo Salvalaggio

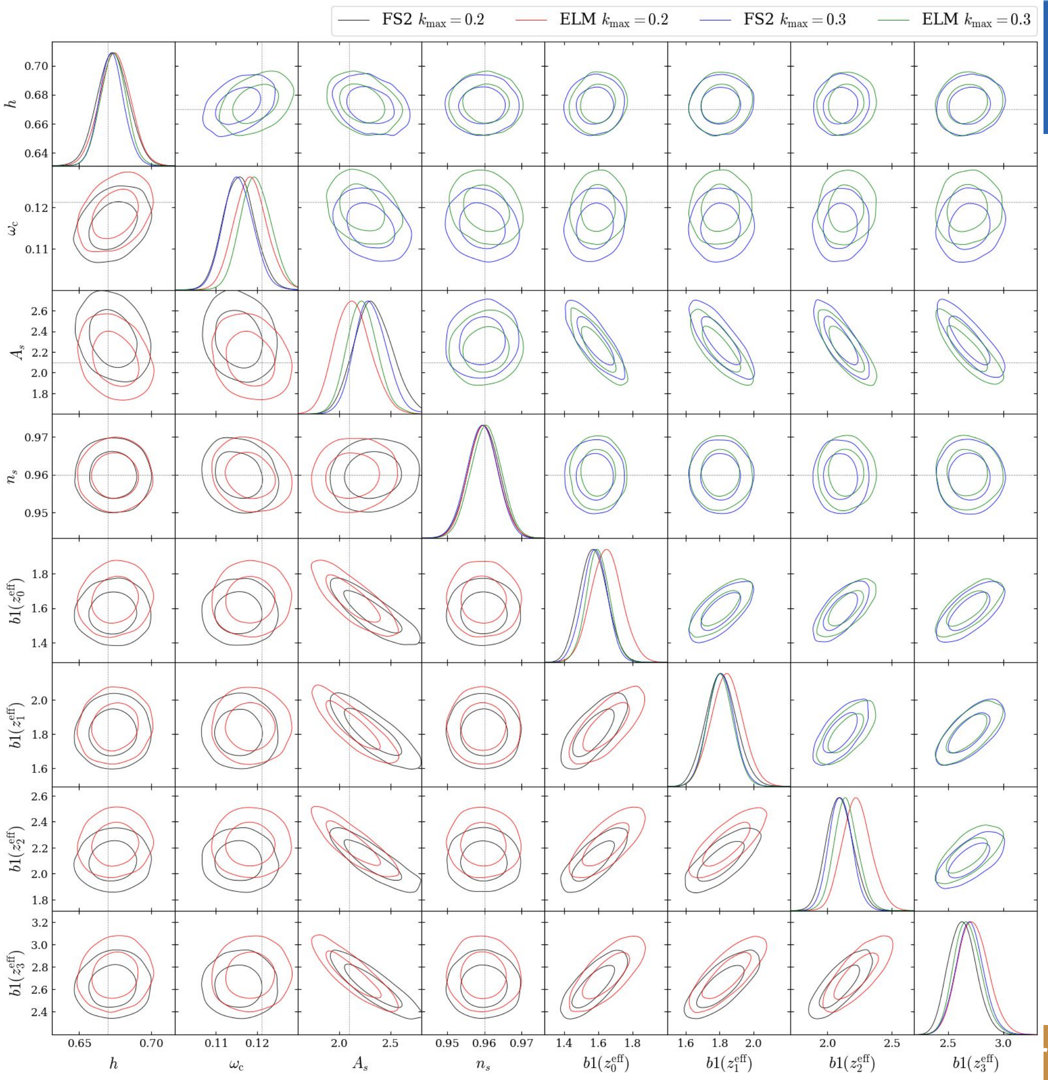
The definitive test

+ take the **Flagship mock** as data vector
+ use **EuclidLargeMocks** for covariance
+ fit the data with a model
+ obtain **posteriors** for parameters
+ check consistency of cosmological parameters

Then:

+ take one of the **EuclidLargeMocks**
+ use it as a data vector
+ repeat the fit
+ check **consistency** of posteriors

Credit: Yousry Elkhashab



Technical Objectives, Methodologies and Solutions (2023)

We aim at producing **3500 simulations of a volume of ~4Gpc with ~10^{12} particles, resolving halos of ~10^{11} M_{sun}/h**, with output on a past-light-cone covering **half of the sky** and starting at **z=3**.

Standard N-body codes are too slow to produce such a massive set of simulations
Approximate methods: **PINOCCHIO**

Development is needed (Spoke3 + Spoke1, PRIN PNRR 2022):
SCIENCE:
- improve and optimize the **reconstruction of halos**, especially in filaments
- implement a **Particle-Mesh** code to move halos (instead of 3LPT)
- add **lensing** and relativistic effect

CODING:
- port on **GPUs**
- improve the **fragmentation (deblending, clustering)** algorithm

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Technical Objectives, Methodologies and Solutions

We aim at producing **3500 simulations of a volume of ~4Gpc with ~10^{12} particles, resolving halos of ~10^{11} M_{sun}/h**, with output on a past-light-cone covering **half of the sky** and starting at **z=3**.

Standard N-body codes are too slow to produce such a massive set of simulations
Approximate methods: **PINOCCHIO**

Development is needed (Spoke3 + Spoke1, PRIN PNRR 2022):
SCIENCE:
- improve and optimize the **reconstruction of halos**, especially in filaments -> 20%
- implement a **Particle-Mesh** code to move halos (instead of 3LPT) -> to be started
- add **lensing** and relativistic effect -> 80%

CODING:
- port on **GPUs**
  - **+ collapse times** -> 100%
  - **+ FFTs** (+De Rubeis, Lacopo, Gheller) -> 80%
- improve the **fragmentation (deblending, clustering)** algorithm -> 20%

Technical Objectives, Methodologies and Solutions

We aim at producing **3500 simulations of a volume of ~4Gpc with ~10^{12} particles, resolving halos of ~10^{11} M_{sun}/h**, with output on a past-light-cone covering **half of the sky** and starting at **z=3**.

Standard N-body codes are too slow to produce such a massive set of simulations
Approximate methods: **PINOCCHIO**

Development is needed (Spoke3 + Spoke1, PRIN PNRR 2022):
SCIENCE:
- improve and optimize the **reconstruction of halos**, especially in filaments
- implement a **Particle-Mesh** code to move halos (instead of 3LPT)
- add **lensing** and relativistic effect

CODING:
- port on **GPUs**
  + collapse times
  + FFTs (+De Rubeis, Lacopo, Gheller)
- improve the **fragmentation (deblending, clustering)** algorithm

See talk by Marius Lepinzan

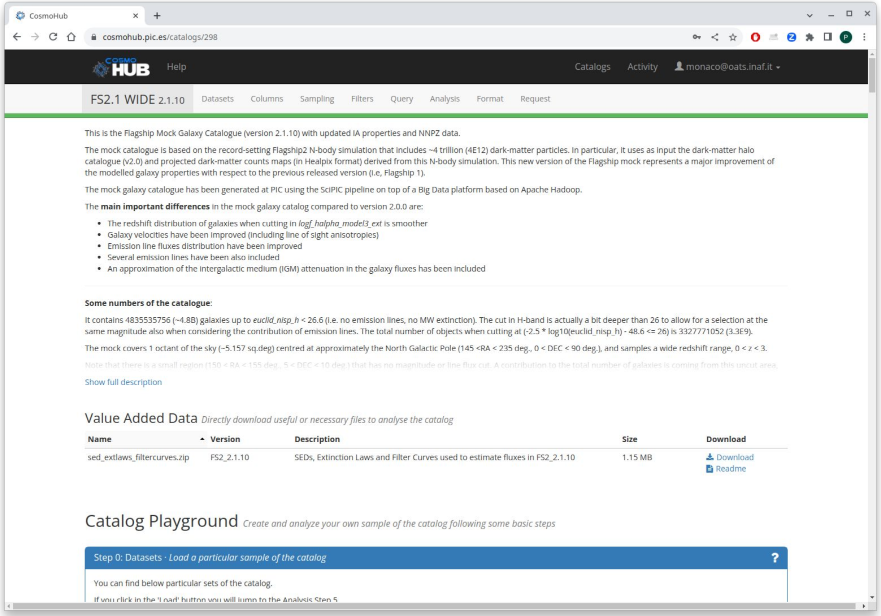
Challenges

Massively parallel code, every step must be optimized before burning so much computing time

Computing time: ~30,000,000 core hours
Memory: ~128 TB
Storage: >~1PB

Petabyte-scale output to be offered to the community
-> National / Interoperable Data Lake

example: Cosmohub.pic.es



Timescale, Milestones... and KPIs (2023)

2023 | **2024** | **2025** | **2026**

launch
July 2023

DR1 data
Sept 2024

DR1 public
Sept 2025

DR2 data
Sept 2026

Milestone 6 | Milestone 7 | Milestone 8 | Milestone 9 | Milestone 10

ICSC
year 1

ICSC
year 2

ICSC
END

ICSC resources available

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

code development........................

available simulations....................

Sept 2024:
available simulations
in the data lake?

March 2025
code ready

March 2025
data lake ready

August 2025
first 100 sims

August 2026
3500 sims

KPI | **KPI** | **KPI**

Timescale, Milestones... and KPIs (2024)

2023 **2024** **2025** **2026**

launch
July 2023

DR1 data
Oct 2025

DR1 public
Oct 2026

Milestone 6 | Milestone 7 | Milestone 8 | Milestone 9 | Milestone 10

ICSC
year 1

ICSC
year 2

ICSC
END

ICSC resources
available

code development.......................

available simulations...................

Sept 2024:
available simulations
in the data lake
STARTED, 20%

KPI

March 2025
code ready
now: 70%

March 2025
data lake ready

KPI

August 2025
first 100 sims

KPI

August 2026
3500 sims

Next Steps and Expected Results

Code development:
-> improved halo definition, by May 2025 (M. Lepinzan)
-> PM code, by December 2025 (P. Monaco, M. Lepinzan)
-> lensing, by August 2025 (Y. Elkhashab, PRIN-PNRR)

-> collapse times to GPUs, completed (M. Lepinzan)
-> FFTs to GPUs, by May 2025 (D. Goz, G. Lacopo)
-> halo construction to GPUs, by August 2025 (M. Lepinzan)

Simulation production:
-> demonstrator set, by August 2025 (M. Lepinzan, P. Monaco)

Simulations in the datalake:
-> available simulations, by May 2025 (...) (with WP5 and WP4)
-> new simulations, as soon as they are produced (...)

(...) means: it depends on resource availability!