



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani

PIANO NAZIONALE
DI RIPRESA E RESILIENZA



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 $z \sim 2$, **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 $\sim 14,000$ sq deg of the sky, detecting the H α line at $0.9 < z < 1.8$

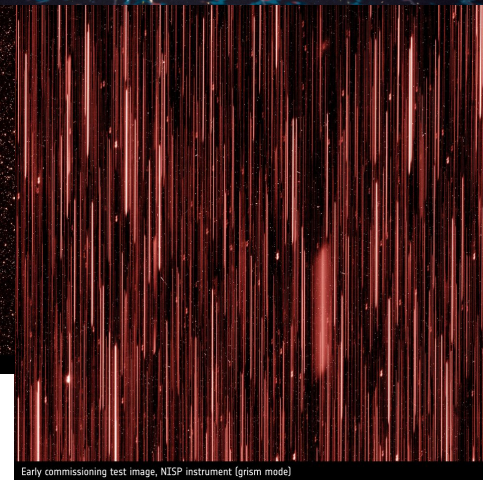
Control of **systematic errors** will be the issue to tackle to provide convincing and potentially groundbreaking results.

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 $\sim 4\text{Gpc}$ with $\sim 10^{12}$ particles, resolving halos of $\sim 10^{11} M_{\text{sun}}/h$** , 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**



Technical Objectives, Methodologies and Solutions

We already produced **1000 simulations of a volume of $\sim 3\text{Gpc}$ with $\sim 2e10^{11}$ particles, resolving halos of $\sim 10^{11} M_{\text{sun}}/h$** , 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 $\sim 3\text{Gpc}$ with $\sim 2e10^{11}$ particles, resolving halos of $\sim 10^{11} M_{\text{sun}}/h$** , 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.



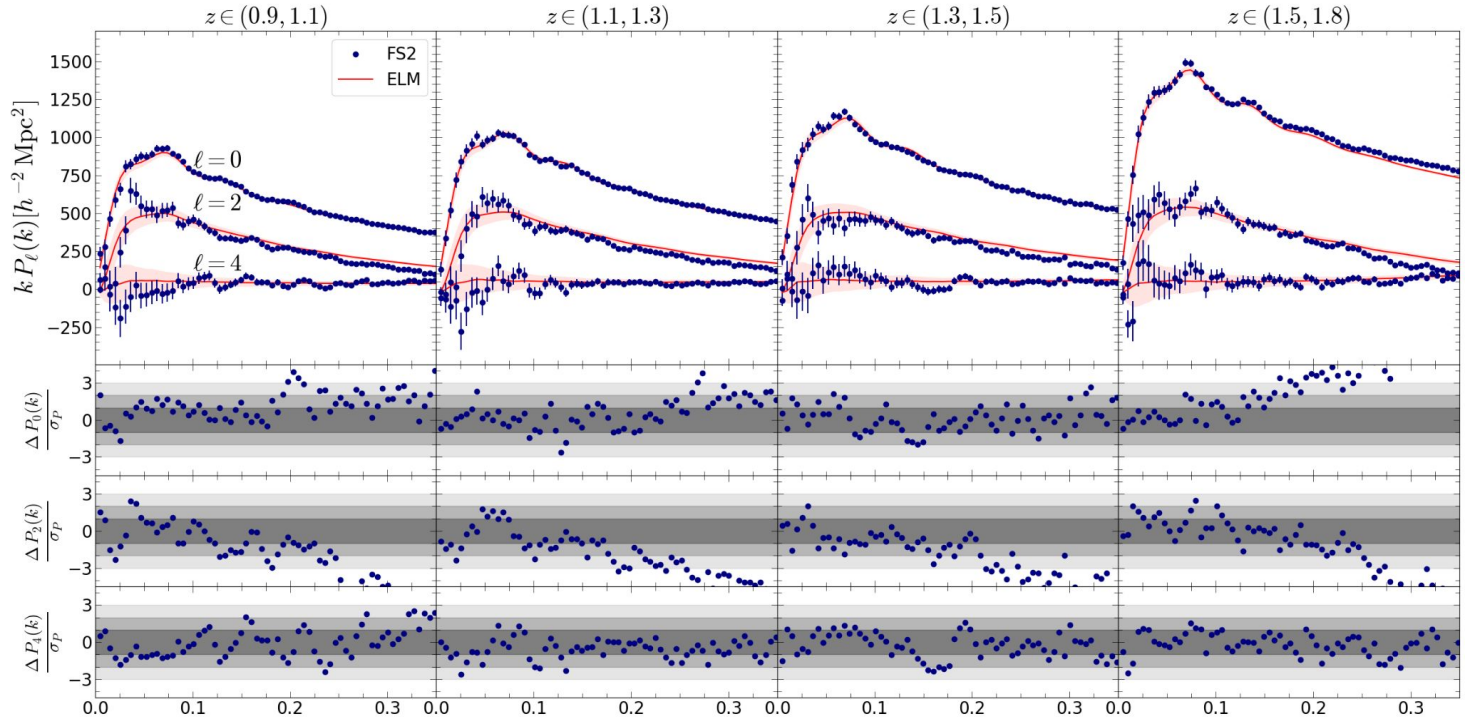
Name	L_s	N_{part}	N_{real}	min. M_h	V_s	tot. V_s	θ	area	z_{start}	V_s	tot. V_{lc}
Geppetto	1.2	2160^3	3500	$1.52e+11$	1.73	6047	30°	2763	2.0	12.72	44520
EuclidLargeMocks	3.38	6144^3	1000	$1.48e+11$	38.61	38614	70°	13571	4.0	163.86	163855
Minerva-like	1.5	1000^3	10000	$2.67e+12$	3.38	33750	—	—	—	—	—
NewClusterMocks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

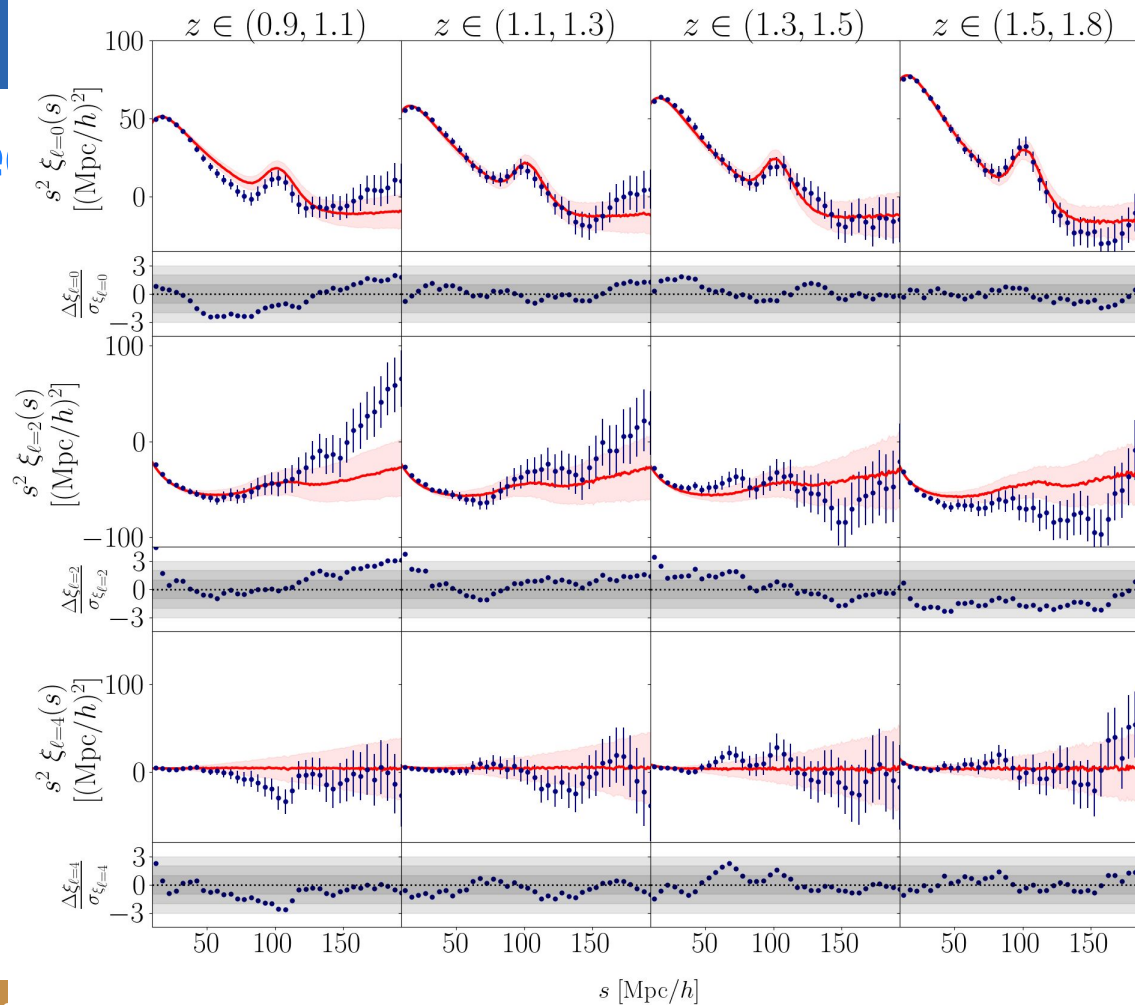
Power spectrum of Flagship galaxy mock catalog (based on N-body) compared to 1000 EuclidLargeMocks calibrated on it



Credit: Yousry Elkhatab

Technical Objective

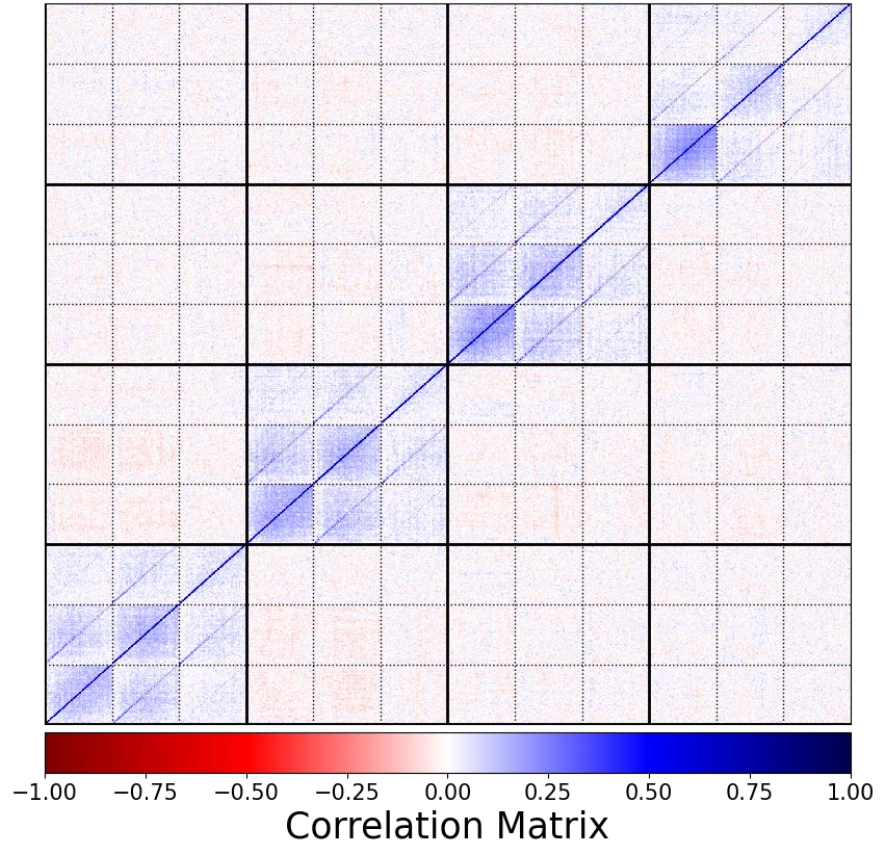
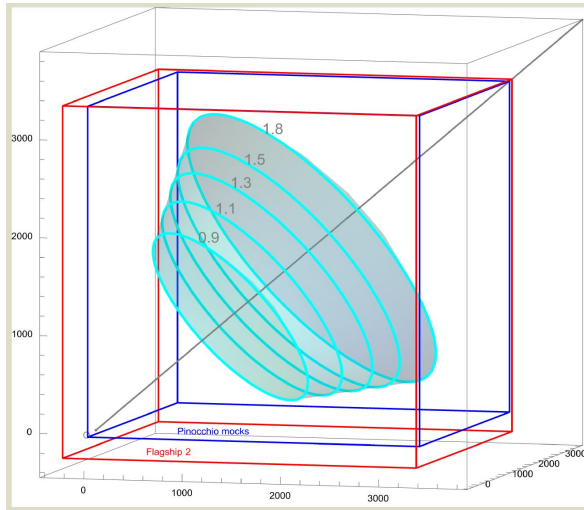
2-point correlation function of **Flagship galaxy mock catalog** (based on N-body) compared to **1000 EuclidLargeMocks** calibrated on it



Credit: Gabriele Paribelli

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

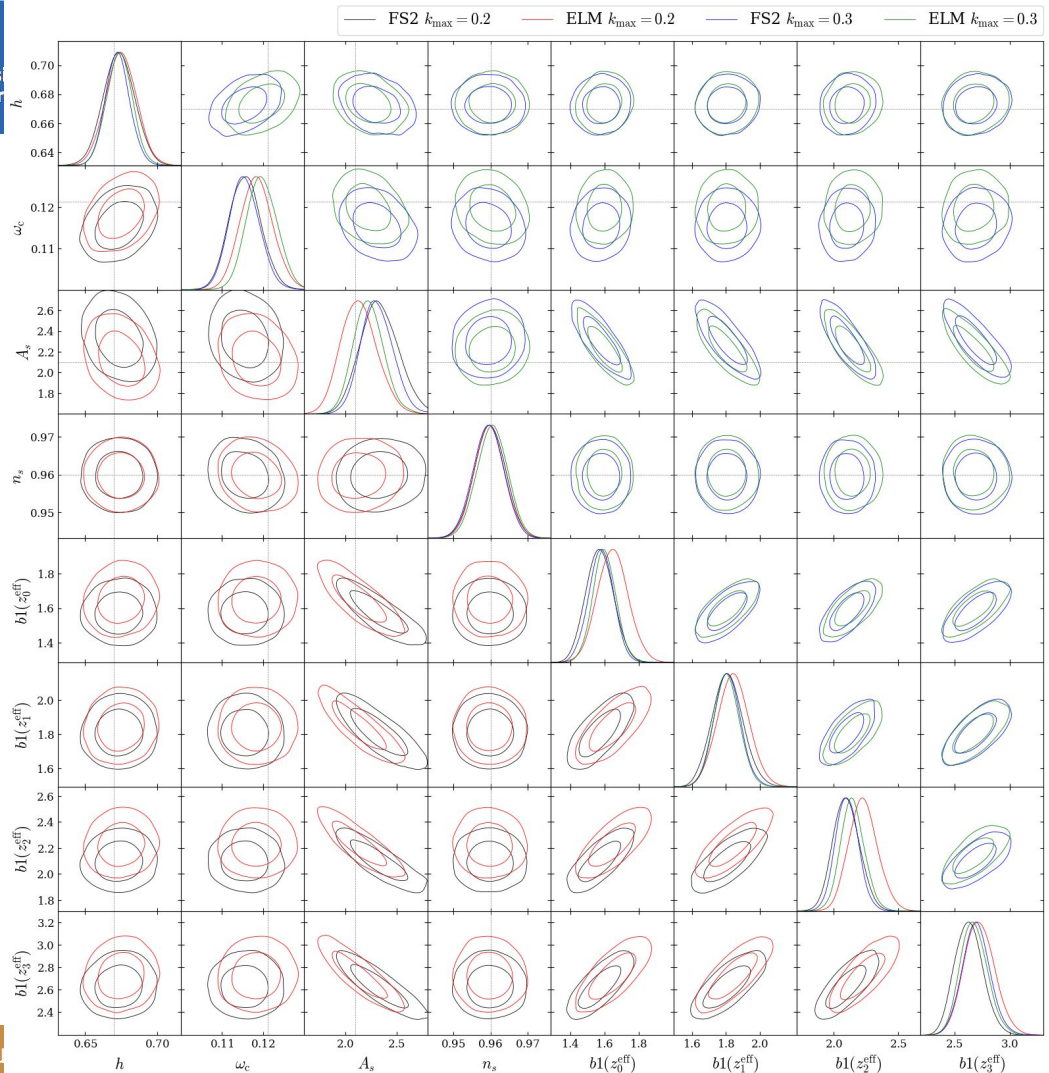
The definitive test

- + 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 Elkhatab



Technical Objectives, Methodologies and Solutions (2023)

We aim at producing **3500 simulations of a volume of ~4Gpc with ~10¹² particles, resolving halos of ~10¹¹ 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



Technical Objectives, Methodologies and Solutions

We aim at producing **3500 simulations of a volume of $\sim 4\text{Gpc}$ with $\sim 10^{12}$ particles, resolving halos of $\sim 10^{11} M_{\text{sun}}/h$** , 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 ~10¹² particles, resolving halos of ~10¹¹ 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

The screenshot shows a web browser window displaying the CosmoHub website. The page title is "FS2.1 WIDE 2.1.1.0". The main content area contains the following text:

This is the Flagship Mock Galaxy Catalogue (version 2.1.1.0) with updated IA properties and NN2 data.

The mock catalogue is based on the record-setting Flagship2 N-body simulation that includes ~4 trillion (4E12) dark-matter particles. In particular, it uses as input the dark-matter halo catalogue (v2.0) and projected dark-matter counts maps (in Healpix format) derived from this N-body simulation. This new version of the Flagship mock represents a major improvement of the modelled galaxy properties with respect to the previous released version (i.e., Flagship 1).

The mock galaxy catalogue has been generated at PIC using the SciPIC pipeline on top of a Big Data platform based on Apache Hadoop.

The **main important differences** in the mock galaxy catalog compared to version 2.0.0 are:

- The redshift distribution of galaxies when cutting in $\log_{10}(\text{halpha_mod}3_ext)$ is smoother
- Galaxy velocities have been improved (including line of sight anisotropies)
- Emission line fluxes distribution have been improved
- Several emission lines have been also included
- An approximation of the intergalactic medium (GM) attenuation in the galaxy fluxes has been included

Some numbers of the catalogue:

It contains 483533756 (~4.8B) galaxies up to $\text{euclid_nisp_h} < 26.6$ (i.e. no emission lines, no MW extinction). The cut in H-band is actually a bit deeper than 26 to allow for a selection at the same magnitude also when considering the contribution of emission lines. The total number of objects when cutting at $(-2.5 * \log_{10}(\text{euclid_nisp_h}) - 48.6 < z < 26)$ is 3327771052 (3.3B).

The mock covers 1 octant of the sky ($-5.157 \text{ rad} < \text{sed} < 10 \text{ deg}$) centred at approximately the North Galactic Pole ($145 < \text{RA} < 235 \text{ deg}$, $0 < \text{DEC} < 90 \text{ deg}$), and samples a wide redshift range: $0 < z < 3$.

Note that there is a small region ($150 < \text{RA} < 155 \text{ deg}$, $5 < \text{DEC} < 10 \text{ deg}$) that has no magnitude or line flux cut. A contribution to the total number of galaxies is coming from this uncut area.

Show full description

Value Added Data *Directly download useful or necessary files to analyse the catalog*

Name	Version	Description	Size	Download
sed_extlaws_filtercurves.zip	FS2_2.1.1.0	SEDs, Extinction Laws and Filter Curves used to estimate fluxes in FS2_2.1.1.0	1,15 MB	Download Readme

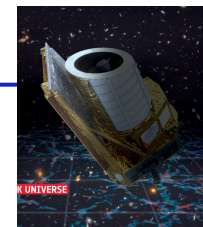
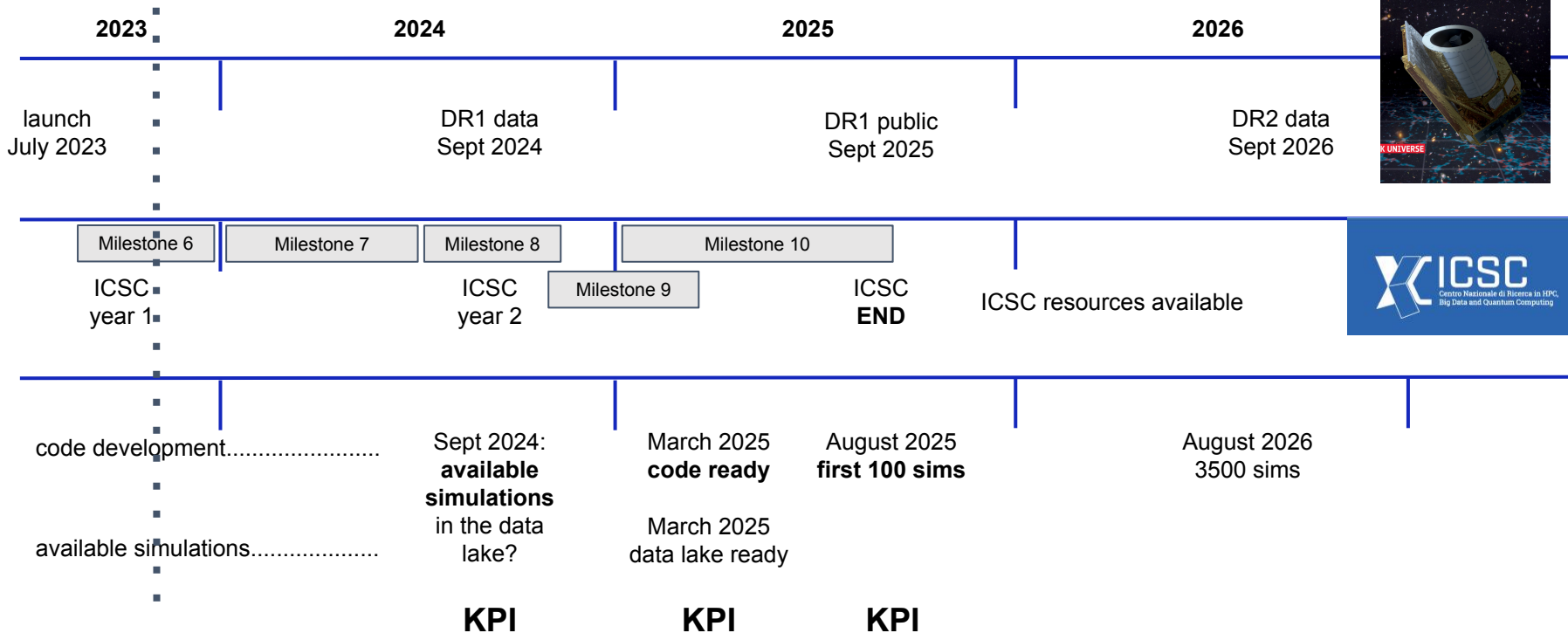
Catalog Playground *Create and analyze your own sample of the catalog following some basic steps*

Step 0: Datasets · Load a particular sample of the catalog

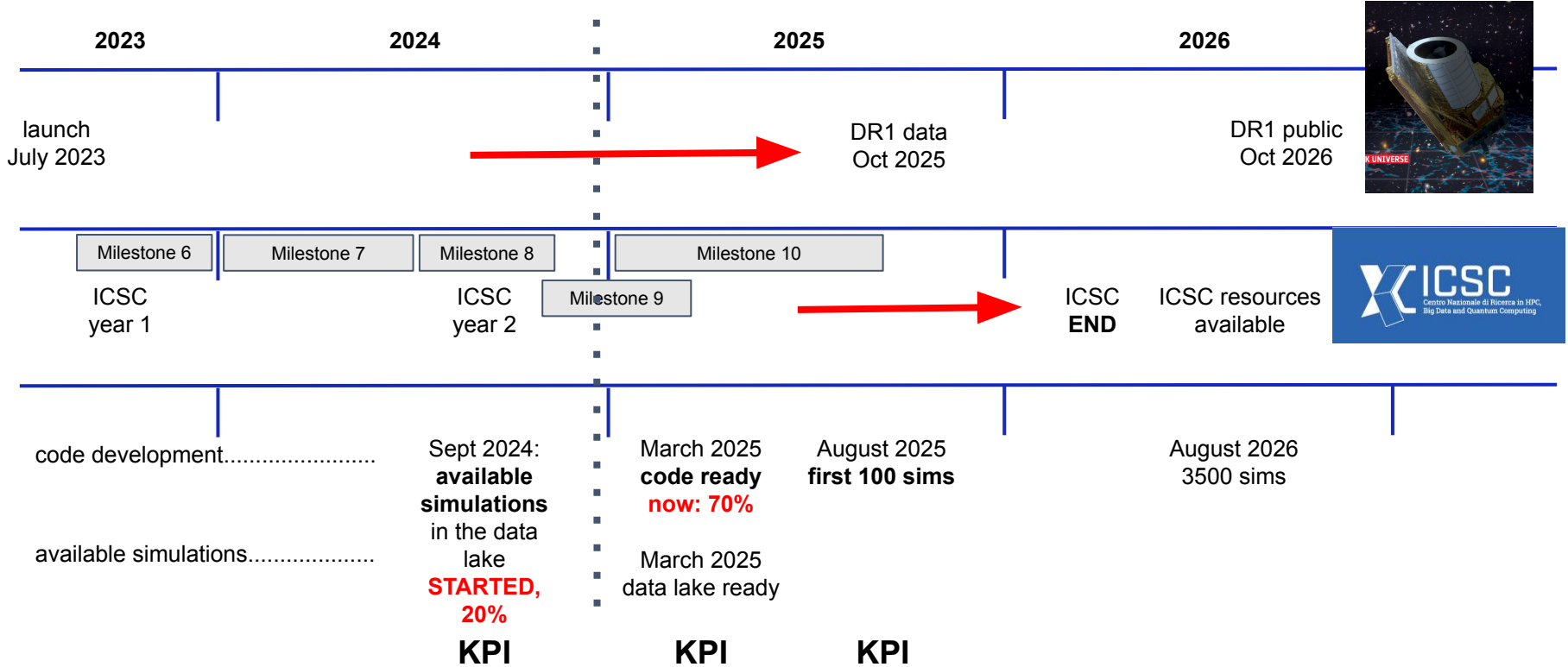
You can find below particular sets of the catalog.

If you click in the "Load" button you will jump to the Analysis Step 1.

Timescale, Milestones... and KPIs (2023)



Timescale, Milestones... and KPIs (2024)



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. Elkhatab, 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!