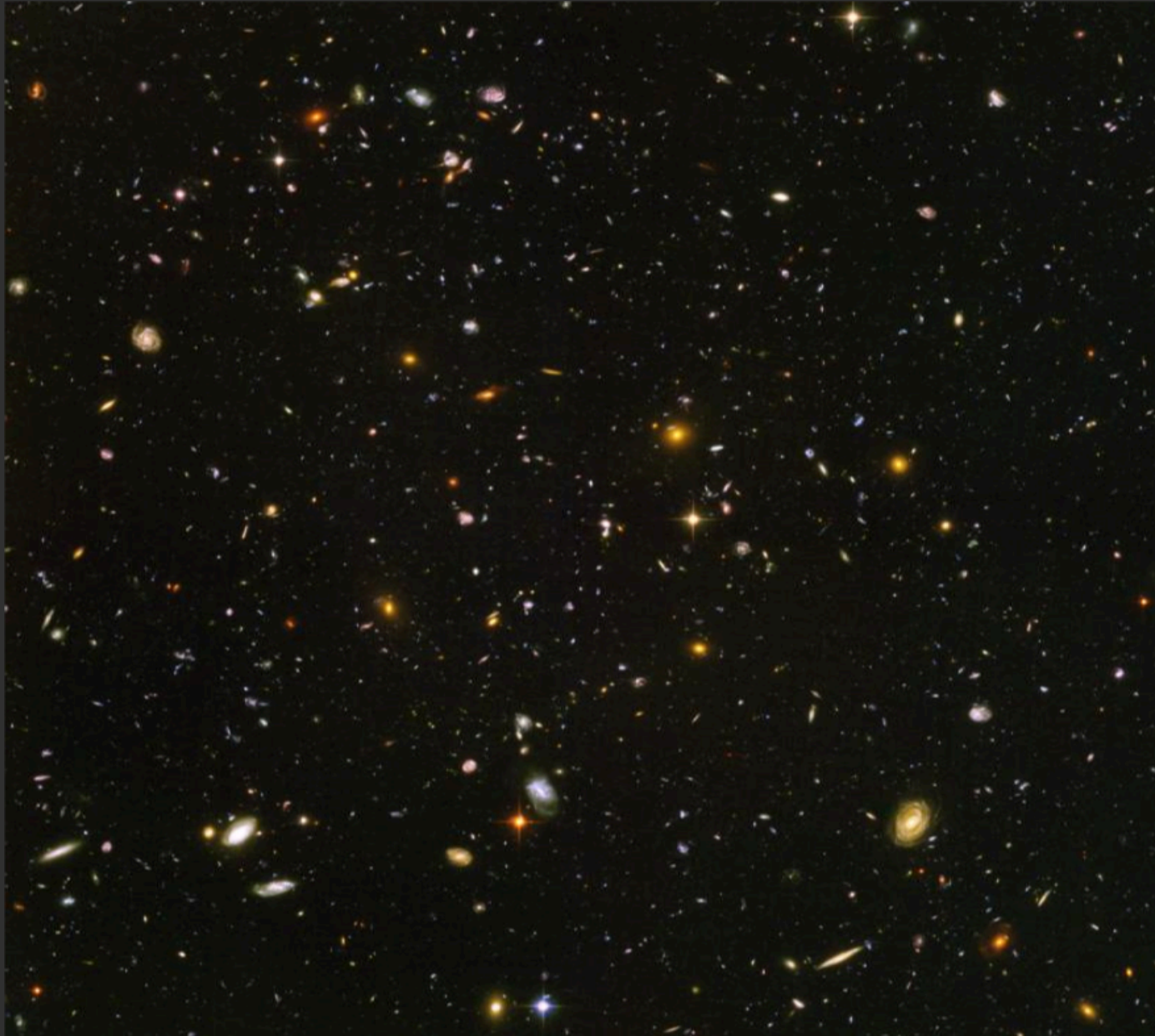


# Cosmology @EOSC

## the HPC Universe in the Cloud

Francisco Prada  
IAA-CSIC

# Our view of the distant Universe:



Is the visible material all there is?

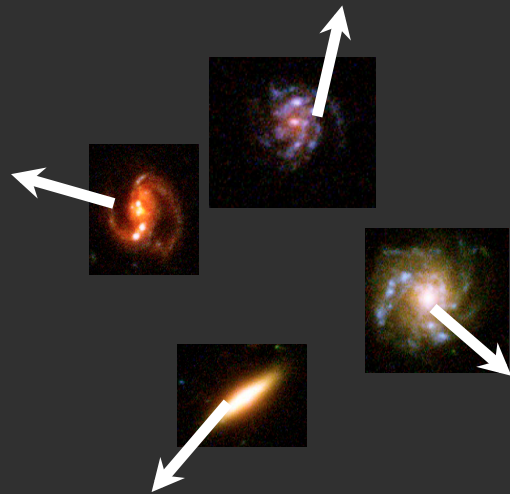
Hubble Ultra-deep Field

# Dark Cosmology



Much more gravity than we would expect:

→ **Dark Matter (27%)**

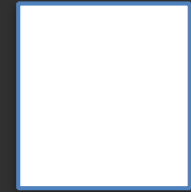


Galaxies moving apart faster and faster:

→ **Dark Energy (68%)**

**New Physics, beyond the Standard Model**

# Observation and Theory interplay



**Dark matter** and **dark energy** have predictable effects on:

- **Cosmic Microwave Background**
- **Patterns** of galaxies
- **Bending of light** in Universe
- Brightness of **supernovae**



The Nobel Prize  
in Physics 2011

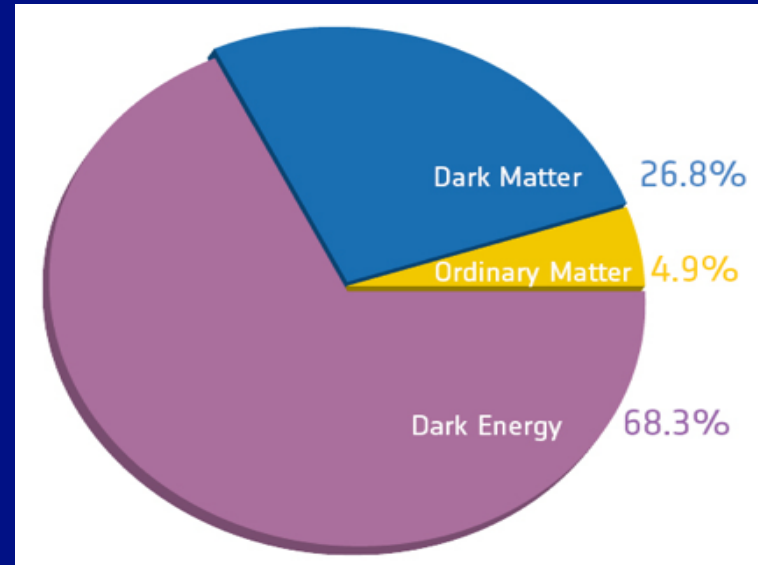
➔ Measure these effects in our **large surveys**

➔ Compare with **theories** and **simulations** containing dark Physics

# Top Scientific Objectives

## Physics of the Universe

### Understanding Scientific Principles



The two highest level questions in the field are the following:

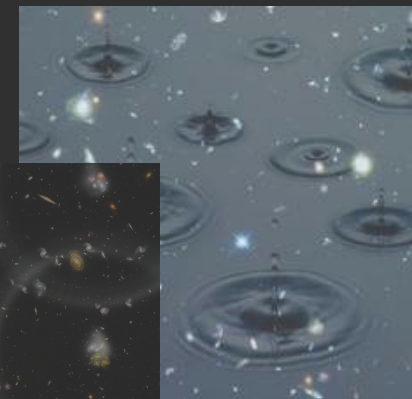
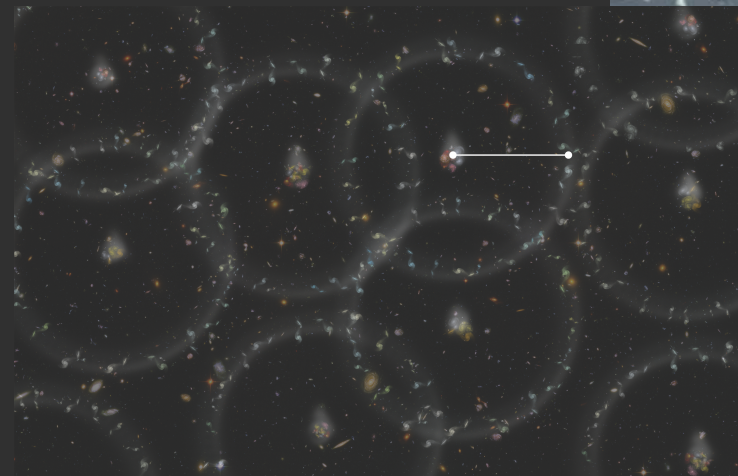
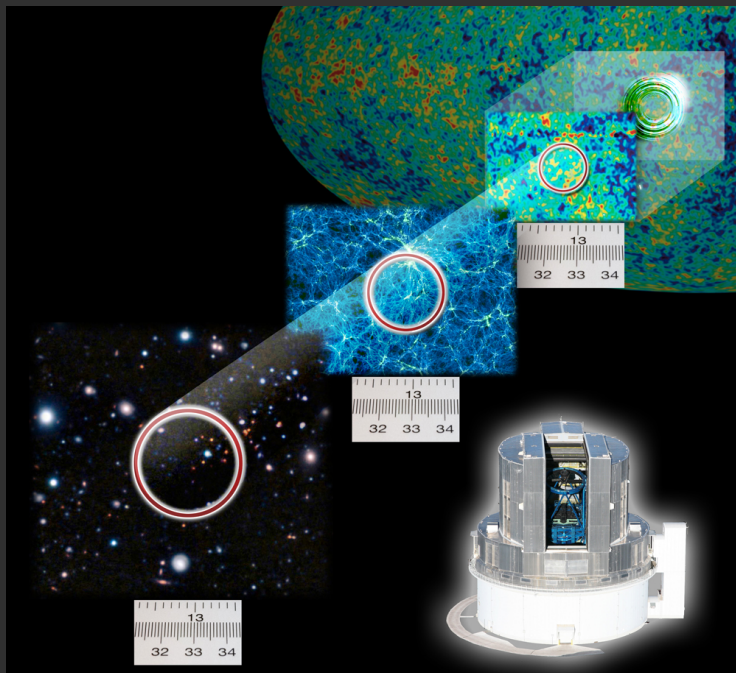
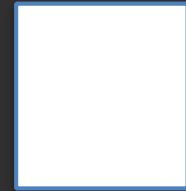
- Is cosmic acceleration caused by a breakdown of Einstein General Relativity on cosmological scales, or is it caused by a new energy component with negative pressure ("dark energy") within General Relativity?
- If the acceleration is caused by "dark energy," is its energy density constant in space and time and thus consistent with quantum vacuum energy or does its energy density evolve in time and/or vary in space?

# From surveys to measurements of dark physics

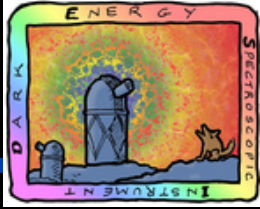
- ➔ Measure statistics of galaxy properties in large surveys
- ➔ Compare with predictions for particular dark energy physical values

e.g. Baryonic Acoustic Oscillations (BAO):

Measure correlation function of galaxy positions – BAO peak is found at a physical scale that depends on **dark energy!**



# Large Survey Projects



A collection of logos for several astronomical survey projects. From left to right: VIPERS (VIMOS PUBLIC EXTRAGALACTIC REDSHIFT SURVEY), LOFAR (Low Frequency Array), Quijote Project (showing a satellite dish), MOST (Multi-Object Spectroscopic Telescope), and Javalambre Physics of the Accelerating Universe Astrophysical Survey (showing a grid of colored squares).

Logos for the Sloan Digital Sky Survey (SDSS-II and SDSS-III) and the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX). HETDEX is described as "Illuminating the Darkness".

Logos for space-based survey missions. Planck is shown with the ESA and European Union flags. eRosita is shown with a satellite dish. COre/PRISM is shown with a satellite dish. Euclid is shown with a satellite dish. A central image of Earth is also present.

Logos for the Dark Energy Survey (DES), the Large Synoptic Survey Telescope (LSST), and the flag of Argentina.

Logos for the Square Kilometre Array Africa (SKA Africa) and the flag of South Africa.

**Table B2.1:** [S] designates a spectroscopic redshift survey, [I] an imaging survey and [R] a radio survey.

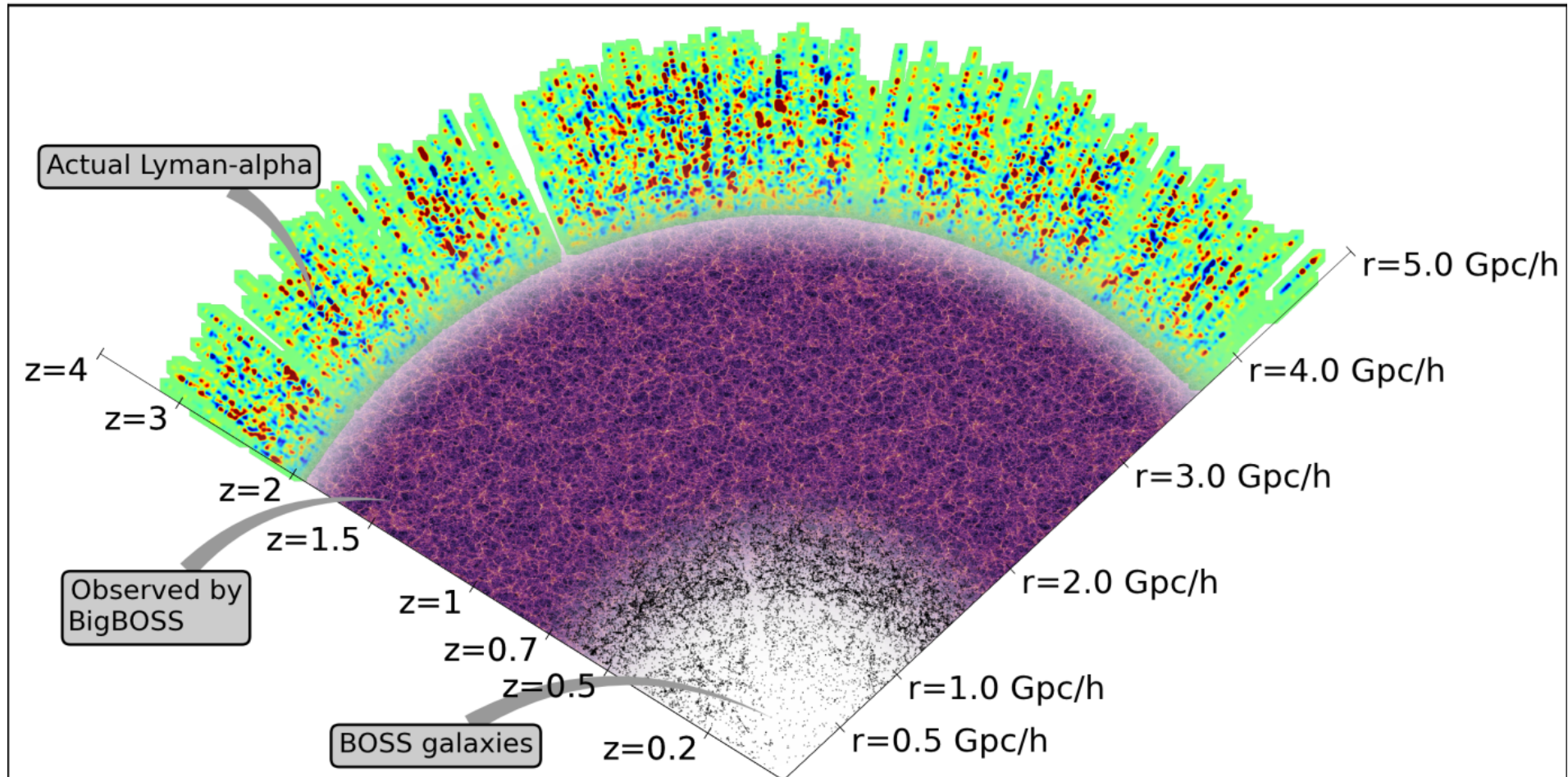
	Projects	Status	Ref.
	[S] VIPERS ✓	2009-2015	<a href="http://vipers.inaf.it/">http://vipers.inaf.it/</a>
→	[S] SDSS-III/BOSS ✓	2009-2014	<a href="http://www.sdss3.org/surveys/boss.php">http://www.sdss3.org/surveys/boss.php</a>
	[I] DES ✓	2012-2017	<a href="http://www.darkenergysurvey.org/">http://www.darkenergysurvey.org/</a>
	[I] VST/KIDS ✓	2011-2016	<a href="http://kids.strw.leidenuniv.nl/">http://kids.strw.leidenuniv.nl/</a>
	[I] eROSITA	2015-2020	<a href="http://www.mpe.mpg.de/erosita/">http://www.mpe.mpg.de/erosita/</a>
	[S] HETDEX	2015-2017	<a href="http://hetdex.org/">http://hetdex.org/</a>
→	[S] SDSS-IV/eBOSS ✓	2014-2020	<a href="http://www.sdss3.org/future/eboss.php">http://www.sdss3.org/future/eboss.php</a>
→	[I+S] Euclid	2021-2027	<a href="http://sci.esa.int/euclid/">http://sci.esa.int/euclid/</a>
→	[S] DESI	2019-2022	<a href="http://desi.lbl.gov/">http://desi.lbl.gov/</a>
→	[I] J-PAS	2015-2020	<a href="http://j-pas.org/">http://j-pas.org/</a>
	[S] 4MOST	2019-2024	<a href="http://www.4most.eu/">http://www.4most.eu/</a>
	[I] VISTA-VHS	2010-2017	<a href="http://www.vista-vhs.org/">http://www.vista-vhs.org/</a>
	[I] iPTF	2013-2015	<a href="http://ptf.caltech.edu/iptf/">http://ptf.caltech.edu/iptf/</a>
	[I] ZTF	2016-2020	-
	[I] LSST	2023-onwards	<a href="http://www.lsst.org/">http://www.lsst.org/</a>
	[R] LOFAR	2013-2018	<a href="http://www.lofar.org/">http://www.lofar.org/</a>
	[R] Meerkat SKA-Pathfinder	2016-onwards	<a href="http://www.ska.ac.za/meerkat/">http://www.ska.ac.za/meerkat/</a>
	[R] SKA	2019-onwards	<a href="http://www.skatelescope.org/">http://www.skatelescope.org/</a>
	[R] CMB (CORe/PRISM)	Proposal	<a href="http://www.prism-mission.org/">http://www.prism-mission.org/</a>
	[R] PLANCK ✓	2009-2014	<a href="http://sci.esa.int/planck/">http://sci.esa.int/planck/</a>



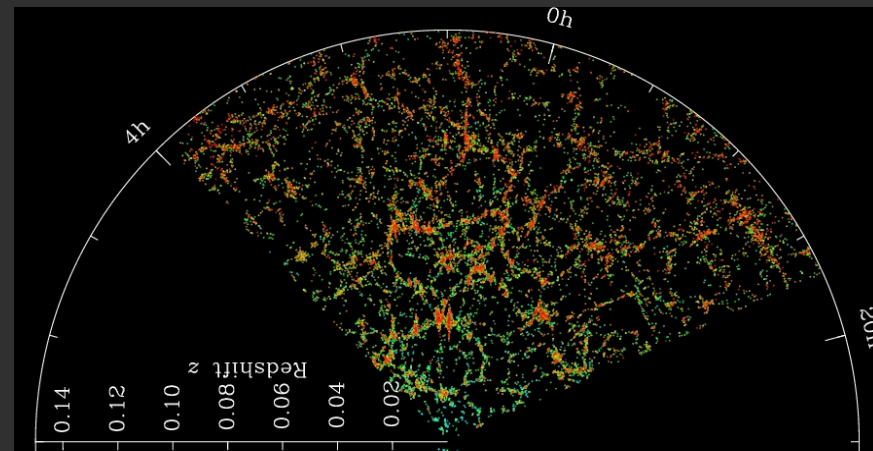
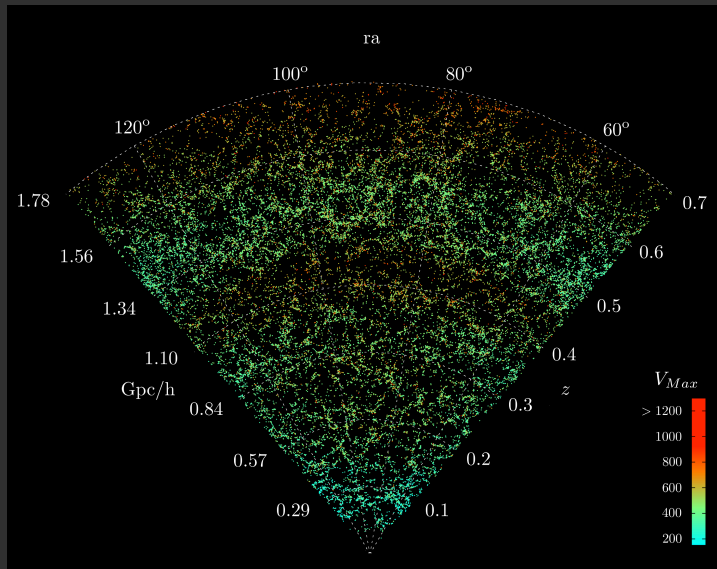
**Dark Energy racing season has started!**



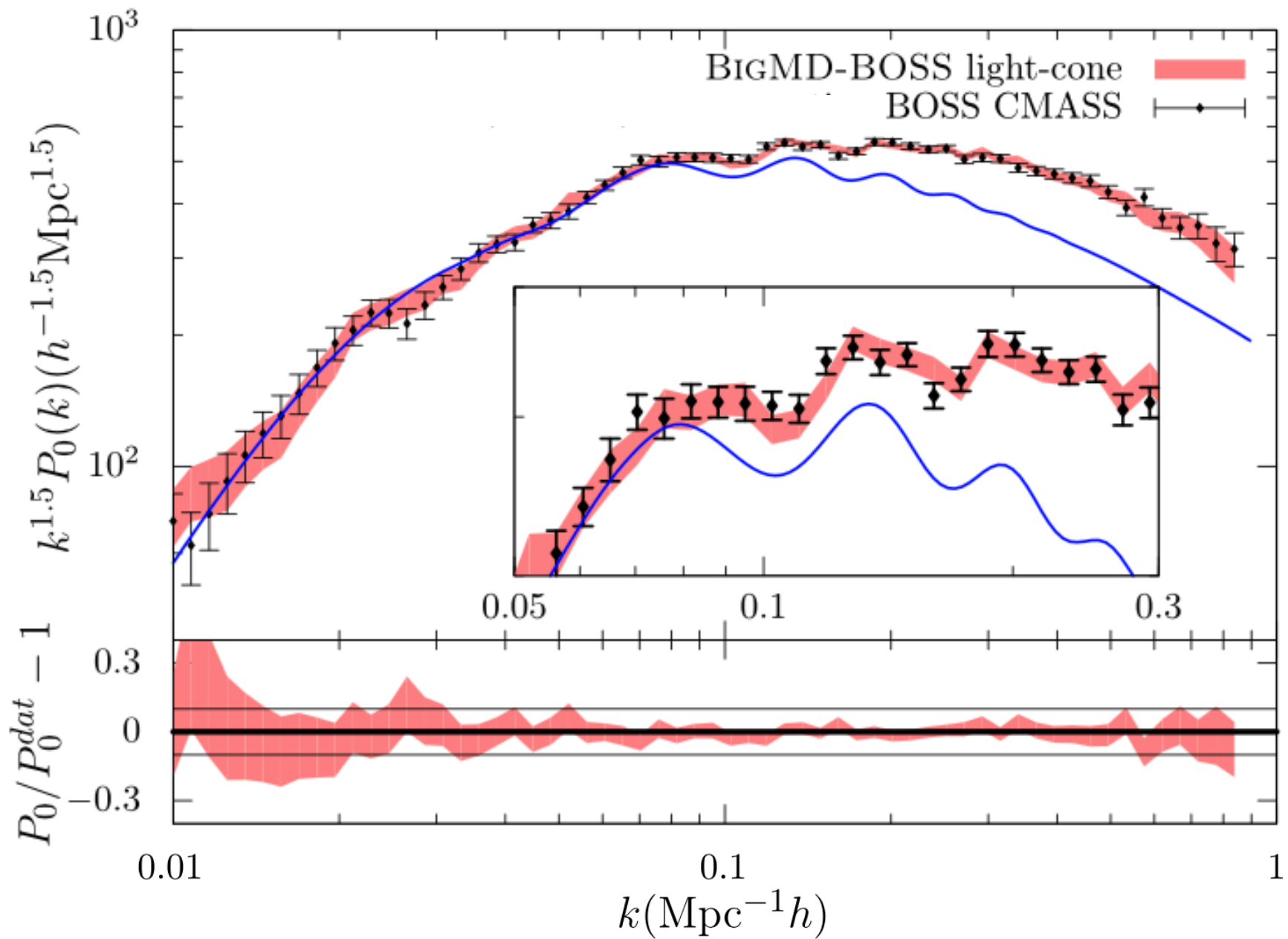
# Large Scale Structure Spectroscopic Surveys



# Simulations calculate consequences of dark physics



We provide cosmological simulations for large surveys  
[www.skiesanduniverses.org](http://www.skiesanduniverses.org)



Power spectrum from the BigMD-BOSS light cone and the BOSS CMASS sample at  $z=0.5$

# The Uchuu Project





# The Uchuu Collaboration

## Collaboration Core Members:

Bruno Altieri, ESAC, **Spain** (\*)

Sofia Cora, Institute of Astrophysics of La Plata, **Argentina** (\*)

Darren Croton, Swinburne University, **Australia** (\*)

Tomoaki Ishiyama, Chiba University, **Japan** (\*)

Eric Jullo, LAM, **France** (\*)

Anatoly Klypin, Virginia University, **USA** (\*)

Ben Metcalf, Bologna University, **Italy**

David Millan, IAA-CSIC, Spain

Francisco Prada, IAA-CSIC, Spain (\*)

Manodeep Sinha, Swinburne University, Australia

Taira Oogi, Kavli IPMU / University of Tokyo, Japan

Cristian Vega-Martínez, University of La Serena, **Chile**

(\*) members of the Collaboration Board (CB);

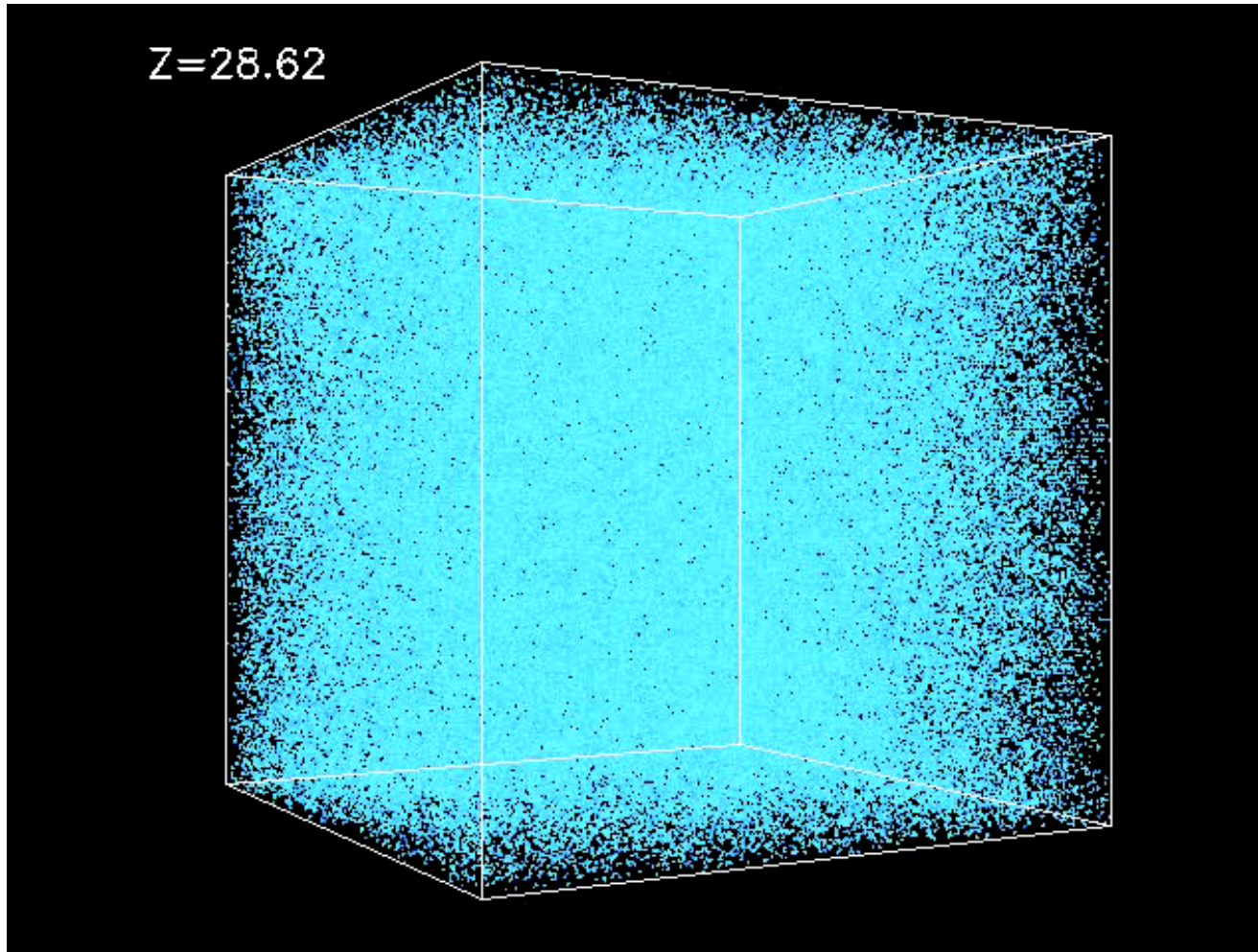
CB Chairs: Tomoaki Ishiyama and Francisco Prada

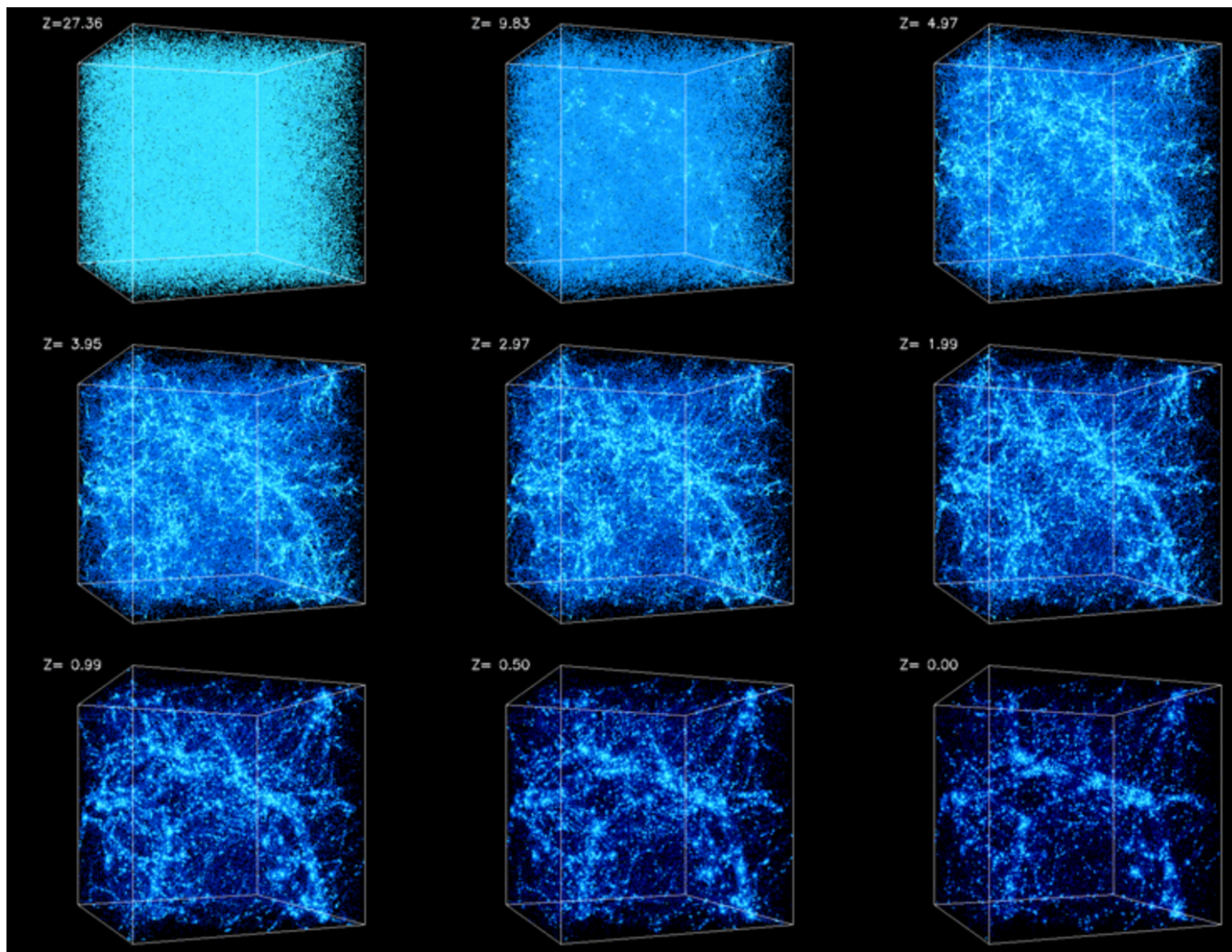
# Our Team Expertise on Computational Cosmology



- 160 years of experience on cosmological simulations and virtual observations:
  - ART, GreeM3, GLAM N-body codes
  - MultiDark, GLAM, v<sup>2</sup>GC suite of simulations (data and products)
  - SAGE & SAG semi-analytical models of galaxy formation
  - Gravitational lensing maps generation
- 60 years of experience building publicly available databases and data platforms for cosmological simulation products and virtual laboratory that host mock observations for massive galaxy survey data (SDSS, DESI, Euclid)
  - MultiDark Simulations Database ([www.cosmosim.org](http://www.cosmosim.org))
  - Skies & Universes ([www.skiesanduniverses.org](http://www.skiesanduniverses.org))
  - Theoretical Astrophysical Observatory (<https://tao.asvo.org.au/tao/>)

# HPC Simulations of the Universe









```
* Number of particles: 12,800^3
* Particle mass: 3.27018e+08 Msun/h
* Box size: 2000.0 Mpc/h
* Force resolution assumed: 0.00427 Mpc/h
* Total number of time steps: 6,000
```

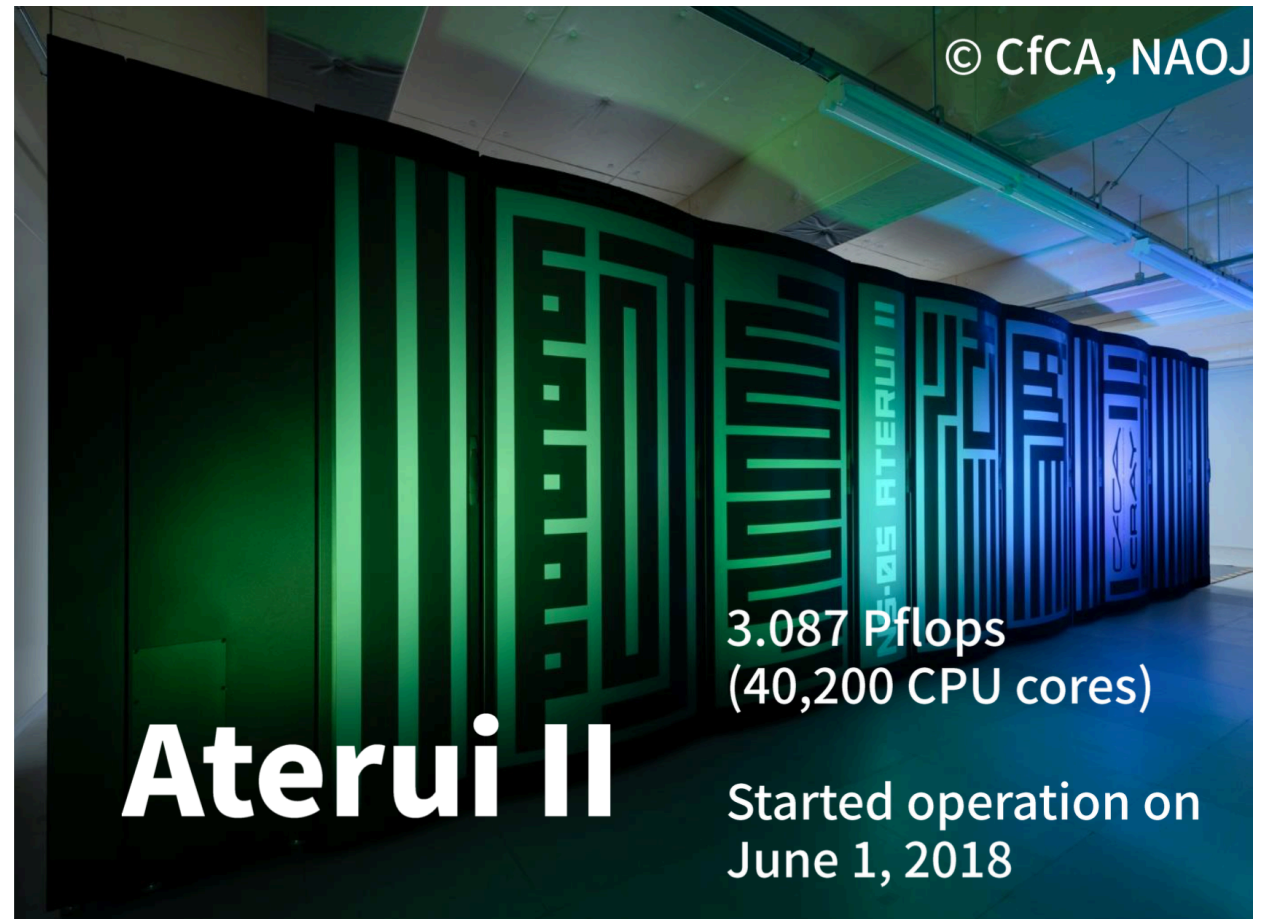
## Uchuu Parameters

**2,097,152,000,000**  
**two trillion particles!**

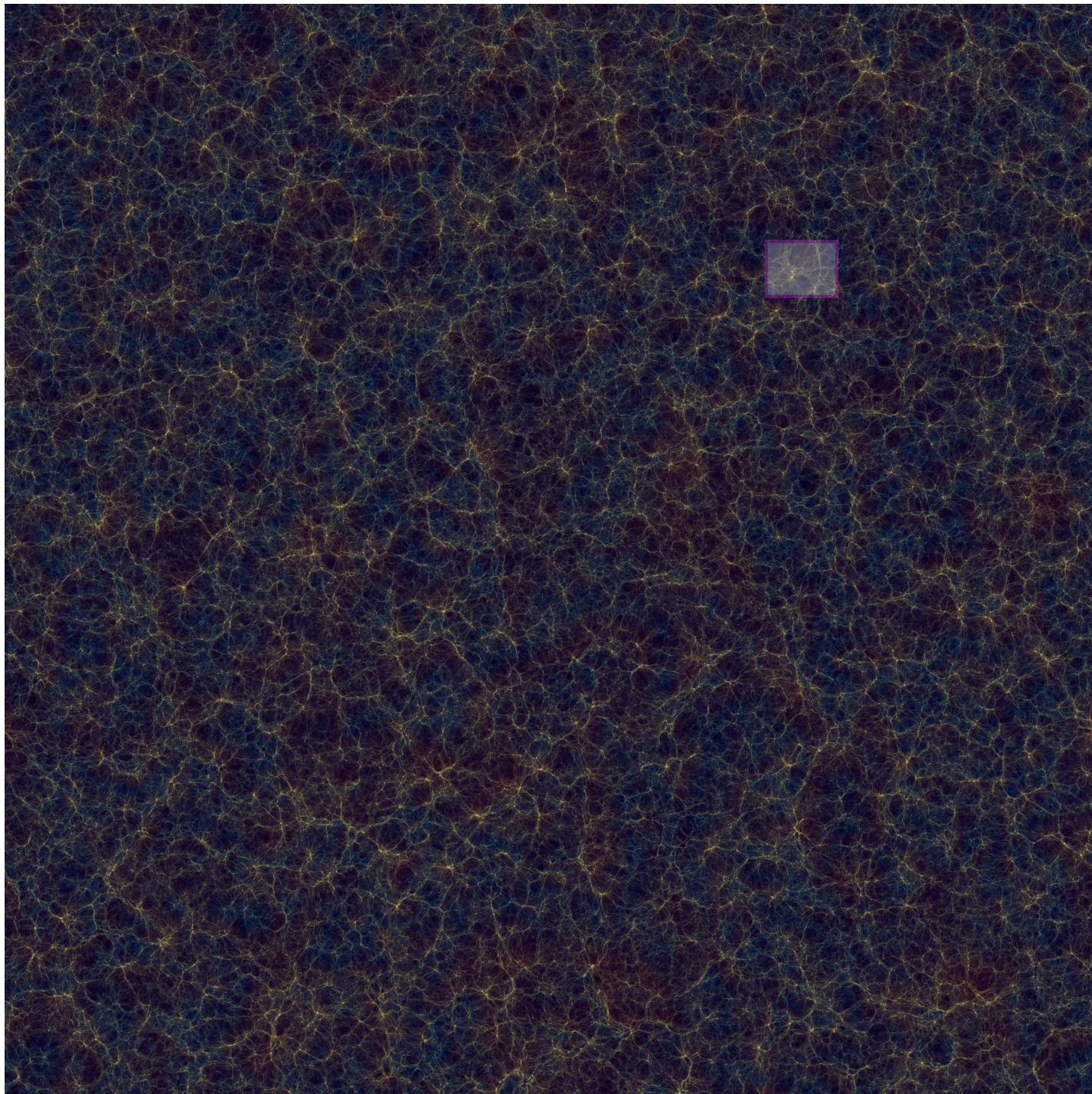
Run with GreeM3 N-body code  
by Tomoaki Ishiyama

**40,000 CPU hours**  
**Exclusive used over 48 hrs / month**  
**July 2018 - October 2019**

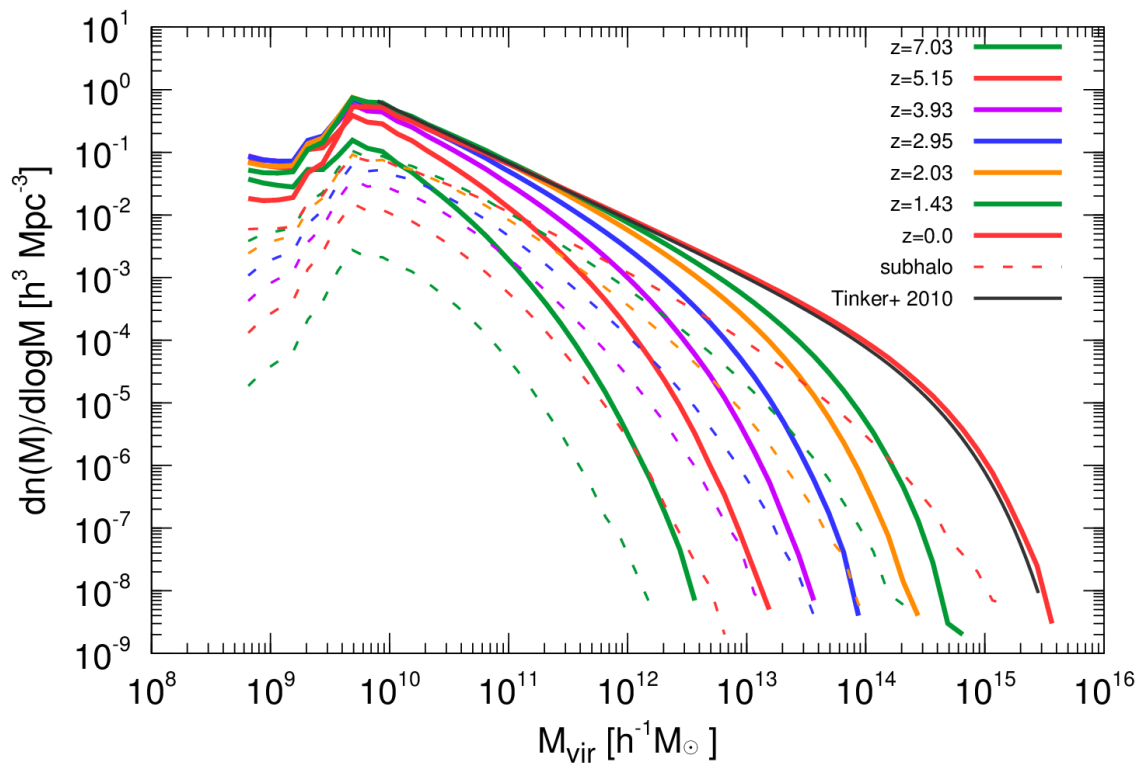
**4 PB of raw data!**



Uchuu is a simulation of the formation of the Large Scale Structure of the Universe based on the mutual gravitational interactions between the dark matter particles, and it can resolve smaller features to depict the Universe in greater detail.

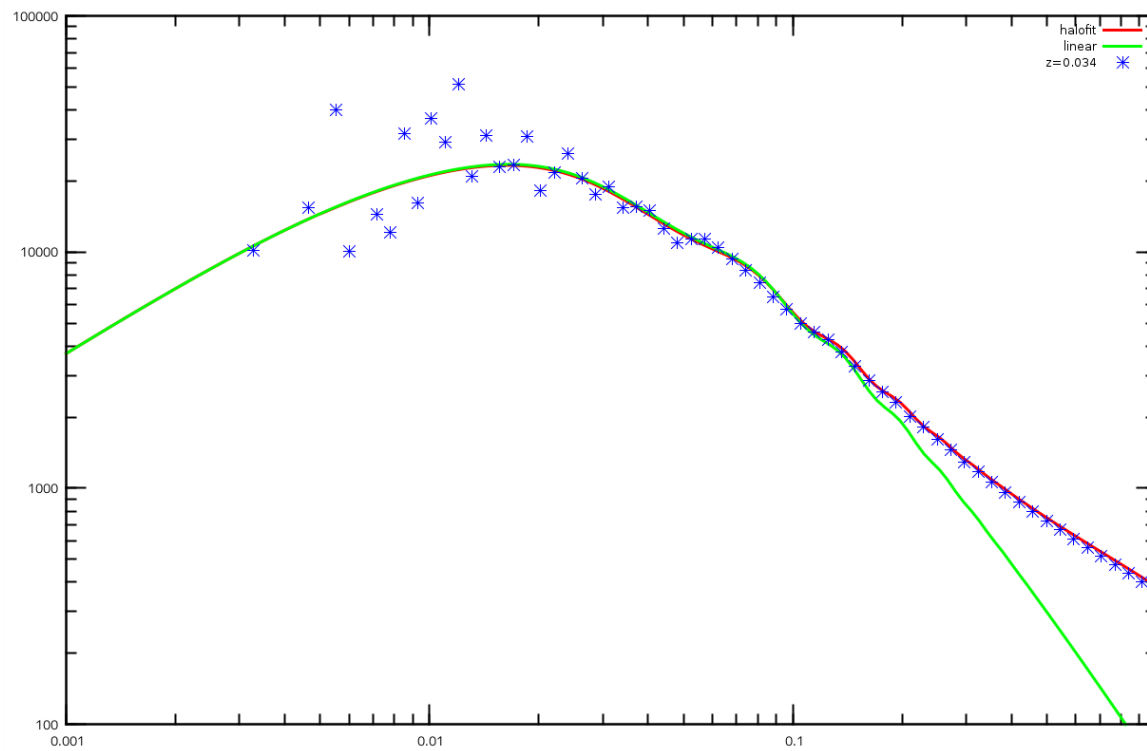


**Uchuu Slice**



**Mass function of dark matter halos at different epochs**

**Power spectrum of matter density at present epoch**





# Uchuu Products on the Cloud

To store and allow analysis with about **300Tb of data:**

- 2 (out of 50) outputs of all particles (128Tb)
- 50 snapshots of halo catalogs (1Tb each)
- 50 snapshots of MergingTrees (1Tb each)
- 50 snapshots of semi-analytical models (1Tb each)
- Gravitational Lensing Maps and Catalogs (20Tb total)

WHAT ARE WE GOING  
TO DO WITH ALL THIS  
FUTURE?

OR -

WHAT IS ALL THIS  
FUTURE GOING TO  
DO WITH US?

# Dissemination of the data to the public: Our operational developments

## MultiDark Project & AIP-Potsdam (German VO)

### CosmoSim

The CosmoSim database provides results from cosmological simulations performed within different projects: [MultiDark and Bolshoi](#), [CLUES](#), and [Galaxies](#).

#### MultiDark Bolshoi

The Spanish MultiDark Consolider project supports efforts to identify and detect matter, including dark matter simulations of the universe.

[MDR1](#) [BigMDPL](#)  
[SMDPL Bolshoi](#)  
[MDPL BolshoiP](#)  
[MDPL2](#)

#### Galaxies

Available now for the MDPL2 simulation - galaxy catalogs contain galaxy properties from different semi-analytical codes.

[MDPL2 Galacticus](#)  
[MDPL2 SAG](#)  
[MDPL2 SAGE](#)

#### CLUES

Constrained Local Universe Simulations

The CLUES project produces constrained simulations of the local universe, partially with gas and star formation.

[Clues3\\_LGDM](#)  
[Clues3\\_LGGas](#)

Please visit the linked sites for more information about the projects and about the appreciated form of acknowledgment, if the data is used in a scientific publication or proposal.

Check out the [Documentation](#) and the [Simulations](#) section for more information or the [CosmoSim blog](#) for latest news, additional materials, tutorials and much more.

#### Database access

The database can be queried by entering SQL statements directly into the [Query Form](#) or via [Scripted access](#). If you haven't done so, please register first via the [Registration Form](#) to get your own private database where the results of your queries will be stored for you. You can also submit queries as a guest, but the result data can then be accessed and removed by any other guest as well.

More information on the simulations and their projects is available in the [Simulations](#) section. Details about the database, its design and the possibilities to access the data are described in the [Documentation](#).

[Register to CosmoSim](#)

AIP

CosmoSim.org is hosted and maintained by the Leibniz-Institute for Astrophysics Potsdam (AIP).

GAVO  
GERMAN ASTROPHYSICAL VIRTUAL OBSERVATORY

It is a contribution to the German Astrophysical Virtual Observatory.

The MultiDark and Bolshoi simulations were run on the NASA's Pleiades supercomputer at the NASA Ames Research Center.

PRACE

The MultiDark-Planck (MDPL) and the BigMD simulation suite have been performed in the Supermuc supercomputer at LRZ using time granted by PRACE.

The CLUES simulations were performed using the supercomputing facilities at LRZ Munich, BSC Barcelona and JSC Jülich.

## SQL queries

**Skies & Universes**   ▾ Home   ▾ Simulations   ▾ Products   Surveys

**The main goal** of the Skies and Universes website is to provide access for the astronomical community with *results from cosmological simulations*. These results are useful on their own to understand processes related with the non-linear evolution of cosmic structures. Some other products are specifically focused on large survey observational projects. To facilitate the interaction between the theory and observations we also provide links to recent observational projects. There are other databases and websites that provide access to results of simulations. Their links can also be found here.

**Specifics** of our portal is that it provides raw data of simulations. For example, one can find here *coordinates and velocities* of simulations with billions of particles. We also provide *halo catalogs*, *light-cones for mock galaxy/qso samples*, *power spectra for many thousands of N-body realizations* and some observational products. While we provide some assistance in the form of codes to read the data, it is responsibility of users to write their own analysis software. Unlike the [CosmoSim.org](#) and [Millennium](#) databases, we do not support SQL queries. Users of *Skies & Universes* are expected to download data and write their own routines to analyze the data. At present we do not provide computing facilities to do the analysis.

**Downloading files:** You are welcome to use more than 30Tb of raw data on this site. However, special care should be given when downloading large files. Please, do not download all of the large files at once. You will not get anything, if you do. Get 1-2 files, wait until the downloading finishes and then get another bunch files.

More info at [arXiv:1711.01453](#)

**Simulations:**

- \* Main Page
- \* GLAM
- \* MultiDark
- \* Bolshoi
- \* **Guest simulations:**
  - \* Lomonosov
  - \* MassiveNuS
  - \* v2GC

**Products:**

- \* Main page
- \* Mock Catalogs

**Other databases and websites:**

- \* CosmoSim
- \* Millennium
- \* Bolshoi
- \* DarkSky
- \* TAO
- \* v2GC
- \* Abacus

**Credits and Registration:**

- \* Credits
- \* Registration

**Alternative Skies & Universes sites:**

- Europe
- USA

If you use our results, acknowledge [www.skiesanduniverses.org](#)   Contact us

Swinburne

Theoretical Astrophysical Observatory

**Queryable Data** from multiple popular cosmological simulations and galaxy formation models **which can be funneled through higher-level modules to build custom mock galaxy catalogues and images.**

DATA QUERY MODULES MOCK

GET STARTED

\* TAO is accessible from anywhere you can access the internet.

Cloud oriented

ACKNOWLEDGEMENTS

TAO is part of the All-Sky Virtual Observatory (ASVO) and is funded and supported by Astronomy Australia Limited, Swinburne University of Technology and



the Australian Government. The latter is provided through the Commonwealth's Education Investment Fund and National Collaborative Research Infrastructure

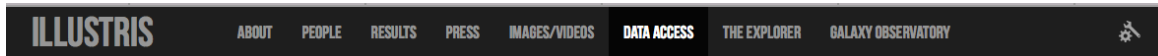


Strategy, particularly the National eResearch Collaboration Tools and Resources (NeCTAR) and the Australian National Data Service Projects.





# Other simulation public platforms



## Public Data Access [Overview](#) / [Web API Documentation](#)

Welcome! You are currently not logged in. In order to access any Illustris[TNG] data, you need to first **Login**.

Don't have an account yet? [New User Registration](#).

## A web-based interface (API)

The web-based interface (API) can respond to a variety of user requests and queries, and can be used in addition to, or in place of, the download and local analysis of large data files. At a high level, the API allows a user to search, extract, visualize, and analyze. In each case, the goal is to reduce the data response size, either by extracting an unmodified subset, or by calculating a derivative quantity.

This page has three sections: [getting started guide](#), [cookbook](#), [reference](#).

We provide examples of accessing the API in a few languages. Select one to show all the content on this page specifically for that language.

- Python
- IDL
- Matlab
- Julia

## API Getting Started Guide

First, start up your interface of choice and define a helper function, whose purpose is to make a request to the API and verify that the response is successful. If the response type is **JSON**, then automatically

```
>>> import requests
>>>
>>> baseUrl = 'http://www.illustris-project.org/api/'
>>> headers = {"api-key": "INSERT_API_KEY_HERE"}
>>>
>>> def get(path, params=None):
>>>     # make HTTP GET request to path
>>>     r = requests.get(path, params=params, headers=headers)
>>>
>>>     # raise exception if response code is not HTTP SUCCESS
>>>     r.raise_for_status()
>>>
>>>     if r.headers['content-type'] == 'application/json':
>>>         return r.json() # parse json responses automatically
>>>     return r
```

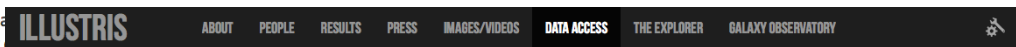
Issue a request to the API root.

```
>>> r = get(baseUrl)
```

The response is a dictionary object with one key, **"simulations"**, which is a list of **N** (c

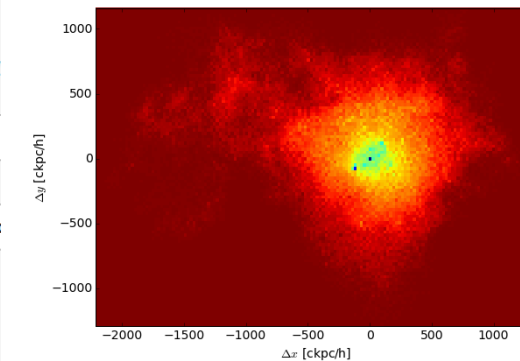
```
>>> r.keys()
['simulations']

>>> len(r['simulations'])
18
```



Make a quick 2d histogram visualization of the distribution of gas bound to this subhalo, weighted by the log of gas cell mass, and with position relative to the subhalo center.

```
>>> with h5py.File(cutout, 'r') as f:
>>>     x = f['PartType0']['Coordinates'][:,0] - sub_prog['pos_x']
>>>     y = f['PartType0']['Coordinates'][:,1] - sub_prog['pos_y']
>>>     dens = np.log10(f['PartType0']['Masses'][:,])
>>>
>>> plt.hist2d(x,y,weights=dens,bins=[150,100])
>>> plt.xlabel('$\Delta x$ [ckpc/h]')
>>> plt.ylabel('$\Delta y$ [ckpc/h]');
```



That's enough! You can also explore the independent examples below.



# Build your own Universe

Interactive data analysis of massive cosmological data without any SQL knowledge



Billions of  
observed and  
simulated  
galaxies



Superfast queries  
means superfast  
results



Features to make  
you work faster  
and easier



Online plotting  
preview and data  
download

Learn more



# How does it work?

Name	Version	Description	Origin	Date
Rockstar Halos	0.1	First version of the rockstar halo catalogue of the first (+,+)+ octant of the light cone in a cube of side length 1 h <sup>-1</sup> Gpc. Coordinate (0,0,0) is z=0.	Simulated	2016-10-11
DES Y1A1 Gold	1.0.3	Gold catalog for DES Y1A1 coadds (science-ready)	Observed	2016-08-19
MICECAT	2	MICE-Grand Challenge (MICE-GC) Galaxy and Halo Light-cone catalog	Simulated	2016-06-25
COSMOS	1	COSMOS photometric redshifts	Observed	2016-05-06
PAUdm FA-Coadd	1.0	PAUdm forced_aperture_coadd table	Observed	2015-11-08
SDSS Spec-Photo	1	SDSS spectroscopic catalog used by PAUdm	Observed	2015-11-08
CFHTLenS	1	CFHTLenS photometry and shear	Observed	2015-08-04
DEEP2	DR4	DEEP2 galaxy redshift survey	Observed	2015-07-30
DES SVA1 Gold	1.0.0	Gold version for the DES-SG1-LSS Data (aka Infrastructure). Some files that are part of the activity of the DES-LSS working group are available too.	Observed	2014-01-28
MICECAT	1	MICE-Grand Challenge (MICE-GC) Galaxy and Halo Light-cone catalog	Simulated	2013-12-09

1. Select a catalog

**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 5.

Name	Version	Type	Description	Rows	Load
Cosmos galaxy magnitudes	1.0	Basic	Cosmos (true) absolute and observed magnitudes	499,609,997	Load Readme
DES galaxy magnitudes	1.0	Basic	DES (true) absolute and observed magnitudes	499,609,997	Load Readme
Euclid galaxy magnitudes	1.0	Basic	Euclid (true) absolute and observed magnitudes	499,609,997	Load Readme
Galaxy & Halo catalog	1.0	Basic	Main galaxy fields from HOD algorithm and halo properties	499,609,997	Load Readme
Galaxy shape	2.0	Basic	Galaxy morphological properties	499,609,997	Load Readme
Galaxy shear	1.0	Basic	Galaxy shear properties	499,609,997	Load Readme
Halo catalog	1.0	Basic	Halo catalog properties	329,004,990	Load Readme

2a. Select one of the prebuilt datasets...

**Step 1: Columns - Select the fields you need**

unique\_gal\_id unique galaxy id  
 unique\_halo\_id unique halo id  
 ra\_gal galaxy right ascension (degrees)  
 dec\_gal galaxy declination (degrees)  
 ra\_gal\_mag galaxy magnified right ascension (degree)  
 dec\_gal\_mag galaxy magnified declination (degree)  
 kappa convergence  
 gamma1 shear  
 gamma2 shear  
 z\_cgal galaxy true redshift  
 cgal galaxy comoving distance (Mpc/h)  
 z\_cgal\_v galaxy observed redshift (including peculiar velocity)  
 cgal\_v galaxy comoving distance corresponding to z\_cgal\_v (Mpc/h)  
 z\_desdm\_mc desdm photometric redshift monte carlo assigned to follow the benchmark N(z)  
 xgal galaxy comoving distance x-axis (Mpc/h)  
 ygal galaxy comoving distance y-axis (Mpc/h)  
 zgal galaxy comoving distance z-axis (Mpc/h)  
 vvcml galaxy radial peculiar velocity in environment (km/s)

2b. Or build your own custom set

**Step 2: Sampling - Select a subset and get faster results**

Size: 1/256  1/1 ~ 7.81 M rows

Seed: 1  Random

**Step 3: Filters - Add conditions to refine your search**

ra_gal	>	0	✕
dec_gal	>	0	✕
dec_gal	<	90	✕
mr_gal	>	-23	✕
gr_gal	>	0	✕

+ Add

3. Select the sampling and filters

CosmoHub portal uses the Apache Hive infrastructure, which provides a very fast data query and is built on top of Hadoop

# Our vision for Uchuu in the Cloud @ EOSC



(NEED YOUR FEEDBACK & HELP!)

- Previous examples (TAO, CosmoHub) provide a convenient interface for “SQL” queries and allow preview the catalogs you are building with online plot generator
- Do we want the users to be able to carry out a certain set of pre-defined actions in the Cloud or do we want them to be able to run arbitrary analysis snippets?
  - Uchuu @ EOSC requires the latter
  - Technologies? Spark, **Vaex** ...
  - Cloud platform with necessary computing and storage resources (RAM is key 2TB, 128 cores, 500TB storage)
  - Legacy: Long live preservation (critical)

## About vaex.io

Power up your business with our data driven solutions. With our unique, state-of-the-art technology, we provide fast and scalable solutions that will make you more agile, while limiting unnecessary resources.



### BE AGILE

Cut development time by 80%. Your prototype is your solution. Create automatic pipelines for any model.



### MAKE IMPACT

Improve your business right away. Deploy your models to Amazon Web Services or Google Cloud Platform with a single command.



### BE RESOURCEFUL

Empower your data scientists. Turn any laptop into a big data powerhouse. No clusters, no engineers.



# What is Vaex?

Vaex is a python library for lazy **Out-of-Core DataFrames** (similar to Pandas), to visualize and explore big tabular datasets. It can calculate *statistics* such as mean, sum, count, standard deviation etc, on an *N-dimensional grid* up to a **billion** ( $10^9$ ) objects/rows **per second**. Visualization is done using **histograms, density plots** and **3d volume rendering**, allowing interactive exploration of big data. Vaex uses memory mapping, a zero memory copy policy, and lazy computations for best performance (no memory wasted).

## Why vaex

- **Performance:** works with huge tabular data, processes  $> 10^9$  rows/second
- **Lazy / Virtual columns:** compute on the fly, without wasting ram
- **Memory efficient** no memory copies when doing filtering/selections/subsets.
- **Visualization:** directly supported, a one-liner is often enough.
- **User friendly API:** you will only need to deal with the DataFrame object, and tab completion + docstring will help you out: `ds.mean<tab>`, feels very similar to Pandas.
- **Lean:** separated into multiple packages
  - `vaex-core`: DataFrame and core algorithms, takes numpy arrays as input columns.
  - `vaex-hdf5`: Provides memory mapped numpy arrays to a DataFrame.
  - `vaex-arrow`: **Arrow** support for cross language data sharing.
  - `vaex-viz`: Visualization based on matplotlib.
  - `vaex-jupyter`: Interactive visualization based on Jupyter widgets / ipywidgets, bqplot, ipyvolum and ipyleaflet.
  - `vaex-astro`: Astronomy related transformations and FITS file support.
  - `vaex-server`: Provides a server to access a DataFrame remotely.
  - `vaex-distributed`: (Proof of concept) combined multiple servers / cluster into a single DataFrame for distributed computations.
  - `vaex-qt`: Program written using Qt GUI.
  - `vaex`: Meta package that installs all of the above.
  - `vaex-ml`: **Machine learning**
- **Jupyter integration:** vaex-jupyter will give you interactive visualization and selection in the Jupyter notebook and Jupyter lab.

<https://vaex.io>

<http://vaex.astro.rug.nl/>

<https://vaex.readthedocs.io/>

# Remarks



- **Uchuu** is coming public on Spring 2020
- Need desperately your HELP!

Grazas!  
Obrigado!