The illustris simulation: Public data release

Citation: Nelson, D. et al. "The Illustris Simulation: Public Data Release." Astronomy and Computing 13 (November 2015): 12–37 © 2015 Elsevier B.V.

As Published: http://dx.doi.org/10.1016/j.ascom.2015.09.003

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/111982

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

The Illustris Simulation: Public Data Release☆

Dylan Nelson[a,*], Annalisa Pillepich[a], Shy Genel[b,a,1], Mark Vogelsberger[c], Volker Springel[d,e], Paul Torrey[c,g], Vicente Rodriguez-Gomez[a], Debora Sijacki[f], Gregory F. Snyder[h], Brendan Griffen[c], Federico Marinacci[c], Laura Blecha[j,2], Laura Sales[i], Dandan Xu[d], Lars Hernquist[a]

[a] *Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138, USA*
[b] *Department of Astronomy, Columbia University, 550 West 120th Street, New York, NY, 10027, USA*
[c] *Kavli Institute for Astrophysics and Space Research, Department of Physics, MIT, Cambridge, MA, 02139, USA*
[d] *Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany*
[e] *Zentrum für Astronomie der Universität Heidelberg, ARI, Mönchhofstr. 12-14, 69120 Heidelberg, Germany*
[f] *Institute of Astronomy and Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[g] *TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA*
[h] *Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218*
[i] *Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA*
[j] *University of Maryland, College Park, Department of Astronomy and Joint Space Science Institute*

Abstract

We present the full public release of all data from the Illustris simulation project. Illustris is a suite of large volume, cosmological hydrodynamical simulations run with the moving-mesh code Arepo and including a comprehensive set of physical models critical for following the formation and evolution of galaxies across cosmic time. Each simulates a volume of $(106.5\ \mathrm{Mpc})^3$ and self-consistently evolves five different types of resolution elements from a starting redshift of $z = 127$ to the present day, $z = 0$. These components are: dark matter particles, gas cells, passive gas tracers, stars and stellar wind particles, and supermassive black holes. This data release includes the snapshots at all 136 available redshifts, halo and subhalo catalogs at each snapshot, and two distinct merger trees. Six primary realizations of the Illustris volume are released, including the flagship Illustris-1 run. These include three resolution levels with the fiducial "full" baryonic physics model, and a dark matter only analog for each. In addition, we provide four distinct, high time resolution, smaller volume "subboxes". The total data volume is ∼265 TB, including ∼800 full volume snapshots and ∼30,000 subbox snapshots. We describe the released data products as well as tools we have developed for their analysis. All data may be directly downloaded in its native HDF5 format. Additionally, we release a comprehensive, web-based API which allows programmatic access to search and data processing tasks. In both cases we provide example scripts and a getting-started guide in several languages: currently, IDL, Python, and Matlab. This paper addresses scientific issues relevant for the interpretation of the simulations, serves as a pointer to published and on-line documentation of the project, describes planned future additional data releases, and discusses technical aspects of the release.

Keywords: methods: data analysis, methods: numerical, galaxies: formation, galaxies: evolution, data management systems, data access methods

1. Introduction

Our theoretical understanding of the origin and evolution of cosmic structure throughout the universe is increasingly propelled forward by large, numerical simulations. From humble beginnings (e.g. Press and Schechter, 1974; Davis et al., 1985), dark matter only N-body simulations of pure gravitational dynamics have reached a state of maturity and extreme scale (e.g. Kim et al., 2011; Skillman et al., 2014). They form a foundation in our understanding of the ΛCDM cosmological model, including the nature of both dark matter and dark energy. Yet, such DM-only simulations have a fundamental limitation – they cannot provide any direct predictions for baryonic components of the universe: gas, stars, and black holes. While dark matter halo collapse forms the back bone of structure formation, the majority of observational astronomy is based on the properties of the baryons.

The natural successor to dark matter only N-body simulations are cosmological hydrodynamical simulations (e.g. Katz et al., 1992), which model the coupled evolution of dark matter and cosmic gas. Hydrodynamical simulations can also account for diverse phenomena such as the formation of stars, the growth of supermassive black holes, the energetic feedback processes arising from both populations, the production and distribution of heavy elements,

☆ Permanently available at www.illustris-project.org/data
* Corresponding author
Email address: dnelson@cfa.harvard.edu (Dylan Nelson)
[1] Hubble Fellow
[2] Einstein Fellow

and so forth. Modern efforts are now able to capture cosmological scales of $\gtrsim$100 Mpc, while simultaneously resolving the internal structure of individual galaxies at $\lesssim$1 kpc scales (Horizon-AGN: Dubois et al. 2014, MassiveBlack-II: Khandai et al. 2014, Illustris: Vogelsberger et al. 2014a, EAGLE: Schaye et al. 2015). These simulations yield verifiable predictions or models for a wide range of interesting astrophysical problems including the spin alignment of galaxies on large scales (e.g. Hahn et al., 2010), the distribution of neutral hydrogen (e.g. Bird et al., 2014; Rahmati et al., 2015), or the impact of baryons on the structure of dark matter haloes (e.g. Schaller et al., 2014).

Observational data focused on the large-scale structure of the universe and the properties of galaxies across cosmic time also continue to increase. Surveys such as SDSS (York et al., 2000), DEEP2 (Davis et al., 2003), CANDELS (Grogin et al., 2011), and 3D-HST (Brammer et al., 2012) provide local and high redshift measurements of the statistical properties of galaxy populations. Future instruments such as LSST (LSST Science Collaboration et al., 2009) and surveys such as DES (The Dark Energy Survey Collaboration, 2005) will provide increasingly precise observational constraints for theoretical models.

To confront theory and observation, the public dissemination of data from both sides is crucial. Efforts based on the availability of ubiquitous international networks began with the highly successful SDSS SkyServer (Szalay et al., 2000, 2002a), which addressed the problems of how remote users could mine data from large datasets (Gray et al., 2002; Szalay et al., 2002b). The approach, which continues to this day, is based on user written SQL queries executed against a large relational database system – query responses can be thought of as both search results and data extraction. Simple queries with near-instantaneous return, as well as long, queued job queries with results saved into temporary storage are supported.

The Millennium simulation (Springel et al., 2005b) public data release was the first large effort from the theoretical side. Modeled on the SDSS approach, the primary data products were stored in a relational database, which users could search and extract data from using raw SQL queries (Lemson and Virgo Consortium, 2006). The focus is on the halo and subhalo catalogs, their merger trees, and various post-processed galaxy property catalogs computed with semi-analytical models. It has been continually extended with additional simulations, data products, and capabilities. The Millennium-II simulation (Boylan-Kolchin et al., 2009; Guo et al., 2011) was included, and the idea of the "virtual observatory" (VO) was realized with Overzier et al. (2013). These efforts have occasionally implemented ideas for incorporating theory within the existing VO framework (Lemson and Zuther, 2009; Lemson et al., 2014). More generally, the Theoretical Astrophysical Observatory (TAO Bernyk et al., 2014) was also targeted at providing mock observations of simulated galaxy and galaxy survey data.

Other dark matter only simulations have adopted similar approaches. The Bolshoi and MultiDark simulations (Klypin et al., 2011) were released under a common database (Riebe et al., 2013), now called CosmoSim. The Dark Energy Universe Simulation (DEUS Rasera et al., 2010) data is available online, as are some data from the MICE simulations (Crocce et al., 2010) through the CosmoHub database. In contrast, the MassiveBlack-II (hydrodynamical) simulation (Khandai et al., 2014) made group catalogs available for direct download. Most recently, the Dark Sky simulation has likewise avoided the database and SQL query framework in favor of direct web access to binary data (Skillman et al., 2014).

In releasing the Illustris simulation data, we adopt a similar approach, offering direct online access to all snapshot, group catalog, merger tree, and supplementary data catalog files. In addition, we develop a web-based API which allows users to perform many common tasks without the need to download any full data files. These include searching over the group catalogs, extracting particle data from the snapshots, accessing individual merger trees, and requesting visualization and further data analysis functions. Extensive documentation and programmatic examples (in IDL, Python, and Matlab) are provided.

This paper is intended primarily as a guide for users of the Illustris simulation data. In Section 2 we give an overview of the simulations. Section 3 describes the data products, and Section 4 discusses methods for data access. Section 5 describes technical details related to the architecture and implementation of the data release itself. In Section 6 we present some scientific remarks and cautions for Illustris, while in Section 7 we discuss community considerations including citation. In Section 8 we summarize. Appendices A through C provide descriptions of all relevant data fields, while Appendix D presents several code examples for the API.

2. Description of the Simulations

The Illustris Project is a series of hydrodynamical simulations of a $(106.5\ \mathrm{Mpc})^3$ cosmological volume that follow the evolution of dark matter, cosmic gas, stars, and super massive black holes from a starting redshift of $z = 127$ to the present day, $z = 0$. It includes three runs at increasing resolution levels, Illustris-(1,2,3), where Illustris-1 is the flagship, highest-resolution box. Each has been simulated including a fiducial "full" baryonic physics model, as well as a dark-matter only analog, Illustris-(1,2,3)-Dark. Vogelsberger et al. (2014a,b); Genel et al. (2014); Sijacki et al. (2014) have presented the Illustris simulations and their galaxy and black hole populations, both at $z = 0$ as well as at higher redshifts. In what follows, we summarize the most relevant features.

In Table 1 we provide an overview of the specifications of the six Illustris runs, including the computational volume, gravitational softening lengths, and masses of the different particle/cell types, which collectively indicate the resolution and dynamic range achieved. To emphasize

Table 1: The most important numerical parameters for the six full volume runs. Gravitational softenings for all particle types other than DM are comoving kpc (with value equal to that of the DM) until $z = 1$ after which they are fixed to their $z = 1$ values, such that at $z = 0$ they have half the softening length as the DM. $m_{\rm baryon}$ is the "target gas mass" (i.e. only the mean mass). The number of gas cells equals the $N_{\rm GAS}$ value only in the initial conditions, the number will then drop as stars and black holes form. Moreover, the total number of baryonic particles (gas cells + star particles + wind particles + black holes) is also not conserved since gas cells can be refined/de-refined to keep their mass within a factor of 2 around $m_{\rm baryon}$. In contrast, the total number of tracers and dark matter particles are both conserved for the duration of the simulation.

Run Name	Alt. Name	Volume [Mpc^3]	$L_{\rm box}$ [Mpc/h]	$N_{\rm GAS}$	$N_{\rm TR}$	$N_{\rm DM}$	$\epsilon_{\rm baryon}$ [kpc]	ϵ_{DM} [kpc]	m_{baryon} [M$_\odot$]	m_{DM} [M$_\odot$]
Illustris-1	L75n1820FP	106.5^3	75	1820^3	1820^3	1820^3	0.7	1.4	1.6×10^6	6.3×10^6
Illustris-2	L75n910FP	106.5^3	75	910^3	910^3	910^3	1.4	2.8	1.0×10^7	5.0×10^7
Illustris-3	L75n455FP	106.5^3	75	455^3	455^3	455^3	2.8	5.7	8.0×10^8	4.0×10^8
Illustris-1-Dark	L75n1820DM	106.5^3	75	0	0	1820^3	-	1.4	-	7.6×10^6
Illustris-2-Dark	L75n910DM	106.5^3	75	0	0	910^3	-	2.8	-	6.0×10^7
Illustris-3-Dark	L75n455DM	106.5^3	75	0	0	455^3	-	5.7	-	4.8×10^8

the variety of galaxy formation and evolution phenomena which can be addressed with the Illustris simulations, in Figure 1 we give the approximate number of a selection of interesting astrophysical objects that can be found in the simulated box, from dark-matter dominated halos at $z = 0$ to luminous active galactic nuclei (AGN) at higher redshifts.

A series of analyses based on the Illustris suite have already been performed. These include 1) comparisons to observations and studies of the impact of different feedback models on the distribution and content of gas on large scales, within halos and in the circumgalactic regime (Bird et al., 2014, 2015; Nelson et al., 2015; Suresh et al., 2015; Bogdan et al., 2015); 2) characterizations of the properties of galactic stellar halos (Pillepich et al., 2014), of the satellite populations across host masses (Sales et al., 2015), of the star formation histories (Sparre et al., 2015) and of the morphologies and angular-momentum build up of Illustris galaxies (Torrey et al., 2015; Snyder et al., 2015; Genel et al., 2015); 3) applications of shock finder algorithms (Schaal and Springel, 2015); 4) analyses on the formation of massive, compact galaxies at high redshifts (Wellons et al., 2015); 5) quantification of the galaxy merger rates (Rodriguez-Gomez et al., 2015), and 6) applications of post-processing radiative transfer algorithms in the study of cosmic reionization (Bauer et al., 2015).

2.1. Physical Models and Numerical Methods

All of the "full physics" Illustris runs contain the following physical components: (1) Primordial and metal-line radiative cooling in the presence of a redshift-dependent, spatially uniform, ionizing UV background field, with self-shielding corrections. (2) Stochastic star formation in dense gas. (3) Pressurization of the ISM due to unresolved supernovae using an effective equation of state model of a two-phase medium. (4) Stellar evolution with the associated mass loss (gas recycling) and chemical enrichment, taking into account SN Ia/II and AGB stars. (5) Galactic-scale outflows with an energy-driven, kinetic wind scheme. (6) Seeding and growth of supermassive black holes. (7) Feedback from AGN in both quasar and radio (bubble) modes, as well as modifications to the cooling curve of nearby gas due to radiation proximity effects. For complete details on the behavior, implementation, parameter selection, and validation of these physical models, see Vogelsberger et al. (2013), which describes the feedback models, and Torrey et al. (2014), which compares the model output with observations from $z = 0$ to $z = 3$.

The Illustris simulations employ the Arepo code (Springel, 2010) which evolves the equations of continuum hydrodynamics coupled with self-gravity. The spatial discretization of the fluid is provided by an unstructured, moving, Voronoi tessellation. On the volumes defined by individual cells Godunov's method is employed, with a directionally unsplit MUSCL-Hancock scheme and an exact Riemann solver. The Voronoi mesh is generated from a set of control points which move with the local fluid velocity modulo mesh regularization corrections. Gravitational forces are computed using the Tree-PM approach, with long-range forces calculated with a Fourier particle-mesh method, and short-range forces with a hierarchical tree algorithm. The code is second order in space, and with hierarchical adaptive time-stepping, also second order in time. During the simulation we employ the Monte Carlo tracer particle scheme (Genel et al., 2013) to follow the Lagrangian evolution of baryons.

In terms of both physical models and numerical methods, the Illustris simulations rely on a substantial foundation of previous work. In Figure 2, we provide an abridged reference tree covering both the physical models and numerical methods. The papers along any given branch are

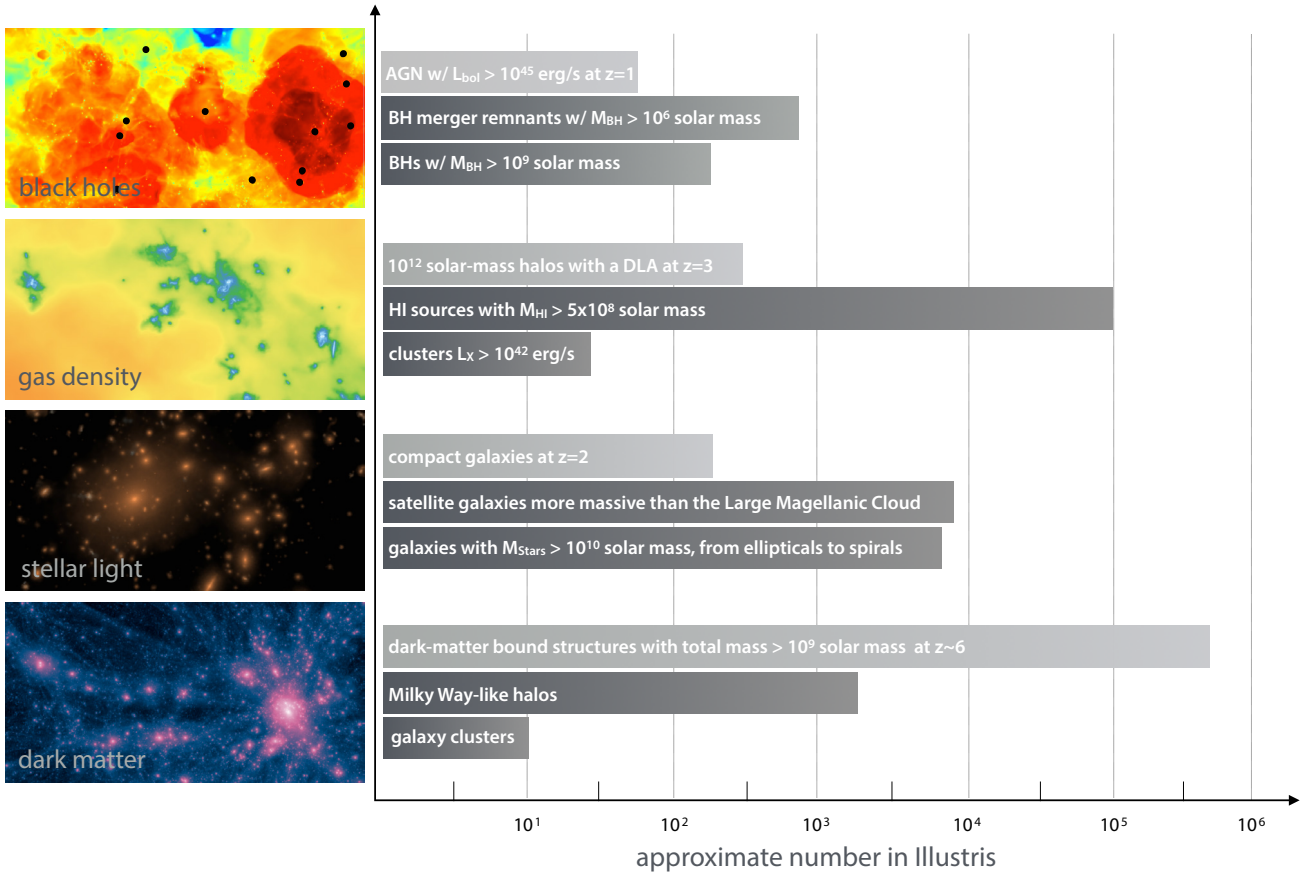

Figure 1: Overview of the variety of galaxy formation and evolution phenomena accessible in the Illustris simulations. A few classes of interesting objects are listed for each of the four mass components present in the simulation: dark matter, stars, gas, and black holes. These are visualized on the left column, for different volumes and spatial scales, as dark-matter density, stellar light, gas density and gas temperature maps, with black holes denoted as black dots. The approximate number present in the Illustris-1 volume is given (from bottom to top), for a) galaxy clusters at $z = 0$ with total mass $M_{200c} > 10^{14}\mathrm{M}_{\odot}$; b) Milky Way-like halos at $z = 0$ ($6 \times 10^{11} < M_{200c} < 2 \times 10^{12}\mathrm{M}_{\odot}$); c) gravitationally-bound objects (dark or luminous) resolved with more than a thousand particles at the end of the reionization epoch; d) galaxies at $z = 0$ with stellar mass exceeding $10^{10}\mathrm{M}_{\odot}$, including both centrals and satellites, from elliptical to disk morphologies; e) satellite galaxies at $z = 0$ more massive than the Large Magellanic Cloud (stellar mass $> 1.5 \times 10^{9}\mathrm{M}_{\odot}$), in any mass host; f) massive, compact galaxies at $z = 2$ according to the selection of Barro et al. (2013); g) clusters of galaxies at $z = 0$ emitting in the X-rays with luminosity exceeding 10^{42} erg/s; h) sources at $z = 0$ with neutral hydrogen mass exceeding $5 \times 10^{8}\mathrm{M}_{\odot}$; i) $10^{12}\mathrm{M}_{\odot}$ halos at $z = 3$ with at least a damped Lyman-alpha system (HI column density $> 10^{20.3}\mathrm{cm}^{-2}$) within 50kpc; j) black holes at $z = 0$ more massive than $10^{9}\mathrm{M}_{\odot}$; k) black-hole merger remnants at $z = 0$, i.e. sub grid black-hole binaries with $M_{\mathrm{BH}} > 10^{6}\mathrm{M}_{\odot}$ for each BH and 1 Gyr delay between the simulation BH merger time and the actual BH merger; l) AGNs at $z = 1$ with bolometric luminosity greater than 10^{45} erg/s.

essential for understanding the details and limitations of the data released here.

3. Data Products

In this data release we give public access to all 136 snapshots between redshift $z = 40$ and redshift zero of the Illustris cosmological volume. This is a periodic box of 106.5 Mpc per side, including up to five types of resolution elements (dark matter particles, gas cells, gas tracers, stellar and stellar wind particles, and black hole sinks). The same volume is available at high (Illustris-1), intermediate (Illustris-2), and low (Illustris-3) resolution. For each resolution, realizations exist with our fiducial, full physics models ("Illustris"), as well as dark matter only analogs ("Illustris Dark"). For all six runs, at every snapshot, two types of group catalogs are provided: friends-of-friends (FoF) halo catalogs, and SUBFIND subhalo catalogs. In postprocessing, these catalogs are used to generate two distinct merger trees, which are both released: SUBLINK, and LHALOTREE. Finally, supplementary data catalogs are released for selected snapshots and runs. At present, these are focused on the stellar properties of Illustris-1 galaxies at $z = 0$, and include mock multi-band images, photometric non-parametric morphological estimates, circularities, angular momenta, and axis ratio measurements. All these data types are described below (snapshots, group catalogs, merger trees, and supplementary catalogs). In the near future we plan to release ROCKSTAR group catalogs and the associated CONSISTENT-TREES merger histories,

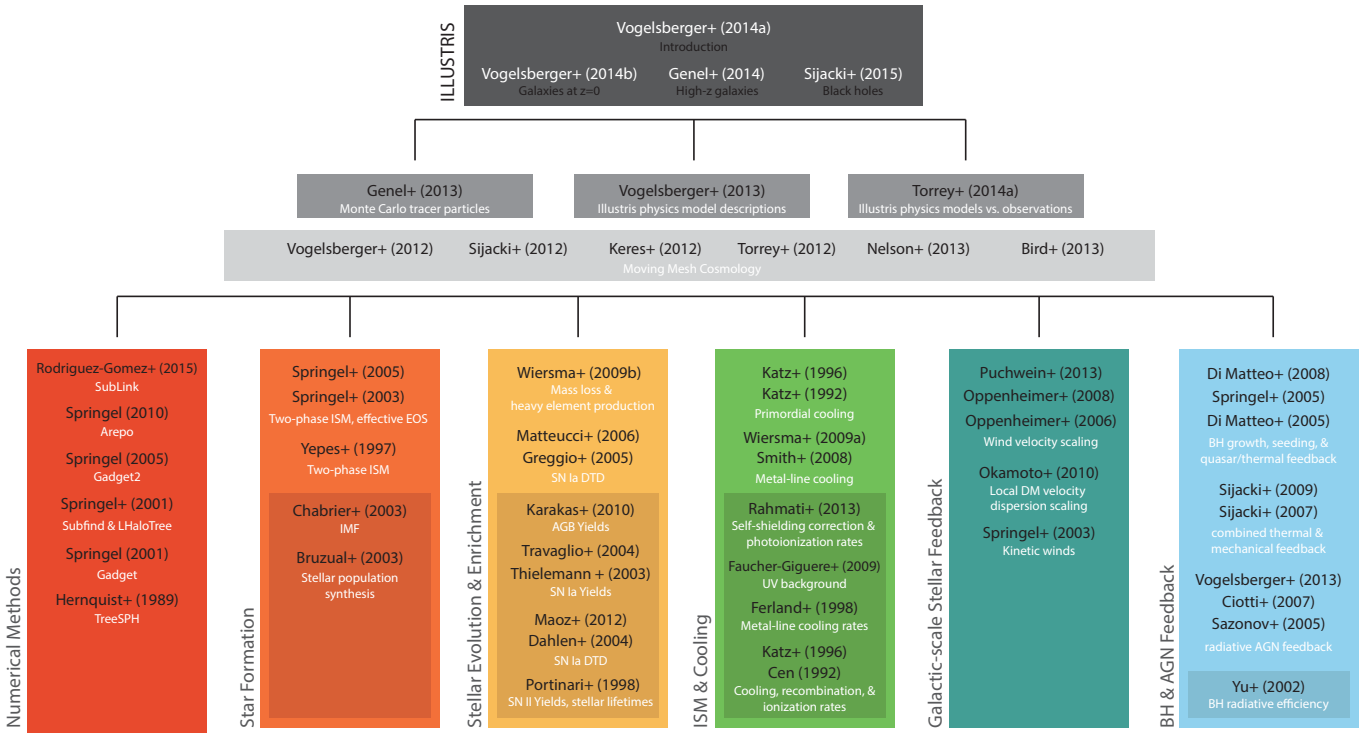

Figure 2: Reference tree for the major components of Illustris, including both numerical methods and physical models. Each paper links to its arXiv or ADS entry. We generally include both models and methods which were directly implemented in Illustris, while entries in the dark subboxes indicate model data inputs. The references are, for the second row: Genel et al. (2013); Vogelsberger et al. (2013); Torrey et al. (2014). The moving mesh cosmology series: Vogelsberger et al. (2012); Sijacki et al. (2012); Kereš et al. (2012); Torrey et al. (2012); Nelson et al. (2013); Bird et al. (2013). Numerical methods: Rodriguez-Gomez et al. (2015); Springel (2010, 2005); Springel et al. (2001a,b); Hernquist and Katz (1989). Star formation: Springel et al. (2005a); Springel and Hernquist (2003); Yepes et al. (1997); Chabrier (2003); Bruzual and Charlot (2003). Stellar evolution and enrichment: Wiersma et al. (2009b); Matteucci et al. (2006); Greggio (2005); Karakas (2010); Travaglio et al. (2004); Thielemann et al. (2003); Maoz et al. (2012); Dahlen et al. (2004); Portinari et al. (1998). ISM and cooling: Katz et al. (1992, 1996); Wiersma et al. (2009a); Smith et al. (2008); Rahmati et al. (2013); Faucher-Giguère et al. (2009); Ferland et al. (1998); Katz et al. (1996); Cen (1992). Galactic-scale stellar feedback: Puchwein and Springel (2013); Oppenheimer and Davé (2008, 2006); Okamoto et al. (2010); Springel and Hernquist (2003). BH and AGN feedback: Di Matteo et al. (2008); Springel et al. (2005a); Di Matteo et al. (2005); Sijacki et al. (2007, 2009); Vogelsberger et al. (2013); Ciotti and Ostriker (2007); Sazonov et al. (2005); Yu and Tremaine (2002).

together with expanded and new supplementary catalogs, with corresponding documentation.

3.1. Snapshots

3.1.1. Snapshot Organization

There are 136 snapshots stored for every run. These include all particles/cells in the whole volume. The full snapshot listings, spacings and redshifts can be found online. A partial listing is provided in Table 2. Every snapshot is stored in a series of "chunks", i.e. more manageable, smaller-size files. The number of chunks per snapshots is different for the different runs, and is given in Table A.1.

The snapshot data is **not** organized according to spatial position. Rather, particles within the snapshot files are sorted according to their group/subgroup memberships, according to the FoF or SUBFIND algorithms. Within each particle type, the sort order is: GroupNumber, SubgroupNumber, BindingEnergy, where particles belonging to the group but not to any of its subgroups ("fuzz") are included after the last subgroup. Figure 3 provides a schematic view of the particle organization within a snapshot, for *one particle type*. The truncation of a snapshot in chunks is arbi-

Table 2: Abridged snapshot list for all six runs. The output times correspond to the set of 128 output redshifts used by the Aquarius project (Springel et al., 2008), augmented by 8 additional saves at integer redshifts.

Snapshot	Scale factor	Redshift
0	0.020932	46.773
32	0.090937	9.9966
45	0.14264	6.0108
54	0.19968	4.0079
60	0.24949	3.0081
68	0.33311	2.002
85	0.50068	0.9973
135	1.0	0.0

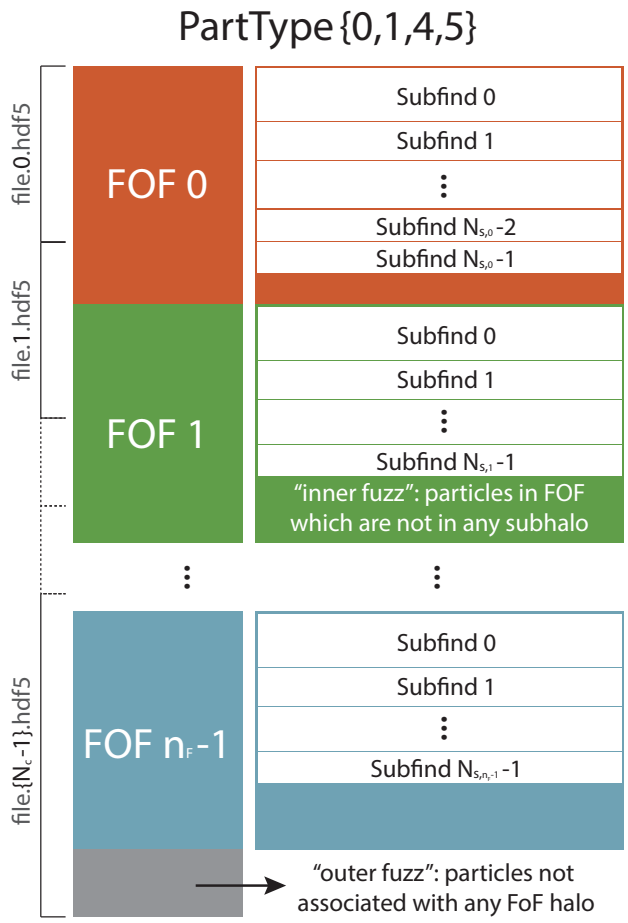

Figure 3: Schematic diagram of the organization of particle/cell data within the snapshots for a single particle type. Within a type, particle order is determined by a global sort of the following fields in this order: FoF group number, SUBFIND subhalo number, binding energy, nearest FoF group number. This implies that FOF halos are contiguous, although they can span file chunks. SUBFIND subhalos are only contiguous within a single group, being separated between groups by an "inner fuzz" of all FOF particles not bound to any subhalo. Here N_c indicates the number of file chunks, n_F the number of FOF groups, and $N_{S,j}$ the number of subhalos in j^{th} FoF group.

trary, thus halos may happen to be stored across multiple, subsequent chunks. Similarly, the different particle types of a halo can be stored in different sets of chunks.

3.1.2. Snapshot Contents

Every HDF5 snapshot contains a "Header" and 5 additional "PartTypeX" groups, for the following particle types (the DM only runs have a single PartType1 group):

- PartType0 - GAS
- PartType1 - DM
- PartType2 - (unused)
- PartType3 - TRACERS
- PartType4 - STARS & WIND PARTICLES
- PartType5 - BLACK HOLES

The most important fields of the header are given in Table A.2. The complete snapshot field listings, including dimensions, units and descriptions, are given for gas in Table A.4, dark matter in A.5, tracers in A.6, stars in A.7, and black holes in A.8.

The general unit system is kpc/h for lengths, $10^{10}\mathrm{M}_\odot/h$ for masses, km/s for velocities. The frequently occurring $(10^{10}\mathrm{M}_\odot/h)/(0.978\mathrm{Gyr}/h)$ represents mass-over-time in this unit system, and multiplying by 10.22 converts to $\mathrm{M}_\odot$/yr. Comoving quantities can be converted in the corresponding physical ones by multiplying for the appropriate power of the scale factor a. For instance, to convert a length in physical units it is sufficient to multiply it by a, volumes need a factor a^3, densities a^{-3} and so on. Note that at redshift $z = 0$ the scale factor is $a = 1$, so that the numerical values of comoving quantities are the same as their physical counterparts.

3.1.3. Tracer Quantities

Each Monte Carlo tracer particle stores 13 auxiliary values. These are updated every timestep where the tracer parent is active. Many are reset to zero immediately after they are written out to a snapshot, such that their recording duration is precisely the time interval between two successive snapshots. Some are only relevant when the tracer resides within a parent of a specific particle type (e.g. gas or star). Table A.9 describes these fields. As the simulations evolve, tracers are exchanged (and can therefore change their parents) in the following ways:

- Gas -> Gas (finite volume fluxes, refinement, derefinement)
- Gas -> Stars (star formation, both spawning new stars and converting cells into stars)
- Stars -> Gas (stellar mass return)
- Gas -> Wind (galactic scale stellar winds)
- Wind -> Gas (recoupling stellar wind)
- Gas -> BHs (black hole accretion)
- BHs -> BHs (black hole mergers)

3.1.4. Subboxes

Four separate "subbox" cutouts exist, for each full physics run. These are spatial cutouts of fixed comoving size and fixed comoving coordinates. They are output at each highest timestep, that is, their time resolution is significantly better than that of the main snapshots – see Table 3. This can be particularly useful for certain types of analysis or particular science questions, or for time evolving visualizations. We point out two notes of caution: first, the time spacing of the subboxes is not uniform in scale factor or redshift, but scales with the time integration hierarchy of the simulation, and is thus variable, with some discrete factor of two jumps at several points during the simulations. Second, the subboxes, unlike the full box, are not periodic.

Table 3: Details of the subbox snapshots. For each resolution level, from lowest to highest, the total number of subbox snapshots saved $N_{\rm snap}$. Each of the four subboxes has the same number of snapshots. The number of file pieces per snapshot N_c, and the approximate time resolution Δt at three redshifts: $z = 6$, $z = 2$, and $z = 0$.

Run	$N_{\rm snap}$	N_c	$\Delta t_{(z=6)}$	$\Delta t_{(z=2)}$	$\Delta t_{(z=0)}$
Illustris-3	1426	1	∼7 Myr	∼12 Myr	∼33 Myr
Illustris-2	2265	16	∼4 Myr	∼6 Myr	∼17 Myr
Illustris-1	3976	512	∼2 Myr	∼3 Myr	∼8 Myr

The four subboxes sample four different areas of the large box, roughly described by the environment column in Table A.3. The particle fields are all identical to the main snapshots. However, the ordering differs. In particular, particles/cells in the subboxes are not ordered according to their group membership, as no group catalogs are available for these cutouts.

3.2. Group Catalogs

There is one group catalog associated with each snapshot, which includes both FoF and SUBFIND objects. The group files are split into a small number of sub-files, just as with the raw snapshots. Every group catalog file contains the following HDF5 groups: Header, Group, Subhalo, Offsets. The IDs of the members of each group/subgroup are not stored in the group catalog files. Rather, particles/cells in the snapshot files are ordered according to group membership. Each group contains its total length, allowing IDs and all other fields of member particles/cells to be accessed using an offset table type approach. This applies to subhalos as well, e.g. the subhalos belonging to group 0 are listed first.

In order to reduce confusion, we adopt the following terminology when referring to different types of objects. "Group", "FoF Group", and "FoF Halo" all refer to halos. "Subgroup", "Subhalo", and "Subfind Group" all refer to subhalos. The first (most massive) subgroup of each halo is the "Primary Subgroup" or "Central Subgroup". All other following subgroups within the same halo are "Secondary Subgroups", or "Satellite Subgroups".

FoF Groups. The Group fields are derived with a standard friends-of-friends (FoF) algorithm with linking length $b = 0.2$. The FoF algorithm is run on the dark matter particles, and the other types (gas, stars, BHs) are attached to the same groups as their nearest DM particle. The fields for the FoF halo catalog are described in Table B.1.

Subfind Groups. The Subhalo fields are derived with the SUBFIND algorithm, last described in Springel et al. (2005a). In identifying gravitationally bound substructures the method considers all particle types and assigns them to subhalos as appropriate. It has undergone many modifications to add additional properties to each subhalo entry. Descriptions of all fields in this subhalo catalog are split across Tables B.2 and B.3.

Header and Offsets. Table B.4 describes the fields in the Header group, while Table B.5 describes the fields in the Offsets group. Note that we simply store the offsets here, which relate to all types of data files and not solely to the group catalogs.

3.3. Merger Trees

Merger trees have been created for the various Illustris simulations using SUBLINK (Rodriguez-Gomez et al., 2015), LHALOTREE (Springel et al., 2005a), and CONSISTENT-TREES (using ROCKSTAR, Behroozi et al. 2013, not discussed in detail here). These codes are all included in the Sussing Merger Trees comparison project (Srisawat et al., 2013). In the population average sense the different merger trees give similar results. In more detail, the exact merger history or mass assembly history for any given halo may differ. For a particular science goal, one type of tree may be more or less useful, and users are free to use whichever they prefer. The explicit differences between the otherwise similar LHALOTREE and SUBLINK algorithms are noted below, here we detail their common features.

Figure 4 shows a schematic of the structure of both the SUBLINK and LHALOTREE merger trees. It is not necessary to understand the complete details of the trees to practically use them. In particular, the only critical links are the 'descendant' (black), 'first progenitor' (green), and 'next progenitor' (red) associations. These are shown for all tree nodes in the diagram. For their exact definitions, see Tables B.6 and B.7, the LHALOTREE and SUBLINK tables. Walking back in time following along the main (most massive) progenitor branch consists of following the first progenitor links until they end (value equals -1). Similarly, walking forward in time along the descendants branch consists of following the descendant links until they end (value equals -1), which typically occurs at $z = 0$. The full progenitor history, and not just the main branch, requires following both the first and next progenitor links. In this way the user can identify all subhalos at a previous snapshot which have a common descendant. Examples of walking the tree are provided in the example scripts.

The number inside each circle from the figure is the unique ID (within the whole simulation) of the corresponding subhalo, which is assigned in a depth-first fashion. Numbering also indicates the on-disk storage ordering for the SUBLINK trees, which adopt the approach of Lemson and Virgo Consortium (2006); Lemson and Springel (2006). For example, the main progenitor branch (from 5-7 in the example) and the full progenitor tree (from 5-13 in the example) are both contiguous subsets of each merger tree field, whose location and size can be calculated using these links. The ordering within a single tree in the LHALOTREE is not guaranteed to follow this scheme.

The 'root descendant' (purple), 'last progenitor' (blue),

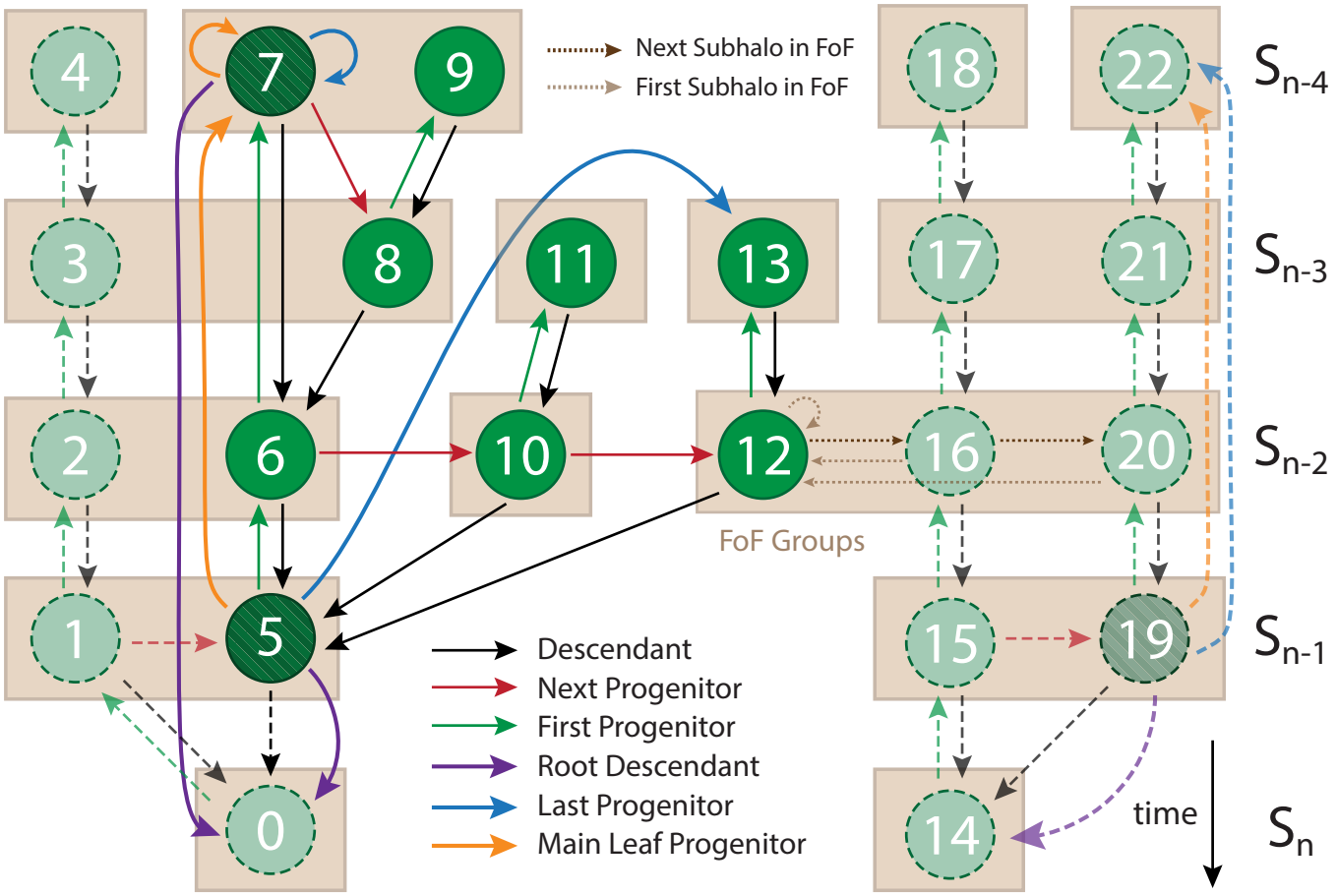

Figure 4: Schematic diagram of the merger tree structure for both SubLink and LHaloTree. Both algorithms connect subhalos (i.e., Subfind halos) across different snapshots in the simulation. Rows indicate discrete snapshots, with time increasing downwards towards redshift zero (the horizontal axis is arbitrary). Green circles represent subhalos (the nodes of the merger tree), while beige boxes indicate the grouping of the subhalos into their parent FoF groups. The most important links are for the descendant (black), first progenitor (green), and next progenitor (red), which are shown for all subhalos. The root descendant (purple), last progenitor (blue), and main leaf progenitor (orange) links exist only for the SubLink trees, and for simplicity these last three link types are shown only for subhalos 5, 7, and 19 (darker striped circles). For exact definitions of each link type, see the corresponding tables. For more information about this figure, consult the text.

and 'main leaf progenitor' (orange) links exist only for the SubLink trees. For simplicity, these last three link types are shown only for nodes 5, 7, and 19 (darker striped circles). Using these links is optional, but allows efficient extraction of main progenitor branches, subtrees (i.e., the set containing a subhalo and "all" its progenitors), "forward" descendant branches, and other subsets of the tree. For their full definitions, see Table B.6 with the SubLink details.

Each subhalo spans a "subtree" consisting of the subhalo itself and all its progenitors. As an example, the subhalos belonging to the subtree of subhalo 5 are shown in darker green in the figure. Other subhalos not belonging to this subtree are shown in lighter green, and their links are indicated with dashed arrows. In the SubLink trees, the subtree of any subhalo can be extracted easily using the 'last progenitor' pointer. As shown in the figure, since subhalo 13 is the 'last progenitor' of subhalo 5, the subtree of subhalo 5 consists of all subhalos with IDs between 5 and 13. Similarly, the main progenitor branch of any subhalo can be retrieved efficiently using the 'main leaf progenitor' link.

Both SubLink and LHaloTree contain the links 'first subhalo in FoF group' (light brown dotted arrow) and 'next subhalo in FoF group' (dark brown dotted arrow), which connect subhalos that belong to the same FoF group. The FoF groups do not play a direct role in the construction of the merger tree. However, subhalos that belong to the same FoF group are also considered to be part of the same tree. As a result, two otherwise independent trees (based on the progenitor and descendant links) are considered to be the same tree if they are "connected" by a FoF group. This is exemplified in the figure by the FoF group containing subhalos 12, 16, and 20. This FoF group acts

as a "bridge" between the left and right trees.

Between the otherwise similar LHaloTree and SubLink algorithms there are three explicit differences, in (i) the merit function used to rank descendants, (ii) the method for skipping snapshots, and (iii) the definition of the main progenitor. In both cases, descendant candidates are identified for each subhalo as those subhalos in the following snapshot(s) that have common particles with the subhalo in question. These candidates are given a score based on a merit function which takes into account the binding energy rank of each matched particle. In this way, preference is given to tracking the fate of the inner parts of a structure, which may survive for a long time upon infall into a bigger halo, even though much of the mass in the outer parts can be quickly stripped. The unique descendant of the subhalo is then the descendant candidate with the highest score. Finally, the halo finder may not detect a small subhalo that is passing through a larger structure in the subsequent snapshot, because the density contrast is not high enough. Descendants are therefore identified also by skipping one snapshot and considering candidates two snapshots apart.

3.3.1. SubLink

SubLink constructs merger trees at the subhalo level (see Rodriguez-Gomez et al. 2015), using a merit function equal to the sum of the binding energy ranks of matched particles, raised to a power of -1. For handling snapshot skipping, it allows some subhalos to skip a snapshot when finding a descendant. In particular, if the highest ranked descendant two snapshots forward differs from the 'descendant of the descendant' found through adjacent snapshots, the former is selected (see Fig. 1 in Rodriguez-Gomez et al. 2015). Once all descendant connections have been made, the main progenitor of each subhalo is defined as the one with the "most massive history" behind it (following De Lucia and Blaizot 2007).

The SubLink merger tree is one large data structure split across several sequential HDF5 files named `tree_extended.[fileNum].hdf5`, where [fileNum] goes from e.g. 0 to 9 for the Illustris-1 run. These files store the data on a per tree basis, and therefore are completely independent from each other. More specifically, any two subhalos that are connected by any of the pointers described in the SubLink table are guaranteed to belong to the same tree, and, therefore, their data is found in the same file. Table B.6 lists the fields which are present in each file.

3.3.2. LHaloTree

The LHaloTree algorithm is virtually identical to that used for the Millennium, Aquarius, and Phoenix simulations, but in HDF5 format. It also constructs trees based on subhalos instead of main halos, and described fully in the supplementary information of Springel et al. (2005b). The unique descendant is selected as the subhalo with the highest score, which as before equals the sum of the binding energy ranks of matched particles, raised in this case to a power of $-2/3$. To allow for the possibility that halos may temporarily disappear for one snapshot, the process is repeated for snapshot n to snapshot $n+2$. If either there is a descendant found in snapshot $n+2$ but none found in snapshot $n+1$, or, if the descendant in snapshot $n+1$ has several direct progenitors and the descendant in snapshot $n+2$ has only one, then a link is made that skips the intervening snapshot. Finally, the main progenitor of each subhalo is selected as the most massive, rather than the one with the most massive history behind it.

The LHaloTree merger tree is one large data structure split across several HDF5 files named `trees_sf1_135.[chunkNum].hdf5`, where [chunkNum] goes from e.g. 0 to 511 for the Illustris-1 run. Within each file there are a number of groups named "TreeX", where X corresponds to the FoF group number in the group catalogs at the final snapshot. However, note that the number X starts over at zero for each tree file chunk, so the FoF group number is recovered by summing of the number of trees in all previous tree file chunks. The pair (SubhaloNumber,SnapNum) provides the indexing into the Subfind group catalog. The five other indices for each entry in a TreeX group index into that same group in the tree file. Table B.7 describes the fields in the Header and TreeX groups.

3.4. Supplementary Data Catalogs

The following additional data products have been computed in post-processing, based on the raw simulation outputs. They are either already available, and now unified under the Illustris data release and made available through the API, or are now made available. In the current effort we focus exclusively on additional properties derived for Illustris-1 galaxies, exclusively at $z = 0$ and above a stellar mass limit of $M_\star \gtrsim 10^9 \mathrm{M}_\odot$.

3.4.1. Stellar Mocks: Multi-band Images and SEDs

A catalog of synthetic stellar images and integrated spectra of galaxies in Illustris-1 at $z = 0$, produced using the radiative transfer code SUNRISE. For complete details on this data product, see Torrey et al. (2015) where it was first described and made available. For all galaxies with stellar masses $M_\star > 10^{10}\mathrm{M}_\odot$ ($\sim 10^4$ star particles and above), both integrated SEDs and spatially resolved photometric maps in 36 broadband filters are computed. There are approximately 7000 galaxies above this limit. For all galaxies with smaller stellar masses, down to 500 star particles, only integrated SEDs are calculated. The 36 bands include GALEX, SDSS, IRAC, Johnson, 2MASS, ACS, and preliminary NIRCAM filters. Note that this is the only data product which is in a format other than HDF5 (namely, FITS). However, the API provides extractions of individual bands and viewing angles in HDF5 format, as well as SEDs in text format, if requested. Finally,

we have developed the Python code SUNPY[3] to add observational realism and make figures based on the raw stellar mock image FITS files.

3.5. Photometric Non-Parametric Stellar Morphologies

A catalog of photometric non-parametric morphologies of Illustris-1 galaxies at $z = 0$. This is meant to replicate automated diagnostics of galaxy stellar structure commonly used observationally, and is calculated by first adding observational realism to the idealized 'stellar mock' images from Torrey et al. (2015), then measuring $(G_{\rm ini}, M_{20}, C, r_P, r_E)$ statistics in four bands, rest-frame u, g, i, and H, each from four directions. For full details on the calculation of each value, see Table C.1 and Snyder et al. (2015) (following Lotz et al., 2004). This data is available for essentially all subhalos with $M_\star > 10^{9.7}{\rm M}_\odot$ at $z = 0$ in Illustris-1. Treating each viewing direction as an independent object, values have been computed for a uniform set of 42531 sources per filter.

3.6. Stellar Circularities, Angular Momenta, Axis Ratios

A catalog for the circularities, angular momenta and axis ratios of the stellar component, for Illustris-1 galaxies. Data is available for all subhalos with stellar mass (inside twice the stellar half mass radius) bigger than $10^9{\rm M}_\odot$. For complete definitions on the calculation of each value, see Table C.2 and Genel et al. (2015), where they were presented and used. The first four quantities in Table C.2 are calculated after alignment with the angular momentum vector of the stars within 10 times the stellar half-mass radius, and measure the quantities inside that radius. The "Circ*" fields are based on the distribution of the circularity parameter ϵ of the individual stars, as defined in Equation (1) of Marinacci et al. (2014). Finally, an analogous calculation including the full stellar content of the subhalos is also provided.

4. Data Access

There are two complementary ways to access the Illustris data products.

1. Raw files can be directly downloaded, and example scripts are provided as a starting point for local analysis.

2. A web-based API can be used, either through a web browser or programmatically in an analysis script, to perform common search and extraction tasks.

These two approaches can be combined. For example, a user may be forced to download the full redshift zero group catalog in order to perform a complex search not supported by the API. After locally determining a sample of interesting galaxies, one could then extract their individual merger trees (and/or raw particle data) without needing to download the full simulation merger tree (or a full snapshot).

Both approaches are documented below, while "getting started" tutorials for several languages (currently: Python, IDL, and Matlab) can be found online.

4.1. Direct File Download and Example Scripts

All of the primary data products for Illustris are released in HDF5 format. This is a portable, self-describing, binary specification suitable for large numerical datasets, for which file access routines are available in all common computing languages. We use only the basic features of the format: groups, attributes, and datasets, with one and two dimensional numeric arrays.

In order to maintain reasonable filesizes, most outputs are split across multiple file "pieces" (or "chunks"). For example, each snapshot of Illustris-1 is split into 512 sequentially numbered files. Individual links to each file chunk are available through the web-based API, and a snapshot can be downloaded in its entirety with a single `wget` command. Direct download links for other snapshots, simulations, and file types (such as group catalogs or merger trees) can be found at the appropriate URLs, as described below. Pre-computed `sha256` checksums are provided for all files so that their integrity can be verified.

The provided example scripts (in IDL, Python, and Matlab) give basic I/O functionality such as: (i) reading a given particle type and/or data field from the snapshot files, (ii) reading only the particle subset from the snapshot corresponding to a halo or subhalo, (iii) extracting the full subtree or main progenitor branch from either SUBLINK or LHALOTREE for a given subhalo, (iv) walking a tree to count the number of mergers, (v) reading the entire group catalog at one snapshot, (vi) reading specific fields from the group catalog, or the entries for a single halo or subhalo. We expect they will serve as a useful starting point for writing any analysis task, and intend them as a 'minimal working examples' which are short and simple enough that they can be quickly understood and extended.

4.2. Web-based API

We have implemented a web-based interface (API) which can respond to a variety of user requests and queries. It is a well-defined interface between the user and the Illustris data products, which is expressed in terms of the required input(s) and expected output(s) for each type of request. The provided functionality is independent, as much as possible, from the underlying data structure, heterogeneity, format, and access methods. The API 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

[3] `http://github.com/ptorrey/sunpy`

by extracting an unmodified subset, or by calculating a derivative quantity.

By specific example, the following types of requests can be handled through the current API, for any simulation at any snapshot:

- List the available simulations, their snapshots, and all associated metadata.
- List all objects in the SUBFIND group catalog and their properties.
- Search with numeric range(s) over any field(s) present in the SUBFIND group catalogs.
- Return all fields from the group catalog for a specific halo or subhalo.
- Return a full snapshot cutout of the particle/cell data for a given halo or subhalo.
- Return a subset of this 'group cutout' containing only specified particle/cell type(s), and/or specific field(s) for each type.
- Return the complete merger history, or just the main progenitor branch, for a given subhalo, for any of the merger trees.
- Download all raw snapshot, group catalog, merger tree, and supplementary data catalog files which exist.
- Download subsets of raw snapshot files, containing only specified particle/cell type(s), and/or specific field(s) for each type.
- Crossmatch subhalos between full physics runs and their dark matter only analogues.
- Traverse relationships between halos and subhalos, for instance from a satellite subhalo to its parent FoF group to the primary (central) subhalo of that group.
- Traverse descendant and primary progenitor links across adjacent snapshots, as available in the SUBLINK merger trees.
- View or render visualizations of the different components (e.g. dark matter, gas, stars) of halos and subhalos, when available.
- Retrieve or calculate additional properties, beyond what is available in the group catalogs, for halos and subhalos, when available.

The Illustris data access API is available at the following permanent URL:

- `http://www.illustris-project.org/api/`

Simple Python examples for working with the API are provided in Appendix D. We provide a list of endpoints, their descriptions, and return types. All accept only GET requests. To provide long-term consistency, we anticipate that the API structure described herein will never change. As additional data products, simulations, tools, and analysis tasks are developed and released, new endpoints will be added. In order to take advantage of new features as they are introduced, we recommend a user consult the up to date API reference available on the website. Tables D.1 and D.2 provide descriptions of each currently available endpoint.

4.2.1. API Access Details

Each API endpoint can return a response in one or more data types. When multiple options exist, a specific return format can be requested through one of the following methods.

- "(?format=)" indicates that the return type is chosen by supplying such a querystring, appended to the URL.
- "(.ext)" indicates that the return type is chosen by supplying the desired file extension in the URL.

Search and Cutout Requests. Several API functions accept additional, optional parameters, which are described here.

{search_query} is an AND combination of restrictions over any of the supported fields, where the relations supported are 'greater than' (gt), 'greater or equal to' (gte), 'less than' (lt), 'less than or equal to' (lte), 'equal to'. The first four work by appending e.g. '__gt=val' to the field name (using a double underscore). For example:

- mass_dm__gt=90.0
- mass__gt=10.0&mass__lte=20.0
- vmax__lt=100.0&len__gas=0&vmaxrad__gt=20.0

{cutout_query} is a concatenated list of particle fields, separated by particle type. The allowed particle types are 'dm','gas','stars','bhs'. The field names are exactly as in the snapshots ("all" is allowed). Omitting all particle types will return the full cutout: all types, all fields. For example:

- gas=Masses,Coordinates,Velocities
- dm=Coordinates&stars=all

Authentication. All API requests require authentication, and therefore also user registration. Each request must provide, along with the details of the request itself, the unique "API Key" of the user making the request. A user can send their API key in the querystring, by appending it to the URL as:

- ?api_key=d22d1f16b894a0b894ec31

A user can alternatively send their API key in HTTP header. This is particularly useful for wget commands or within scripts (see the API tutorial). Note that if a user is logged in to the website, then requests *from the browser* are automatically authenticated. Navigating the Browsable API works in this way.

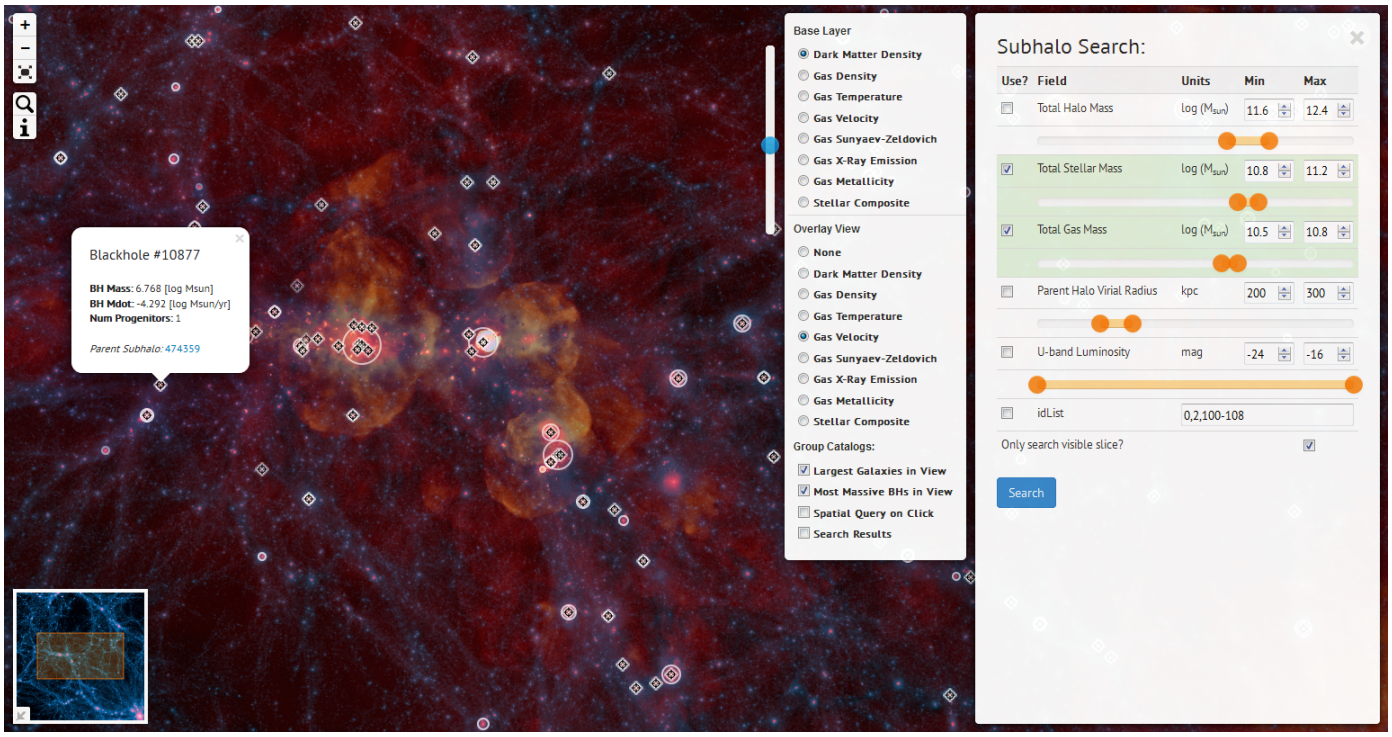

Figure 5: The current Illustris Explorer interface. The main view shows a gas velocity projection overlaid on the dark matter density field. The most massive galaxies currently visible are shown with circles, while black holes are represented with crosshairs. The overview in the lower left corner provides orientation on larger scales. Clicking at any location will launch a spatial search for the nearest subhalos, while clicking on a BH particle will query its details, including a link to its parent subhalo. The central panel controls image layer selection. The right panel presents a simple search interface over subhalo properties.

4.3. Further Online Tools

4.3.1. Subhalo Search Form

We provide a simple search form through which users can query the subhalo database. The search capabilities that exist in the API are exposed in a more human-friendly interface, to enable exploration without the need to write code or write URLs by hand. For example, objects can be selected based on total mass, stellar mass, star formation rate, gas metallicity, or size. The output is a familiar spreadsheet type format, which lists properties from the group catalogs. In addition, each subhalo row provides links to a common set of web-based tools for introspection. These include the canonical link to the object within the API, a form for selecting particle types and initiating an extraction of particles from the snapshot, merger tree visualization, and links to pre-rendered images, when available.

4.3.2. Explorer

The Illustris Explorer[4] is an experiment in the visualization, exploration, and dissemination of large data sets – in particular, those generated by large, astrophysical simulations such as Illustris. It uses the approach of thin-client interaction with derived data products, in this case, precomputed imagery layered under group catalog information. In Figure 5 a full box slice of the simulation is shown in projection, with a depth of 15 Mpc/h, revealing a fifth of the total volume of Illustris at $z = 0$. All the imagery is rendered and saved as hierarchical image pyramids (see also Overzier et al. (2013); Khandai et al. (2014); Bertin et al. (2015)), while rapid search over group properties spatially overlays the results within this volume. All mass components of the simulation are present: the continuous gas and dark matter fields, stellar light from individual stars, and black holes. We have found the interface particularly useful in exploring the spatial relationships between these four components and the discrete halos and subhalos identified with substructure finding algorithms.

4.3.3. Merger Tree

As a demonstration of the potential of rich client applications built on top of the Illustris API, we show in Figure 6 the currently available interface for interactively exploring the merger trees.[5] A zoomed-in portion of the SubLink tree for the 500th most massive central subhalo of Illustris-1 at $z = 0$ is shown. For any run, snapshot,

[4] `www.illustris-project.org/explorer/`

[5] If logged in, this viewer can be launched from inside the Explorer, by selecting a subhalo ID or subhalo circle marker after a search, or through the general subhalo search form.

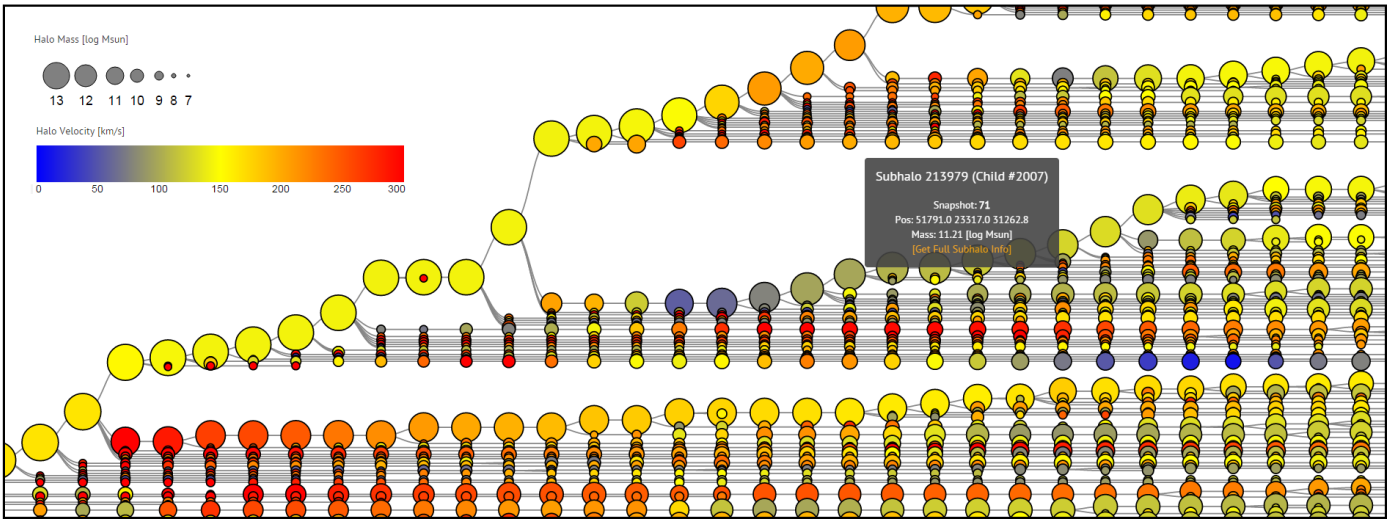

Figure 6: Example of interactive merger tree exploration. We show a zoomed-in portion of the SubLink tree for the 500th most massive central subhalo of Illustris-1 at $z = 0$ (ID 395444). Vector based, client-side rendering means that each node can be interacted with individually. One is shown displaying an informational popup, which includes a link back into the API for inspecting that particular progenitor subhalo. Here we show tree node size scaled with total halo mass in $\log \mathrm{M}_\odot$, and color mapped to subhalo velocity magnitude in km/s.

and subhalo combination, the browser requests a parseable representation of the merger tree from the API (in JSON format), and renders it using the scalable vector graphics (SVG) backend of the d3 javascript visualization library. Because the tree is vector based, and client side, each node can be interacted with individually. Here the informational popup provides a link, back into the API, where the details of the selected progenitor subhalo can be interrogated.

5. Architectural and Implementation Details

In the development of the Illustris public data release, many design decisions were made. Here we discuss technical details related to the release effort, focusing on the relationship between (i) expected use cases with preferred methods of data analysis, and (ii) the specific decisions made to enable those goals, balanced against practical considerations and the need for efficiency. We also contrast with other methodologies, as implemented in other large simulation data releases, and attempt to justify the particular balance struck in the case of Illustris. The details in this section are not necessary for scientific uses of the simulation data.

5.1. Relational Databases

The vast majority of past simulation data releases have made use of relational database systems (i.e. MySQL, PostgreSQL, or commercial options) as the primary mechanism for user interaction as well as data distribution. Following the impressive success of the SDSS Skyserver (Szalay et al., 2000), and starting notably for theory with the Millenium simulation database (Lemson and Virgo Consortium, 2006), users were invited to write and submit raw SQL queries to these databases. Most non-trivial tasks require complex queries which can join multiple tables together across foreign key relations, as well as an awareness of the indexing systems and their use. The power of the query language is offset for most non-experts by the unusual approach, which requires abandoning common methods for the local analysis of astronomical data sets: most notably, the writing of small code snippets, which can have loops and if-else type decision branches. Although many science questions relevant for these projects can be answered by writing suitable SQL queries – as in the "20 typical queries" the SDSS system was designed around (Gray et al., 2002) – it is easy to think up a complex analysis routine which would be unwieldy, if possible at all, with such queries.

In the present effort we have consequently made more limited use of a relational database in the usual way, to hold the full outputs of the group finding algorithms (and not the raw particle data). We exported all group catalogs into the database, with one InnoDB table per run. Each table is partitioned on snapshot number, and has only a single composite B-Tree index on (snapshot,subhalo_id). The goal was to enable rapid search over arbitrary parameter combinations, primarily at a single snapshot. Therefore we did not adopt a merger tree centric ordering (as in Lemson and Virgo Consortium, 2006). In fact, by releasing multiple merger trees we wished to emphasize the fact that there is no ground truth for the merger history of any object, where by definition such an ordering is useful for only one tree. Our snapshot ordering scheme suffers the same limitation – it is specifically reflective of the SubFind group finder employed on-the-fly. However, based on previous experiences within the collaboration, we have adopted this snapshot ordering scheme as being particu-

larly effective for galaxy-centric analyses. Replication of the particle level data using a different ordering (e.g. along a space-filling curve, as has been typically done) would be prohibitively expensive, and so we offer it only in its single existing format.

For interactions with the group catalogs we hide the existence of the database behind an API facade, instead of allowing the direct submission of SQL queries. This approach implies that each piece of functionality must be exposed through an API endpoint. The trade-offs are clear: common tasks which are supported are much easier to accomplish, while more complex or specialized queries are simply not possible. Our motivations for this decision arose out of several considerations.

First, the complexity of hydrodynamical simulation data, as opposed to the dark matter only case, is substantially higher. The number of properties for each halo or galaxy is larger, and the number of possible analysis and post-processing tasks even more so. Therefore, our expectation from the outset was that users would primarily want to process simulation data using their local computing resources and familiar environments. Given this preference towards local data acquisition and analysis, a large focus of the API is on data volume reduction prior to transfer – for example, the ability to download particle data for a single galaxy without having to acquire an entire snapshot, or its merger history without having to download the entire merger tree. This is similar in spirit to Rasera et al. (2010) where particle extraction by halo was also made available, as were sub-volume tilings which together encompass the whole box. Our willingness to promote this approach is driven in part by the increasing availability of high bandwidth network connections, and so the ability to easily download large data volumes. This has undoubtedly also influenced the "raw data download" approaches of other recent, large simulation data releases, Skillman et al. (2014) in particular. One 1.5 TB full snapshot of Illustris can be downloaded in a little under two days at 10 MB/s, a realistic goal for U.S. institutional connections. In reality, then, the only prohibitively large data transfer is the entire set of snapshots.

Our second consideration relates to the use of SQL itself. Previous dark matter simulations implementing the "raw SQL" approach (Lemson and Virgo Consortium, 2006; Crocce et al., 2010; Riebe et al., 2013) demonstrated considerable success in converting users to the language and workflow as a whole, despite it being a relatively unknown tool within the field. The impact of these projects decisively demonstrates the usefulness of this methodology for such projects. Yet, for most users this tool is still foreign, and many uses of the query interface are to simply export data from the database for ingestion into a more familiar data analysis environment. To estimate interest within the community, we conducted an informal survey prior to the design of the Illustris public data release. We report here in brief the most relevant results. Of 125 responses approximately 70% were graduate students, postdocs, or faculty in the field, evenly split between observers and theorists. Given the wordings of the questions, the majority opinion was that accessing astronomical data sets by writing SQL queries worked ok, but was not their primary choice. Given the options, the favored approaches were search, cutout, and data download interfaces which were programmatically accessible. The least favored options involved writing SQL queries or interacting with temporary storage or intermediate outputs stored on remote servers. For data download, the majority preferred direct download over HTTP, in FITS (∼55%) or HDF5 (∼35%) for large binary data and plain text for smaller data sets. We used this input, in combination with our previous experience and the relevant restrictions, to shape the structure of the API and the data release as a whole.

5.2. API Design and Data Formats

The Illustris API is based on a representational state transfer architecture (REST, see Fielding, 2000). Requests and responses are transferred over HTTP, and GET is the only supported request verb (meaning that the system is read-only from the user perspective). Individual resources, or "endpoints", are identified by their unique URL. The system is stateless, meaning that each request is independent of any previous requests, and must include sufficient information to handle it. The default response type is JSON, a human-readable text format which can be parsed by all modern languages and clients. Because the primary purpose of the API is to serve scientific data sets, HDF5 is chosen as the default response type for binary data. For many resources, the response can be requested in any number of supported formats, which currently include CSV, JSON, HDF5, FITS, PNG, and plain text. All are easily digestable by any modern scripting language, and we consider the exact choices rather unimportant, so long as they are widely supported.

In particular, our choice of HDF5 for the primary data products is driven mainly by practicality – whatever output format a simulation writes in, and which the simulators therefore interact with for their own science, will be chosen for the broader release. For example, SDF in the case of Skillman et al. (2014), or raw binary arrays with metadata in numpy saves for Khandai et al. (2014). The only essential requirement is a self-describing binary format, although more sophisticated extraction tasks may be enabled by the features of a specific format. A particularly nice case of this is the use of SDF for direct array slicing through the HTTP protocol (which already supports file subset requests via starting and ending byte positions). Although HDF5 is sufficiently complicated at the bytestream level to make this same approach impossible, our in-memory hyperslab selection method (described below) offers the same functionality with no apparent difference to the user. The only drawback is that responses from the server cannot be blocked (streamed), so the entire requested data set must be temporarily loaded into

memory. Given the small size of our community and the expectation of a correspondingly low number of concurrent requests, this has proven to be a non-issue in practice.

The ability of a client to navigate the API and discover available resources is crucial. We generally adopted the principle of Hypermedia as the Engine of Application State (HATEOAS), meaning that users can discover and request resources in the API without needing to know its structure in advance. This is achieved by stating all relationships between objects in terms of the absolute URL at which each object can be found. For example, the final code listing in Appendix D uses the hyperlinked relationship from a given subhalo to its descendant at a different redshift to walk through a merger tree. In addition to the subhalo catalogs, we also export all relevant metadata for simulation runs and snapshots into the database, which enables the overall API structure. In particular, it allows users to freely discover all available resources (e.g. simulations, snapshots, and types of catalogs or particle data available for each) from the common and fixed API root address. This will enable us to seamlessly include new simulations, as well as new data for existing simulations, as later additions to this initial data release.

In terms of the types of interactions with the API, we aim to support only relatively light queries, which the user should anticipate will complete in a few seconds at most. There is no queued or batch query system, where long running queries can be submitted and their progress periodically polled. There is no per-user remote storage (e.g. "MyDB", **?**). Together, this greatly simplifies the design of the system and maximizes its ease of use, with the implied thought that the typical user workflow will be to download and process specific datasets on their local machine. The ability to offer a remote, persisent, and familiar analysis environment for end users would be a significant though feasible extension of this approach, which we discuss in the following subsection.

As currently designed, users have no need to consider the actual details of where data resides, or how to access it, at the filesystem level. This design goal motivated a system with a split between a front end, which is exposed to the user, and (one or more) back end resources. The separation allows for the two to be in different locations, and for multiple back ends to be supported. In particular, our division is such that the front end handles (i) the Illustris website itself, including (ii) all user details: registration, management, authentication. (iii) All statistics and record keeping. (iv) The full API structure, and responding to API requests at all endpoints. (v) The database, holding both simulation metadata, and the group catalogs. Currently only one back end is in use, and consists of a public-facing machine on the same local network as the data, which is mounted via NFS. It handles:

- Serving raw data files. In this case, several distributed filesystems are locally mounted. Requests are translated into the appropriate system path, and given back to Apache to serve directly via XSendFile.
- Extracted subsets of data files are also served. In this case, the pre-calculated offsets are used in order to only read the requested data from disk. This data is either read into a memory structure in the format requested by the client, or subsequently converted to the requested format. In particular, binary extractions from HDF5 containers are read into an in-memory HDF5 "image". The raw bytestream of this image is then transferred to the client from memory, such that no temporary copy of the data subset need be saved.

The back end is stateless, has no database or persistent local storage of any kind, and no knowledge of the user making each request. This simplifies the addition or transfer of data sources. In order to provide authentication, which forms the basis of usage monitoring, permission levels, bandwidth throttling and rate limits, the following steps are taken:

1. The user makes a request to the API on the front end, including their API-Key.
2. The front end authenticates (verifies their identity) and authorizes (checks sufficient permissions) the user.
3. The front end verifies the validity of the request, including the existence of the requested data.
4. If the request can be satisfied from data available in the front end database (e.g. simulation metadata, subhalo fields), the response is returned directly.
5. If the request requires data from the back end, the appropriate path (URL) is constructed.
6. The front end generates a hash-based message authentication code (HMAC) by concatenating a time-based one-time password (TOTP, see RFC 6238) with a pre-shared secret key and the request URL itself.
7. This token is appended to the back end request URL, which is then sent to the client with a REDIRECT request.
8. The client makes the request to the back end.
9. The back end verifies the request by computing the current TOTP and constructing the same hash using the pre-shared secret key.

The use of the time-varying key means that each request to the back end is attached to a specific request from a specific user. The advantage of this approach is that the front end can redirect clients to data at any back end resource while avoiding the bandwidth burden of making the request itself and forwarding the data on to the client. Although the authentication process is somewhat complex, from the perspective of the user the additional burden is

minimal. We find each of its uses important: (i) usage monitoring is needed for our accurate assessement of impact within the community, (ii) different permission levels allow us to include private or pre-release data for specific collaborators within the same framework, while (iii) bandwidth and rate limits can enforce fair use if necessary.

5.3. Software Stack and Future Directions

At the software level, the Illustris data release makes use of a large number of projects. It is realized on a common open source software stack: CentOS, Apache, and MySQL. On the front end, Python is used to handle all dynamic web content through the Django web framework with several packages including the Django REST framework. The website uses the Bootstrap framework, the jQuery javascript library, MathJax and pygments rendering. The Explorer interface uses the Leaflet tile map engine, as well as the two-dimensional R-Tree indexing capabilities in MySQL to locate subhalos and black holes inside in the visible bounding box. Currently there is no support for spatial indexing in higher dimensions, so using the database for 3D (periodic) distance queries would require a custom solution (Lemson et al., 2011).

Client-side visualizations, currently for the merger trees, use the d3 javascript data visualization library, and three.js for WebGL. There is significant room for the development of additional features in these areas. In particular, for (i) on-demand visualization tasks, (ii) on-demand analysis tasks, and (iii) client-side, browser based tools for data exploration and visualization. For example, (i) requesting an image of projected gas density for a given halo, (ii) requesting a power-law radial slope measurement of a stellar halo or best-fit NFW parameters, and (iii) an interactive 3D representation of the subhalos within a given halo. We welcome community input and direct contributions in any of these directions. On the back end, the HDF5 library with the h5py, numpy, and fitsio Python packages provide the bulk of the data interaction layer.

This back end is currently only focused on storage and data delivery, and we do not yet have any system in place to allow temporary, guest access to compute resources which are local to the data itself. However, we envision that this could change in the future. The data delivery portal has access to the compute resources of the cluster, and instead of defining specific, pre-written analysis functions, we would like to provide a familiar environment for the execution of arbitrary user programs. There has been significant recent development related to remote, multi-user, rich interfaces to computational kernels. In particular, the Jupyter notebook environment (previously called IPython, Pérez and Granger 2007) can be spawned, on demand, inside sand-boxed Docker instances, through a web-based portal with authentication provided by the existing user registration system. This means that users could develop analysis routines in any language (Jupyter support includes Python, IDL, Matlab, Julia, and many others) and execute them, in the same interface, on the remote cluster. We view this possibility as a promising future direction, particularly for researchers who require such remote resources, and otherwise would be unable to use the data for their science.

Finally, the read-only, highly structured nature of simulation output motivates different and more efficient approaches for data search and processing. As an alternative to search within a relational database, one could consider bitmap indexing over HDF5 as in FastQuery (Chou et al., 2011; Byna et al., 2012) together with a SQL-like query layer (Wang et al., 2013). When these technologies are slightly more mature, the need to place a copy of raw simulation data into a database will be removed. Instead, the DB can be used only to handle meta-data, and fast indexed search and queries can be made directly against structured binary data on disk. We anticipate that such an approach might be relevant for future data release efforts, although the sophistication of existing software building blocks already enables an effective way to broadly release both large data sets and rich tools for subsequent data interrogation and analysis.

6. Scientific Remarks and Cautions

The Illustris Simulations (particularly Illustris-1) have been shown to resolve many details of the small-scale properties of galaxies, as well as the evolution of stars and gas within the cosmic web. Illustris-1 reproduces many observational facts on the demographics and properties of the galaxy populations at various epochs, and on the distribution of gas on large scales. As described in Section 2, this has been achieved with a comprehensive galaxy formation model which is intended to account for all the primary processes that are believed to be important for the formation and evolution of galaxies.

However, the enormous dynamical range and the variety and complexity of physics phenomena involved in these numerical endeavours necessarily involve some modeling uncertainties. We have identified below the known problems and points of caution in the Illustris simulated output that any user of the public data must be aware of before embarking on the analysis of the released products. These points should be carefully taken into account before advancing scientific conclusions or making comparisons to observational results.

6.1. Caveats with the Illustris Galaxy Formation Model

Limitations in the Illustris implementations of the stellar and AGN feedback, and possibly of the adopted star-formation recipe, determine a series of issues in the simulated galaxy populations and gas content of halos in comparison to observational constraints. These all point to an inefficient quenching of the star formation in galaxies at different masses and regimes, and in some cases also to

qualitatively not-realistic behaviors of the feedback models. In particular, we note the following issues applicable to the highest-resolution realization (Illustris-1).

- The cosmic star formation rate density is too high at $z \lesssim 1$, possibly because of an inefficient quenching of galaxies residing in halos of $10^{11-12}\mathrm{M}_\odot$ (see Figs. 8 and 2 in Vogelsberger et al., 2014b; Genel et al., 2014, respectively).
- The stellar mass function at $z \lesssim 1$ is too high both at the high and the low ends of the sampled stellar mass range, $M_\star \lesssim 10^{10}\mathrm{M}_\odot$ and $M_\star \gtrsim 10^{11.5}\mathrm{M}_\odot$, see Fig.11, Vogelsberger et al. (2014b) and Fig.3, Genel et al. (2014).
- The physical extent of galaxies can be a factor of a few larger than observed for $M_\star \lesssim 10^{10.7}\mathrm{M}_\odot$ (see Fig. 9 in Snyder et al., 2015).
- The galaxy color distribution deviates from observations in that it does not exhibit a clear bimodality between red and blue galaxies, and the green-valley and the blue cloud appear over populated with respect to to the red sequence (especially for $M_\star \gtrsim 10^{10}\mathrm{M}_\odot$ (see Fig.14 in Vogelsberger et al., 2014b).
- About 10 percent of disk galaxies in the mass range $M_\star \sim 10^{10.5-11}\mathrm{M}_\odot$ at $z = 0$ exhibit strong stellar and gaseous ring-like features, and appear as an additional sub-population in the $G_{\mathrm{ini}} - M_{20}$ plane (see Fig. 5 in Snyder et al., 2015); such features appear to be even more frequent at higher redshifts. Via fragmentation, stellar rings may give rise to spurious stellar clumps that the SUBFIND algorithm identifies as subhalos but whose origin and existence is not necessarily physically well motivated (see also below). Furthermore, these stellar rings are often associated with cores in the stellar and dark matter components, visible in the inner radial density profiles. These cores can extend up $\sim$10 kpc in radius and are likely not realistic in detail.
- The total gas within $R_{500\mathrm{c}}$ is underestimated at late times by a factor 3-10 in halos with $M_{500\mathrm{c}} \sim 10^{13-14}\mathrm{M}_\odot$, because of the too violent operation mode of the Illustris radio-mode feedback (see Fig. 10 in Genel et al., 2014).
- For similar reasons, the bolometric X-ray luminosity in the hot coronae of elliptical galaxies is by many factors lower than in spiral galaxies, contradicting observational constraints (see Section 5.2 of Bogdan et al., 2015); and the predictions for the Sunyaev-Zel'dovich signals from Illustris clusters are not reliable (Popa et al. 2015, in prep).

For some items of this list we have intentionally omitted more specific quantifications of the tensions with observations for two reasons: on the one side, not all observational results are in agreement among each other, making quantitative statements necessarily partial; on the other side, excruciating care is necessary to properly map simulated variables into observationally-derived quantities. For example, we notice that the adopted low star-formation density threshold value and the low thermal energy content of galactic winds may be the cause for spurious star-formation in the circumgalactic medium around Milky Way-like galaxies, at large distances from the natural, dense sites of star formation activity (i.e. disks, see Marinacci et al. 2014). However, no observational data are available to properly quantify such phenomenon. Similarly, the impact of the AGN feedback on the dark-matter distribution within Illustris halos might be overestimated, but direct observational constraints are lacking. Furthermore, while a first analysis of the stellar ages of Illustris galaxies seemed to reveal an overestimation of the predicted stellar ages for $M_\star \lesssim 10^{10.5}\mathrm{M}_\odot$ galaxies (see Fig. 25, Vogelsberger et al. 2014b), we have now recognized that such a comparison to observations is rather inconclusive, as the shape of the age-mass relation of galaxies strongly depends, in the first place, on whether stellar ages are measured by mass- or light- weighting.

To better inform which features of the simulations should be trusted when making science conclusions, we note also following points more directly related to numerical choices:

- In both the snapshots and halo catalogs, metallicity values should be used and interpreted with care. These depend on the underlying choices for stellar evolution and metal enrichment, with tabulated yields being uncertain and continuously updated. Furthermore, no metallicity floor has been imposed to the output data, so that metallicities of a small fraction of gas and star elements adopt minuscule, unrealistic values. In this case, a convenient and appropriate metallicity floor can be adopted, as necessary.
- In the SUBFIND catalogs, relatively-low mass, stellar- or gas-dominated objects at small galactocentric distances from their host halos may be artifacts and should be considered with care. These may be the results of the fragmentation of aforementioned stellar rings in disk galaxies, and may appear as outliers in halos/galaxies scaling relations involving sizes, masses, metallicities and mass-to-light ratios.
- Low-mass BHs in relatively low-mass subhalos should also be considered with care, particularly those hosted in satellite subhalos of more massive galaxies or at low redshifts. Because spurious motions of BH particles are prevented by repositioning the BH on halo potential minimum, in some cases, low-mass BHs in satellite galaxies are repositioned on the central halo on artificially short timescales. These "empty" satellites may then be repopulated with new BH seeds, regardless of redshift. The vast majority of these late-forming, satellite-hosted seeds do not grow significantly before merging with the central BH, so the effects are largely confined to BHs with mass $< 10^6 M_\odot$.

7. Community Considerations

7.1. Citation

To support proper attribution, recognize the effort of individuals involved, and monitor ongoing usage and impact, we request the following. Any publication making use of data from the Illustris simulations should cite this release paper (Nelson et al. 2015b) as well as the original paper introducing the project (Vogelsberger et al., 2014a). Furthermore, extensive use of the data, or studies of galaxy properties and populations, should cite if appropriate Vogelsberger et al. (2014b) as well as Genel et al. (2014). Any investigation of the black hole population should cite if appropriate Sijacki et al. (2014).

Finally, use of any of the supplementary data products should include the relevant citation. A full and up to date list is maintained on the Illustris website. At the time of publication, this includes use of the SubLink merger trees (Rodriguez-Gomez et al., 2015), the redshift zero synthetic stellar images (Torrey et al., 2015), the subsequently derived morphological parameters (Snyder et al., 2015), and the stellar angular momentum, circularity measurements, and axis ratios (Genel et al., 2015).

7.2. Collaboration and Contributions

The full snapshots of Illustris-1 are sufficiently large that it will be prohibitive for most users to acquire or store a large number. As a result, projects which require access to the entire snapshot set may benefit from closer interaction with members of the Illustris collaboration. In particular, many team members are open to more direct collaboration, which can include guest access to compute resources which are local to full copies of the data. We welcome ideas for joint projects, so long as they intersect with the interests of collaboration members and do not overlap with existing efforts. We suggest, practically, to contact the author(s) who have already published work using Illustris data in related scientific topics.[6]

We also welcome contributions to the data release. These can take the form of either analysis code, or computed data products. For example, with the development of an (expensive) analysis routine, we can run it against one or all simulations or snapshots. The resulting data can be made immediately public through the Illustris API. Alternatively, the resulting data can be made privately available until an initial publication is released, and then released publicly. With the development of an (inexpensive, fast) analysis routine, we can integrate it into the Illustris API, such that it can be requested on demand for any object. In this case, analysis should be restricted to subhalo or halo particles, and take at most a few seconds. For the production of a data set derived from the Illustris simulations, in order to make it publicly available, we can host and distribute it alongside the other supplementary data catalogs.

[6] See `http://www.illustris-project.org/results/` for a list.

7.3. Future Data Releases

We anticipate release of additional data in the near future, for which further documentation will be provided online.

7.3.1. Rockstar and Consistent-Trees

We plan to release Rockstar group catalogs and the Consistent-Trees merger trees built upon them for the six Illustris boxes in the near future, and will provide further documentation at that time. These group catalogs can include a different subhalo population than identified with the Subfind algorithm, particularly during mergers. The algorithm used to construct the C-Trees also has fundamental differences to both LHaloTree and SubLink, inserting 'ghost' nodes or modifying properties of existing nodes such that objects in the tree may not map 1-to-1 to the group catalogs from which they were constructed. The output format and structure also differ substantially from either of the two other trees.

These additional catalogs can provide a powerful comparison and consistency check for any scientific analysis. We also anticipate that some users will simply be more familiar with these outputs, or need them as inputs to other tools.

7.3.2. Additional Supplementary Data Catalogs

The $z = 0$ "stellar mocks" multi-band images are being generated for twelve additional snapshots of Illustris-1 at $0.5 < z < 9$. These will include two sets of mock images in 47 common filters, one observing galaxies redshifted to the appropriate epoch and the other observing galaxies in their rest frame. In addition, we expect to add maps of mass, metallicity, gas and stellar velocity, and gas and stellar velocity dispersion in the same projections as these synthetic images. Subsequently, we will also release the non-parametric morphology catalogs for the high redshift galaxy populations.

We expect to release a mock strong lensing catalog, which includes properties of galaxies that most resemble the observed lenses in term of mass/velocity dispersion.The following properties will be available: the Einstein radius R_E, the projected and 3d radial profile slopes, dark matter fraction within R_E, central stellar velocity dispersion, anisotropic parameters, effective radius, Sersic index, light ellipticity and orientation. This data will be available at several redshifts from $z = 0$ to $z = 1$, assuming fiducial source redshifts (**?**).

Additional details on the black holes will be provided: high time resolution outputs of black hole properties, and enumeration of all black hole merger events. This data is new and independent from the snapshots (**?**).

Stellar assembly and merger history catalogs will be released, including details such as in-situ/ex-situ fractions, stellar mass formed pre/post infall, number of major and minor mergers in different time intervals and time since

recent merger events. This data will be available for all subhalos at all snapshots of Illustris-123.

Dark-matter halo catalogs at selected snapshots will be released including dark-matter density profiles fit parameters, fit-independent concentration estimates, halo formation times, and halo shapes.

Mock images and property catalogs of Illustris-1 stellar halos will be released, at a selection of snapshots between z=0 to z=2.

We plan to publish lightcone images, whereby we transform raw simulation data from all snapshots into self-consistent mock-observed survey fields, in HST and JWST filters.

7.3.3. Additional Simulations

Several smaller simulations related to Illustris have been discussed in previous papers, including a series of $25\mathrm{Mpc}/h$ boxes with variations on the input feedback parameters. These can be released in the future if there is community interest. Ongoing and future projects, including higher resolution "zooms" of individual systems, as well as larger volumes, will also be released through this platform in the future.

8. Summary and Conclusions

We have made publicly available all the simulated data associated with the Illustris project at the permanent URL:

- `http://www.illustris-project.org/data/`

The Illustris project includes a series of large-scale, cosmological simulations ideal for studying the formation and evolution of galaxies. The simulation suite consists of three runs at increasing resolution levels of the same $(106.5\ \mathrm{Mpc})^3$ cosmological volume, with and without baryonic physics included. The high-resolution simulations (Illustris-1 and Illustris-1-Dark) include several million gravitationally bound structures, and the $z = 0$ Illustris-1 volume contains ∼7000 well-resolved galaxies with stellar mass exceeding $10^{10}\mathrm{M}_\odot$. The galaxies sampled in this volume span a range of environments and formation histories, allowing for a wide range of science topics to be addressed using the simulation data. For all six realizations, we are releasing the following data products:

- the raw snapshots at all 136 available redshifts down to $z = 0$;
- the friends-of-friends and SUBFIND halo/galaxy catalogs at the same 136 available redshifts down to $z = 0$;
- the SUBLINK and LHALOTREE merger trees;
- the raw snapshots of four sub regions of the full volume, for each full physics run, output with significantly higher time frequency;
- supplementary data catalogs currently focused on properties of the Illustris-1 $z = 0$ galaxy population.

We anticipate release of additional data post-processed products in the near future, for which further documentation will be provided online. Although the total data volume associated with the Illustris project which is presently released is sizeable, ∼265 TB, we have made a significant effort to make this data accessible to the broader community. Specifically, the simulation data is available either via direct download of the raw files or via web-based API queries for common search, extraction, and analysis tasks. Extensive documentation on the format and contents of all released datasets is included both in this paper as well as online, where it will be progressively extended. Additionally, we have made basic I/O scripts and starting examples in IDL, Python, and Matlab available to enable users to analyze and work with the raw data. The resulting data products have widespread applications and provide a powerful tool for the interpretation of extragalactic observations. By making this data publicly available, we hope to maximize the scientific return from the considerable computational resources invested into running the Illustris simulation suite.

Acknowledgements

DN would like to thank Research Computing and the Odyssey cluster at Harvard University for significant computational resources. AP acknowledges support from the HST grant HST-AR-13897. SG acknowledges support provided by NASA through Hubble Fellowship grant HST-HF2-51341 001-A awarded by the STScI, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. VS acknowledges support by the European Research Council under ERC-StG grant EXAGAL-308037, and by the DFG Priority Program SPPEXA through project EXAMAG. PT acknowledges support from NASA ATP Grant NNX14AH35G. GS acknowledges support from HST grants HST-AR-12856.01-A and HST-AR-13887.004-A. Funding for HST programs #12856 and #13887 is provided by NASA through grants from STScI. LB acknowledges support provided by NASA through Einstein Fellowship grant PF2-130093. LH acknowledges support from NASA grant NNX12AC67G and NSF grant AST-1312095. The authors would like to thank many people for contributing to analysis and understanding of the Illustris simulations and their results: Andreas Bauer, Simeon Bird, Akos Bogdan, Aaron Bray, Eddie Chua, Benjamin Cook, Chris Hayward, Rahul Kannan, Luke Kelley, Cristina Popa, Kevin Schaal, Martin Sparre, Joshua Suresh, Sarah Wellons.

The Illustris-1 simulation was run on the CURIE supercomputer at CEA/France as part of PRACE project RA0844, and the SuperMUC computer at the Leibniz Computing Centre, Germany, as part of GCS-project pr85je. The further simulations were run on the Harvard Odyssey and CfA/ITC clusters, the Ranger and Stampede supercomputers at the Texas Advanced Computing Center through

XSEDE, and the Kraken supercomputer at Oak Rridge National Laboratory through XSEDE.

References

References

Barro G. et al., 2013. CANDELS: The Progenitors of Compact Quiescent Galaxies at z ˜ 2. ApJ 765, 104. doi:10.1088/0004-637X/765/2/104, arXiv:1206.5000.

Bauer, A., Springel, V., Vogelsberger, M., Genel, S., Torrey, P., Sijacki, D., Nelson, D., Hernquist, L., 2015. Hydrogen Reionization in the Illustris Universe. ArXiv e-prints arXiv:1503.00734.

Behroozi, P.S., Wechsler, R.H., Wu, H.Y., Busha, M.T., Klypin, A.A., Primack, J.R., 2013. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. ApJ 763, 18. doi:10.1088/0004-637X/763/1/18, arXiv:1110.4370.

Bernyk M. et al., 2014. The Theoretical Astrophysical Observatory: Cloud-Based Mock Galaxy Catalogues. ArXiv e-prints arXiv:1403.5270.

Bertin, E., Pillay, R., Marmo, C., 2015. Web-based visualization of very large scientific astronomy imagery. Astronomy and Computing 10, 43–53. doi:10.1016/j.ascom.2014.12.006, arXiv:1403.6025.

Bird, S., Haehnelt, M., Neeleman, M., Genel, S., Vogelsberger, M., Hernquist, L., 2015. Reproducing the kinematics of damped Lyman α systems. MNRAS 447, 1834–1846. doi:10.1093/mnras/stu2542, arXiv:1407.7858.

Bird, S., Vogelsberger, M., Haehnelt, M., Sijacki, D., Genel, S., Torrey, P., Springel, V., Hernquist, L., 2014. Damped Lyman α absorbers as a probe of stellar feedback. MNRAS 445, 2313–2324. doi:10.1093/mnras/stu1923, arXiv:1405.3994.

Bird, S., Vogelsberger, M., Sijacki, D., Zaldarriaga, M., Springel, V., Hernquist, L., 2013. Moving-mesh cosmology: properties of neutral hydrogen in absorption. MNRAS 429, 3341–3352. doi:10.1093/mnras/sts590, arXiv:1209.2118.

Bogdan A. et al., 2015. Hot Gaseous Coronae around Spiral Galaxies: Probing the Illustris Simulation. ArXiv e-prints arXiv:1503.01107.

Boylan-Kolchin, M., Springel, V., White, S.D.M., Jenkins, A., Lemson, G., 2009. Resolving cosmic structure formation with the Millennium-II Simulation. MNRAS 398, 1150–1164. doi:10.1111/j.1365-2966.2009.15191.x, arXiv:0903.3041.

Brammer G. B. et al., 2012. 3D-HST: A Wide-field Grism Spectroscopic Survey with the Hubble Space Telescope. ApJS 200, 13. doi:10.1088/0067-0049/200/2/13, arXiv:1204.2829.

Bruzual, G., Charlot, S., 2003. Stellar population synthesis at the resolution of 2003. MNRAS 344, 1000–1028. doi:10.1046/j.1365-8711.2003.06897.x, arXiv:astro-ph/0309134.

Bryan, G.L., Norman, M.L., 1998. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. ApJ 495, 80–99. doi:10.1086/305262, arXiv:astro-ph/9710107.

Buser, R., 1978. A systematic investigation of multicolor photometric systems. I - The UBV, RGU and UVBY systems. II - The transformations between the UBV and RGU systems. A&A 62, 411–430.

Byna S. et al., 2012. Parallel i/o, analysis, and visualization of a trillion particle simulation, in: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, IEEE Computer Society Press, Los Alamitos, CA, USA. SC '12. pp. 59:1–59:12. URL: http://dl.acm.org/citation.cfm?id=2388996.2389077.

Cen, R., 1992. A hydrodynamic approach to cosmology - Methodology. ApJS 78, 341–364. doi:10.1086/191630.

Chabrier, G., 2003. Galactic Stellar and Substellar Initial Mass Function. PASP 115, 763–795. doi:10.1086/376392, arXiv:astro-ph/0304382.

Chou, J., Wu, K., Prabhat, 2011. Fastquery: A parallel indexing system for scientific data, in: Cluster Computing (CLUSTER), 2011 IEEE International Conference on, pp. 455–464. doi:10.1109/CLUSTER.2011.86.

Ciotti, L., Ostriker, J.P., 2007. Radiative Feedback from Massive Black Holes in Elliptical Galaxies: AGN Flaring and Central Starburst Fueled by Recycled Gas. ApJ 665, 1038–1056. doi:10.1086/519833, arXiv:astro-ph/0703057.

Crocce, M., Fosalba, P., Castander, F.J., Gaztañaga, E., 2010. Simulating the Universe with MICE: the abundance of massive clusters. MNRAS 403, 1353–1367. doi:10.1111/j.1365-2966.2009.16194.x, arXiv:0907.0019.

Dahlen T. et al., 2004. High-Redshift Supernova Rates. ApJ 613, 189–199. doi:10.1086/422899, arXiv:astro-ph/0406547.

Davis, M., Efstathiou, G., Frenk, C.S., White, S.D.M., 1985. The evolution of large-scale structure in a universe dominated by cold dark matter. ApJ 292, 371–394. doi:10.1086/163168.

Davis M. et al., 2003. Science Objectives and Early Results of the DEEP2 Redshift Survey, in: Guhathakurta, P. (Ed.), Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II, volume 4834 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. pp. 161–172. doi:10.1117/12.457897, arXiv:astro-ph/0209419.

De Lucia, G., Blaizot, J., 2007. The hierarchical formation of the brightest cluster galaxies. MNRAS 375, 2–14. doi:10.1111/j.1365-2966.2006.11287.x, arXiv:astro-ph/0606519.

Di Matteo, T., Colberg, J., Springel, V., Hernquist, L., Sijacki, D., 2008. Direct Cosmological Simulations of the Growth of Black Holes and Galaxies. ApJ 676, 33–53. doi:10.1086/524921, arXiv:0705.2269.

Di Matteo, T., Springel, V., Hernquist, L., 2005. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. Nature 433, 604–607. doi:10.1038/nature03335, arXiv:astro-ph/0502199.

Dubois Y. et al., 2014. Dancing in the dark: galactic properties trace spin swings along the cosmic web. MNRAS 444, 1453–1468. doi:10.1093/mnras/stu1227, arXiv:1402.1165.

Faucher-Giguère, C.A., Lidz, A., Zaldarriaga, M., Hernquist, L., 2009. A New Calculation of the Ionizing Background Spectrum and the Effects of He II Reionization. ApJ 703, 1416–1443. doi:10.1088/0004-637X/703/2/1416, arXiv:0901.4554.

Ferland, G.J., Korista, K.T., Verner, D.A., Ferguson, J.W., Kingdon, J.B., Verner, E.M., 1998. CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra. PASP 110, 761–778. doi:10.1086/316190.

Fielding, R.T., 2000. Architectural Styles and the Design of Network-based Software Architectures. Ph.D. thesis. AAI9980887.

Genel, S., Fall, S.M., Hernquist, L., Vogelsberger, M., Snyder, G.F., Rodriguez-Gomez, V., Sijacki, D., Springel, V., 2015. Galactic Angular Momentum in the Illustris Simulation: Feedback and the Hubble Sequence. ArXiv e-prints arXiv:1503.01117.

Genel, S., Vogelsberger, M., Nelson, D., Sijacki, D., Springel, V., Hernquist, L., 2013. Following the flow: tracer particles in astrophysical fluid simulations. MNRAS 435, 1426–1442. doi:10.1093/mnras/stt1383, arXiv:1305.2195.

Genel S. et al., 2014. Introducing the Illustris project: the evolution of galaxy populations across cosmic time. MNRAS 445, 175–200. doi:10.1093/mnras/stu1654, arXiv:1405.3749.

Gray, J., Szalay, A.S., Thakar, A.R., Kunszt, P.Z., Stoughton, C., Slutz, D., vandenBerg, J., 2002. Data Mining the SDSS SkyServer Database. eprint arXiv:cs/0202014 arXiv:cs/0202014.

Greggio, L., 2005. The rates of type Ia supernovae. I. Analytical formulations. A&A 441, 1055–1078. doi:10.1051/0004-6361:20052926, arXiv:astro-ph/0504376.

Grogin N. A. et al., 2011. CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey. ApJS 197, 35. doi:10.1088/0067-0049/197/2/35, arXiv:1105.3753.

Guo Q. et al., 2011. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. MNRAS 413, 101–131. doi:10.1111/j.1365-2966.2010.18114.x, arXiv:1006.0106.

Hahn, O., Teyssier, R., Carollo, C.M., 2010. The large-scale orientations of disc galaxies. MNRAS 405, 274–290. doi:10.1111/j.

1365-2966.2010.16494.x, arXiv:1002.1964.
Hernquist, L., Katz, N., 1989. TREESPH - A unification of SPH with the hierarchical tree method. ApJS 70, 419–446. doi:10.1086/191344.
Karakas, A.I., 2010. Updated stellar yields from asymptotic giant branch models. MNRAS 403, 1413–1425. doi:10.1111/j.1365-2966.2009.16198.x, arXiv:0912.2142.
Katz, N., Hernquist, L., Weinberg, D.H., 1992. Galaxies and gas in a cold dark matter universe. ApJL 399, L109–L112. doi:10.1086/186619.
Katz, N., Weinberg, D.H., Hernquist, L., 1996. Cosmological Simulations with TreeSPH. ApJS 105, 19. doi:10.1086/192305, arXiv:arXiv:astro-ph/9509107.
Kereš, D., Vogelsberger, M., Sijacki, D., Springel, V., Hernquist, L., 2012. Moving-mesh cosmology: characteristics of galaxies and haloes. MNRAS 425, 2027–2048. doi:10.1111/j.1365-2966.2012.21548.x, arXiv:1109.4638.
Khandai, N., Di Matteo, T., Croft, R., Wilkins, S.M., Feng, Y., Tucker, E., DeGraf, C., Liu, M.S., 2014. The MassiveBlack-II Simulation: The Evolution of Halos and Galaxies to z˜0. preprint, (arXiv:1402.0888) arXiv:1402.0888.
Kim, J., Park, C., Rossi, G., Lee, S.M., Gott, III, J.R., 2011. The New Horizon Run Cosmological N-Body Simulations. Journal of Korean Astronomical Society 44, 217–234. doi:10.5303/JKAS.2011.44.6.217, arXiv:1112.1754.
Klypin, A.A., Trujillo-Gomez, S., Primack, J., 2011. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. ApJ 740, 102. doi:10.1088/0004-637X/740/2/102, arXiv:1002.3660.
Lemson G. et al., 2014. IVOA Recommendation: Simulation Data Model. ArXiv e-prints arXiv:1402.4744.
Lemson, G., Budavári, T., Szalay, A., 2011. Implementing a general spatial indexing library for relational databases of large numerical simulations, in: Proceedings of the 23rd International Conference on Scientific and Statistical Database Management, Springer-Verlag, Berlin, Heidelberg. SSDBM'11. pp. 509–526. URL: http://dl.acm.org/citation.cfm?id=2032397.2032441.
Lemson, G., Springel, V., 2006. Cosmological Simulations in a Relational Database: Modelling and Storing Merger Trees, in: Gabriel, C., Arviset, C., Ponz, D., Enrique, S. (Eds.), Astronomical Data Analysis Software and Systems XV, volume 351 of *Astronomical Society of the Pacific Conference Series*. p. 212.
Lemson, G., Virgo Consortium, t., 2006. Halo and Galaxy Formation Histories from the Millennium Simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the LambdaCDM cosmogony. ArXiv Astrophysics e-prints arXiv:astro-ph/0608019.
Lemson, G., Zuther, J., 2009. Theory in the Virtual Observatory. MEMSAI 80, 342.
Lotz, J.M., Primack, J., Madau, P., 2004. A New Nonparametric Approach to Galaxy Morphological Classification. AJ 128, 163–182. doi:10.1086/421849, arXiv:astro-ph/0311352.
LSST Science Collaborationet al., 2009. LSST Science Book, Version 2.0. ArXiv e-prints arXiv:0912.0201.
Maoz, D., Mannucci, F., Brandt, T.D., 2012. The delay-time distribution of Type Ia supernovae from Sloan II. MNRAS 426, 3282–3294. doi:10.1111/j.1365-2966.2012.21871.x, arXiv:1206.0465.
Marinacci, F., Pakmor, R., Springel, V., 2014. The formation of disc galaxies in high-resolution moving-mesh cosmological simulations. MNRAS 437, 1750–1775. doi:10.1093/mnras/stt2003, arXiv:1305.5360.
Matteucci, F., Panagia, N., Pipino, A., Mannucci, F., Recchi, S., Della Valle, M., 2006. A new formulation of the Type Ia supernova rate and its consequences on galactic chemical evolution. MNRAS 372, 265–275. doi:10.1111/j.1365-2966.2006.10848.x, arXiv:astro-ph/0607504.
Nelson, D., Genel, S., Vogelsberger, M., Springel, V., Sijacki, D., Torrey, P., Hernquist, L., 2015. The impact of feedback on cosmological gas accretion. MNRAS 448, 59–74. doi:10.1093/mnras/stv017.
Nelson, D., Vogelsberger, M., Genel, S., Sijacki, D., Kereš, D., Springel, V., Hernquist, L., 2013. Moving mesh cosmology: tracing cosmological gas accretion. MNRAS 429, 3353–3370. doi:10.1093/mnras/sts595, arXiv:1301.6753.
Okamoto, T., Frenk, C.S., Jenkins, A., Theuns, T., 2010. The properties of satellite galaxies in simulations of galaxy formation. MNRAS 406, 208–222. doi:10.1111/j.1365-2966.2010.16690.x, arXiv:0909.0265.
Oppenheimer, B.D., Davé, R., 2006. Cosmological simulations of intergalactic medium enrichment from galactic outflows. MNRAS 373, 1265–1292. doi:10.1111/j.1365-2966.2006.10989.x, arXiv:astro-ph/0605651.
Oppenheimer, B.D., Davé, R., 2008. Mass, metal, and energy feedback in cosmological simulations. MNRAS 387, 577–600. doi:10.1111/j.1365-2966.2008.13280.x, arXiv:0712.1827.
Overzier, R., Lemson, G., Angulo, R.E., Bertin, E., Blaizot, J., Henriques, B.M.B., Marleau, G.D., White, S.D.M., 2013. The Millennium Run Observatory: first light. MNRAS 428, 778–803. doi:10.1093/mnras/sts076, arXiv:1206.6923.
Pérez, F., Granger, B.E., 2007. IPython: a system for interactive scientific computing. Computing in Science and Engineering 9, 21–29. URL: http://ipython.org, doi:10.1109/MCSE.2007.53.
Pillepich A. et al., 2014. Halo mass and assembly history exposed in the faint outskirts: the stellar and dark matter haloes of Illustris galaxies. MNRAS 444, 237–249. doi:10.1093/mnras/stu1408, arXiv:1406.1174.
Portinari, L., Chiosi, C., Bressan, A., 1998. Galactic chemical enrichment with new metallicity dependent stellar yields. A&A 334, 505–539. arXiv:astro-ph/9711337.
Press, W.H., Schechter, P., 1974. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. ApJ 187, 425–438. doi:10.1086/152650.
Puchwein, E., Springel, V., 2013. Shaping the galaxy stellar mass function with supernova- and AGN-driven winds. MNRAS 428, 2966–2979. doi:10.1093/mnras/sts243, arXiv:1205.2694.
Rahmati, A., Pawlik, A.H., Raičevic, M., Schaye, J., 2013. On the evolution of the H I column density distribution in cosmological simulations. MNRAS 430, 2427–2445. doi:10.1093/mnras/stt066, arXiv:1210.7808.
Rahmati, A., Schaye, J., Bower, R.G., Crain, R.A., Furlong, M., Schaller, M., Theuns, T., 2015. The distribution of neutral hydrogen around high-redshift galaxies and quasars in the EAGLE simulation. ArXiv e-prints arXiv:1503.05553.
Rasera, Y., Alimi, J.M., Courtin, J., Roy, F., Corasaniti, P.S., Füzfa, A., Boucher, V., 2010. Introducing the Dark Energy Universe Simulation Series (DEUSS), in: Alimi, J.M., Fuözfa, A. (Eds.), American Institute of Physics Conference Series, volume 1241 of *American Institute of Physics Conference Series*. pp. 1134–1139. doi:10.1063/1.3462610, arXiv:1002.4950.
Riebe K. et al., 2013. The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. Astronomische Nachrichten 334, 691–708. doi:10.1002/asna.201211900.
Rodriguez-Gomez V. et al., 2015. The merger rate of galaxies in the Illustris Simulation: a comparison with observations and semi-empirical models. ArXiv e-prints arXiv:1502.01339.
Sales L. V. et al., 2015. The colours of satellite galaxies in the Illustris simulation. MNRAS 447, L6–L10. doi:10.1093/mnrasl/slu173, arXiv:1410.7400.
Sazonov, S.Y., Ostriker, J.P., Ciotti, L., Sunyaev, R.A., 2005. Radiative feedback from quasars and the growth of massive black holes in stellar spheroids. MNRAS 358, 168–180. doi:10.1111/j.1365-2966.2005.08763.x, arXiv:astro-ph/0411086.
Schaal, K., Springel, V., 2015. Shock finding on a moving mesh - I. Shock statistics in non-radiative cosmological simulations. MNRAS 446, 3992–4007. doi:10.1093/mnras/stu2386, arXiv:1407.4117.
Schaller M. et al., 2014. The masses and density profiles of halos in a LCDM galaxy formation simulation. ArXiv e-prints arXiv:1409.8617.
Schaye J. et al., 2015. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. MNRAS 446, 521–554. doi:10.1093/mnras/stu2058, arXiv:1407.7040.

Sijacki, D., Springel, V., Di Matteo, T., Hernquist, L., 2007. A unified model for AGN feedback in cosmological simulations of structure formation. MNRAS 380, 877–900. doi:10.1111/j.1365-2966.2007.12153.x, arXiv:0705.2238.

Sijacki, D., Springel, V., Haehnelt, M.G., 2009. Growing the first bright quasars in cosmological simulations of structure formation. MNRAS 400, 100–122. doi:10.1111/j.1365-2966.2009.15452.x, arXiv:0905.1689.

Sijacki, D., Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Snyder, G., Nelson, D., Hernquist, L., 2014. The Illustris simulation: Evolving population of black holes across cosmic time. ArXiv e-prints arXiv:1408.6842.

Sijacki, D., Vogelsberger, M., Kereš, D., Springel, V., Hernquist, L., 2012. Moving mesh cosmology: the hydrodynamics of galaxy formation. MNRAS 424, 2999–3027. doi:10.1111/j.1365-2966.2012.21466.x, arXiv:1109.3468.

Skillman, S.W., Warren, M.S., Turk, M.J., Wechsler, R.H., Holz, D.E., Sutter, P.M., 2014. Dark Sky Simulations: Early Data Release. ArXiv e-prints arXiv:1407.2600.

Smith, B., Sigurdsson, S., Abel, T., 2008. Metal cooling in simulations of cosmic structure formation. MNRAS 385, 1443–1454. doi:10.1111/j.1365-2966.2008.12922.x, arXiv:0706.0754.

Snyder G. F. et al., 2015. Galaxy Morphology and Star Formation in the Illustris Simulation at z=0. ArXiv e-prints arXiv:1502.07747.

Sparre M. et al., 2015. The star formation main sequence and stellar mass assembly of galaxies in the Illustris simulation. MNRAS 447, 3548–3563. doi:10.1093/mnras/stu2713, arXiv:1409.0009.

Springel, V., 2005. The cosmological simulation code GADGET-2. MNRAS 364, 1105–1134. doi:10.1111/j.1365-2966.2005.09655.x, arXiv:arXiv:astro-ph/0505010.

Springel, V., 2010. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. MNRAS 401, 791–851. doi:10.1111/j.1365-2966.2009.15715.x, arXiv:0901.4107.

Springel, V., Di Matteo, T., Hernquist, L., 2005a. Modelling feedback from stars and black holes in galaxy mergers. MNRAS 361, 776–794. doi:10.1111/j.1365-2966.2005.09238.x, arXiv:arXiv:astro-ph/0411108.

Springel, V., Hernquist, L., 2003. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. MNRAS 339, 289–311. doi:10.1046/j.1365-8711.2003.06206.x, arXiv:arXiv:astro-ph/0206393.

Springel V. et al., 2008. The Aquarius Project: the subhaloes of galactic haloes. MNRAS 391, 1685–1711. doi:10.1111/j.1365-2966.2008.14066.x, arXiv:0809.0898.

Springel V. et al., 2005b. Simulations of the formation, evolution and clustering of galaxies and quasars. Nature 435, 629–636. doi:10.1038/nature03597, arXiv:astro-ph/0504097.

Springel, V., White, S.D.M., Tormen, G., Kauffmann, G., 2001a. Populating a cluster of galaxies - I. Results at [formmu2]z=0. MNRAS 328, 726–750. doi:10.1046/j.1365-8711.2001.04912.x, arXiv:arXiv:astro-ph/0012055.

Springel, V., Yoshida, N., White, S.D.M., 2001b. GADGET: a code for collisionless and gasdynamical cosmological simulations. NewA 6, 79–117. doi:10.1016/S1384-1076(01)00042-2, arXiv:astro-ph/0003162.

Srisawat C. et al., 2013. Sussing Merger Trees: The Merger Trees Comparison Project. MNRAS 436, 150–162. doi:10.1093/mnras/stt1545, arXiv:1307.3577.

Stoughton C. et al., 2002. Sloan Digital Sky Survey: Early Data Release. AJ 123, 485–548. doi:10.1086/324741.

Suresh, J., Bird, S., Vogelsberger, M., Genel, S., Torrey, P., Sijacki, D., Springel, V., Hernquist, L., 2015. The Impact of Galactic Feedback on the Circumgalactic Medium. preprint, (arXiv:1501.02267) arXiv:1501.02267.

Szalay, A.S., Gray, J., Thakar, A.R., Kunszt, P.Z., Malik, T., Raddick, J., Stoughton, C., vandenBerg, J., 2002a. The SDSS SkyServer: Public Access to the Sloan Digital Sky Server Data. eprint arXiv:cs/0202013 arXiv:cs/0202013.

Szalay, A.S., Gray, J., VandenBerg, J., 2002b. Petabyte Scale Data Mining: Dream or Reality?, in: Tyson, J.A., Wolff, S. (Eds.), Survey and Other Telescope Technologies and Discoveries, volume 4836 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. pp. 333–338. doi:10.1117/12.461427, arXiv:cs/0208013.

Szalay, A.S., Kunszt, P.Z., Thakar, A.R., Gray, J., Slutz, D., 2000. The Sloan Digital Sky Survey and its Archive, in: Manset, N., Veillet, C., Crabtree, D. (Eds.), Astronomical Data Analysis Software and Systems IX, volume 216 of *Astronomical Society of the Pacific Conference Series*. p. 405. arXiv:astro-ph/9912382.

The Dark Energy Survey Collaboration, 2005. The Dark Energy Survey. ArXiv Astrophysics e-prints arXiv:astro-ph/0510346.

Thielemann F.-K. et al., 2003. Nuclear cross sections, nuclear structure and stellar nucleosynthesis. Nuclear Physics A 718, 139–146. doi:10.1016/S0375-9474(03)00704-8.

Torrey P. et al., 2015. Synthetic galaxy images and spectra from the Illustris simulation. MNRAS 447, 2753–2771. doi:10.1093/mnras/stu2592, arXiv:1411.3717.

Torrey, P., Vogelsberger, M., Genel, S., Sijacki, D., Springel, V., Hernquist, L., 2014. A model for cosmological simulations of galaxy formation physics: multi-epoch validation. MNRAS 438, 1985–2004. doi:10.1093/mnras/stt2295, arXiv:1305.4931.

Torrey, P., Vogelsberger, M., Sijacki, D., Springel, V., Hernquist, L., 2012. Moving-mesh cosmology: properties of gas discs. MNRAS 427, 2224–2238. doi:10.1111/j.1365-2966.2012.22082.x, arXiv:1110.5635.

Travaglio, C., Hillebrandt, W., Reinecke, M., Thielemann, F.K., 2004. Nucleosynthesis in multi-dimensional SN Ia explosions. A&A 425, 1029–1040. doi:10.1051/0004-6361:20041108, arXiv:astro-ph/0406281.

Vogelsberger, M., Genel, S., Sijacki, D., Torrey, P., Springel, V., Hernquist, L., 2013. A model for cosmological simulations of galaxy formation physics. MNRAS 436, 3031–3067. doi:10.1093/mnras/stt1789, arXiv:1305.2913.

Vogelsberger M. et al., 2014a. Properties of galaxies reproduced by a hydrodynamic simulation. Nature 509, 177–182. doi:10.1038/nature13316, arXiv:1405.1418.

Vogelsberger M. et al., 2014b. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. MNRAS 444, 1518–1547. doi:10.1093/mnras/stu1536, arXiv:1405.2921.

Vogelsberger, M., Sijacki, D., Kereš, D., Springel, V., Hernquist, L., 2012. Moving mesh cosmology: numerical techniques and global statistics. MNRAS 425, 3024–3057. doi:10.1111/j.1365-2966.2012.21590.x, arXiv:1109.1281.

Wang, Y., Su, Y., Agrawal, G., 2013. Supporting a Light-Weight Data Management Layer over HDF5 , volume 2013 of *Cluster, Cloud and Grid Computing (CCGrid)*. pp. 335–342. doi:10.1109/CCGrid.2013.9.

Wellons S. et al., 2015. The formation of massive, compact galaxies at z = 2 in the Illustris simulation. MNRAS 449, 361–372. doi:10.1093/mnras/stv303, arXiv:1411.0667.

Wiersma, R.P.C., Schaye, J., Smith, B.D., 2009a. The effect of photoionization on the cooling rates of enriched, astrophysical plasmas. MNRAS 393, 99–107. doi:10.1111/j.1365-2966.2008.14191.x, arXiv:0807.3748.

Wiersma, R.P.C., Schaye, J., Theuns, T., Dalla Vecchia, C., Tornatore, L., 2009b. Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations. MNRAS 399, 574–600. doi:10.1111/j.1365-2966.2009.15331.x, arXiv:0902.1535.

Yepes, G., Kates, R., Khokhlov, A., Klypin, A., 1997. Hydrodynamical simulations of galaxy formation: effects of supernova feedback. MNRAS 284, 235–256. arXiv:astro-ph/9605182.

York D. G. et al., 2000. The Sloan Digital Sky Survey: Technical Summary. AJ 120, 1579–1587. doi:10.1086/301513, arXiv:astro-ph/0006396.

Yu, Q., Tremaine, S., 2002. Observational constraints on growth of massive black holes. MNRAS 335, 965–976. doi:10.1046/j.1365-8711.2002.05532.x, arXiv:astro-ph/0203082.

Appendix A: Snapshot Data Details

Table A.1: Details on the file organization for the six runs. In each case, N_f represents the number of files for each data type, while the provided sizes are the average for that data type. The approximate total data volume for each run is also listed.

Run	Total $N_{\rm DM}$	Snapshot N_f	Groupcat N_f	Snapshot Size	Groupcat Size	Data Volume
Illustris-3	94,196,375	32	2	22 GB	100 MB	3 TB
Illustris-3-Dark	94,196,375	8	2	3.2 GB	50 MB	0.4 TB
Illustris-2	753,571,000	256	4	176 GB	500 MB	24 TB
Illustris-2-Dark	753,571,000	32	4	26 GB	320 MB	3.5 TB
Illustris-1	6,028,568,000	512	8	1.5 TB	3.6 GB	204 TB
Illustris-1-Dark	6,028,568,000	128	8	203 GB	4 GB	28 TB

Table A.2: Details of the Header group in the snapshot files.

Field	Dimensions	Units	Description
BoxSize	1	ckpc/h	Spatial extent of the periodic box (in comoving units).
MassTable	6	$10^{10}{\rm M}_\odot/h$	Masses of particle types which have a constant mass (only DM).
NumPart_ThisFile	6	-	Number of particles (of each type) included in this (sub-)file.
NumPart_Total	6	-	Total number of particles (of each type) in this snapshot, modulo 2^{32}.
NumPart_Total_HighWord	6	-	Total number of particles (of each type) in this snapshot, divided by 2^{32} and rounded downwards.
Omega0	1	-	The cosmological density parameter for matter.
OmegaLambda	1	-	The cosmological density parameter for the cosmological constant.
Redshift	1	-	The redshift corresponding to the current snapshot.
Time	1	-	The scale factor $a = 1/(1+z)$ corresponding to the current snapshot.
NumFilesPerSnapshot	1	-	Number of file chunks per snapshot.

Table A.3: Additional details of the subbox snapshots. For each subbox number, its physical environment, matter overdensity, center position, box size along each coordinate axis, and volume fraction with respect to the full box.

Subbox #	Environment	$\Omega_m^{\rm sub}$	(x_c, y_c, z_c)	$L_{\rm subbox}$	Volume Frac
0	Crowded, one $\sim 5\times10^{13}{\rm M}_\odot$ halo	1.47	(9000, 17000, 63000)	7.5 cMpc/h	0.1%
1	Less crowded, several $> 10^{12}{\rm M}_\odot$ halos	0.16	(43100, 53600, 60800)	8.0 cMpc/h	0.12%
2	Less crowded, several $> 10^{12}{\rm M}_\odot$ halos	0.29	(37000, 43500, 67500)	5.0 cMpc/h	0.03%
3	Least crowded, several $\sim 10^{12}{\rm M}_\odot$ halos	0.25	(64500, 51500, 39500)	5.0 cMpc/h	0.03%

Table A.4: Listing of all snapshot fields for gas (PartType0).

Field	Dimensions	Units	Description
Coordinates	N,3	ckpc/h	Spatial position within the periodic box of size 75000 ckpc/h. Comoving coordinate.
Density	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	Comoving mass density of cell (calculated as mass/volume).
ElectronAbundance	N	-	Fractional electron number density with respect to the total hydrogen number density, so $n_e = \mathrm{ElectronAbundance} * \mathrm{n_H}$ where $n_H = X_H * \rho/m_p$. Use with caution for star-forming gas (see comment below for NeutralHydrogenAbundance).
GFM_ AGNRadiation	N	erg/s/cm^2	Bolometric intensity (physical units) at the position of this cell arising from the radiation fields of nearby AGN.
GFM_ CoolingRate	N	ergcm$^3/s$	The instantaneous net cooling rate experienced by this gas cell, in cgs units (e.g. $\Lambda_{\mathrm{net}}/n_H^2$).
GFM_ Metallicity	N	-	The ratio M_Z/M_{total} where M_Z is the total mass all metal elements (above He). This is not in solar units! To convert to solar metallicity, divide by 0.0127 (the primordial solar metallicity).
GFM_ WindDMVelDisp	N	km/s	Equal to SubfindVelDisp.
InternalEnergy	N	(km/s)^2	Internal (thermal) energy per unit mass for this gas cell.
Masses	N	$10^{10}\mathrm{M}_\odot/h$	Gas mass in this cell. Refinement/derefinement attempts to keep this value within a factor of two of the targetGasMass for every cell.
Neutral Hydrogen Abundance	N	-	Fraction of the hydrogen cell mass (or density) in neutral hydrogen, so $n_{H_0} = \mathrm{NeutralHydrogenAbundance} * \mathrm{n_H}$. (So note that $n_{H^+} = n_H - n_{H_0}$). Use with caution for star-forming gas, as the calculation is based on the 'effective' temperature of the equation of state, which is not a physical temperature.
NumTracers	N	-	The number of child tracers residing within this gas cell.
ParticleIDs	N	-	The unique ID (uint64) of this gas cell. Constant for the duration of the simulation. May cease to exist (as gas) in a future snapshot due to conversion into a star/wind particle, accretion into a BH, or a derefinement event.
Potential	N	(km/s)^2	Gravitational potential energy.
SmoothingLength	N	ckpc/h	Twice the maximum radius of all Delaunay tetrahedra that have this cell at a vertex in comoving units (s_i from Springel et al. 2010).
StarFormationRate	N	$\mathrm{M}_\odot$/yr	Instantaneous star formation rate of this gas cell.
SubfindDensity	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	The local total comoving mass density, estimated using the standard cubic-spline SPH kernel over all particles/cells within a radius of SubfindHsml.
SubfindHsml	N	ckpc/h	The comoving radius of the sphere centered on this cell enclosing the 64$\pm$1 nearest dark matter particles.
SubfindVelDisp	N	km/s	The 3D velocity dispersion of all dark matter particles within a radius of SubfindHsml of this cell.
Velocities	N,3	km$\sqrt{a}$/s	Spatial velocity. The peculiar velocity is obtained by multiplying this value by $\sqrt{a}$.
Volume	N	$1/(\mathrm{ckpc}/h)^3$	Comoving volume of the Voronoi gas cell.

Table A.5: Listing of all snapshot fields for dark matter (PartType1).

Field	Dimensions	Units	Description
Coordinates	N,3	ckpc/h	Spatial position within the periodic box of size 75000 ckpc/h. Comoving coordinate.
ParticleIDs	N	-	The unique ID (uint64) of this DM particle. Constant for the duration of the simulation.
Potential	N	$(\mathrm{km/s})^2$	Gravitational potential energy.
SubfindDensity	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	The local total comoving mass density, estimated using the standard cubic-spline SPH kernel over all particles/cells within a radius of SubfindHsml.
SubfindHsml	N	ckpc/h	The comoving radius of the sphere centered on this particle enclosing the 64±1 nearest dark matter particles.
SubfindVelDisp	N	km/s	The 3D velocity dispersion of all dark matter particles within a radius of SubfindHsml.
Velocities	N,3	km$\sqrt{a}$/s	Spatial velocity. The peculiar velocity is obtained by multiplying this value by $\sqrt{a}$.

Table A.6: Listing of all snapshot fields for tracer particles (PartType3).

Field	Dimensions	Units	Description
FluidQuantities	N,13	Various	Thirteen auxiliary quantities stored for each tracer with differing significance. See Tracer Quantities below.
ParentID	N	-	The unique ID (uint64) of the parent of this tracer. Could be a gas cell, star, wind phase cell, or BH.
TracerID	N	-	The unique ID (uint64) of this tracer. Constant for the duration of the simulation.

Table A.7: Listing of all snapshot fields for stars (PartType4).

Field	Dimensions	Units	Description
Coordinates	N,3	ckpc/h	Spatial position within the periodic box of size 75000 ckpc/h. Comoving coordinate.
GFM_InitialMass	N	$10^{10}\mathrm{M}_\odot/h$	Mass of this star particle when it was formed (will subsequently decrease due to stellar evolution).
GFM_Metallicity	N	-	See entry under PartType0. Inherited from the gas cell spawning/converted into this star, at the time of birth.
GFM_Stellar FormationTime	N	-	The exact time (given as the scale factor) when this star was formed. **Note: The only differentiation between a real star ($>= 0$) and a wind phase gas cell (< 0) is the sign of this quantity.**
GFM_Stellar Photometrics	N,8	*mag*	Stellar magnitudes in eight bands: U, B, V, K, g, r, i, z. In detail, these are: Buser's X filter (Buser, 1978), where X=U,B3,V (Vega magnitudes), then IR K filter + Palomar 200 IR detectors + atmosphere.57 (Vega), then SDSS Camera X Response Function, airmass = 1.3 (June 2001), where X=g,r,i,z (AB magnitudes). They can be found in the filters.log file in the BC03 package[7]. The details on the four SDSS filters can be found in Stoughton et al. (2002), section 3.2.1.
Masses	N	$10^{10}\mathrm{M}_\odot/h$	Mass of this star or wind phase cell.
NumTracers	N	-	Number of child tracers belonging to this star/wind phase cell.
ParticleIDs	N	-	The unique ID (uint64) of this star/wind cell. Constant for the duration of the simulation.
Potential	N	$(\mathrm{km/s})^2$	Gravitational potential energy.
SubfindDensity	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	The local total comoving mass density, estimated using the standard cubic-spline SPH kernel over all particles/cells within a radius of SubfindHsml.
SubfindHsml	N	ckpc/h	The comoving radius of the sphere centered on this star particle enclosing the 64±1 nearest dark matter particles.
SubfindVelDisp	N	km/s	The 3D velocity dispersion of all dark matter particles within a radius of SubfindHsml.
Velocities	N,3	km$\sqrt{a}$/s	Spatial velocity. The peculiar velocity is obtained by multiplying this value by $\sqrt{a}$.

Table A.8: Listing of all snapshot fields for black holes (PartType5).

Field	Dimensions	Units	Description
BH_CumEgy Injection_QM	N	$\frac{10^{10}\mathrm{M}_\odot/h(\mathrm{ckpc}/h)^2}{(0.978\mathrm{Gyr}/h)^2}$	Cumulative amount of thermal AGN feedback energy injected into surrounding gas in the quasar mode.
BH_CumMass Growth_QM	N	$(10^{10}\mathrm{M}_\odot/h)$	Cumulative mass accreted onto the BH in the quasar mode.
BH_Density	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	Local comoving gas density averaged over the nearest neighbors of the BH.
BH_Hsml	N	ckpc/h	The comoving radius of the sphere enclosing the 64 nearest-neighbor gas cells around the BH.
BH_Mass	N	$10^{10}\mathrm{M}_\odot/h$	Actual mass of the BH, does not include gas reservoir. Monotonically increases with time according to the accretion prescription, starting from the seed mass.
BH_Mass_bubbles	N	$10^{10}\mathrm{M}_\odot/h$	Accreted mass in current duty cycle for AGN radio mode bubble feedback. When this value reaches a critical fraction of BH_Mass_ini, the bubble energy is released.
BH_Mass_ini	N	$10^{10}\mathrm{M}_\odot/h$	BH mass at the start of the current duty cycle for AGN radio mode feedback, reset after each duty cycle. See BH_Mass_bubbles.
BH_Mdot	N	$\frac{10^{10}\mathrm{M}_\odot/h}{0.978\mathrm{Gyr}/h}$	The mass accretion rate onto the black hole, instantaneous.
BH_Pressure	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)(0.978\mathrm{Gyr}/h)^2}$	Reference gas pressure (in comoving units) near the BH, defined as $(\gamma-1)\rho_{sfr}u_{eq}$, where ρ_{sfr} is the star-formation threshold and u_{eq} is BH_U (defined below).
BH_Progs	N	-	Total number of BHs that have merged into this BH.
BH_U	N	$(\mathrm{km/s})^2$	Thermal energy per unit mass in quasar-heated bubbles near the BH, assuming equilibrium between radiative cooling and thermal AGN heating near the BH. Used to define the BH_Pressure.
Coordinates	N,3	ckpc/h	Spatial position within the periodic box of size 75000 ckpc/h. Comoving coordinate.
HostHaloMass	N	$10^{10}\mathrm{M}_\odot/h$	Mass of FoF group that hosts the BH.
Masses	N	$10^{10}\mathrm{M}_\odot/h$	Total mass of the black hole particle. Includes the gas reservoir from which accretion is tracked onto the actual BH mass (see BH_Mass).
NumTracers	N	-	The number of child tracers residing within this BH.
ParticleIDs	N	-	The unique ID (uint64) of this black hole. Constant for the duration of the simulation. May cease to exist in a future snapshot due to a BH merger.
Potential	N	$(\mathrm{km/s})^2$	Gravitational potential at the location of the BH.
SubfindDensity	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	The local total comoving mass density, estimated using the standard cubic-spline SPH kernel over all particles/cells within a radius of SubfindHsml.
SubfindHsml	N	ckpc/h	The comoving radius of the sphere centered on this black hole particle enclosing the 64±1 nearest dark matter particles.
SubfindVelDisp	N	km/s	The 3D velocity dispersion of all dark matter particles within a radius of SubfindHsml.
Velocities	N,3	$\mathrm{km}\sqrt{a}/\mathrm{s}$	Spatial velocity. The peculiar velocity is obtained by multiplying this value by $\sqrt{a}$.

Table A.9: Listing of the thirteen auxiliary values stored by the tracer particles. The Reset column indicates whether or not this field is set to zero immediately after each snapshot is written.

Number	Name	Reset?	Units	Description
0	TMax	Y	Kelvin	The maximum past temperature of the parent gas cell, back to the previous snapshot. Only updated when parent is a gas cell.
1	TMax_Time	Y	-	Scale factor of the above TMax event. Only updated when parent is a gas cell.
2	TMax_Time_Rho	Y	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	Density of the parent gas cell when the most recent TMax was recorded. Only updated when parent is a gas cell.
3	RhoMax	Y	$\frac{10^{10}\mathrm{M}_\odot/h}{(\mathrm{ckpc}/h)^3}$	Maximum past density of the parent gas cell, back to the previous snapshot. Only updated when parent is a gas cell.
4	RhoMax_Time	Y	-	Scale factor of the above RhoMax event. Only updated when parent is a gas cell.
5	MachMax	Y	-	Maximum past mach number of the parent gas cell, as set in the Riemann solver. Only updated when parent is a gas cell.
6	EntMax	Y	$P/(\rho/a^3)^\gamma$	Maximum past entropy of the parent gas cell, back to the previous snapshot. Only updated when parent is a gas cell. Note slightly strange units, where P and ρ are pressure and density, as in the snapshots.
7	EntMax_Time	Y	-	Scale factor of the above EntMax event. Only updated when parent is a gas cell.
8	Last_Star_Time	N	-	Scale factor, set only when this tracer exchanges from a star/wind to a gas, or from a gas to a star/wind. These four cases respectively set LST = { a, -a, a+1, a+2 }.
9	Wind_Counter	N	int32	Integer counter initialized to zero, increased by one each time this tracer is moved from a gas cell to a wind particle.
10	Exchange_Counter	N	int32	Integer counter initialized to zero, increased by one each time this tracer is exchanged, regardless of parent type.
11	Exchange_Distance	N	ckpc/h	Cumulative sum of the spatial distance over which this tracer has moved due to Monte Carlo exchange between gas cells. In particular, the sum of the parent gas cell radii when either the originating parent or destination parent is of gas type.
12	Exchange_ Distance_Error	N	ckpc/h	Cumulative sum of $r_{\mathrm{cell}} \times (\sqrt{N_{\mathrm{exch}}} - \sqrt{N_{\mathrm{exch}} - 1})$, when either the originating or destination parent is of gas type.

Appendix B: Group and Merger Tree Data Details

Table B.1: Description of all fields in the FoF halo catalogs. All fields are float32 unless otherwise specified.

Field	Dimensions	Units	Description
GroupBHMass	N	$10^{10}\mathrm{M}_\odot/h$	Sum of the BH_Mass field of all black holes (type 5) in this group.
GroupBHMdot	N	$\frac{10^{10}\mathrm{M}_\odot/h}{(0.978\mathrm{Gyr}/h)}$	Sum of the BH_Mdot field of all black holes (type 5) in this group.
GroupCM	N,3	ckpc/h	Center of mass of the group, computed as the sum of the mass weighted relative coordinates of all particles/cells in the group, of all types. Comoving coordinate. (Available only for the Illustris-3 run)
GroupFirstSub	N	-	Index into the Subhalo table of the first/primary/most massive SUBFIND group within this FoF group (int32).
GroupGasMetallicity	N	-	Mass-weighted average metallicity (Mz/Mtot, where Z = any element above He) of all gas cells in this FOF group.
GroupLen	N	-	Integer counter of the total number of particles/cells of all types in this group (int32).
GroupLenType	N,6	-	Integer counter of the total number of particles/cells, split into the six different types, in this group. Note: Wind phase cells are counted as stars (type 4) for GroupLenType (int32).
GroupMass	N	$10^{10}\mathrm{M}_\odot/h$	Sum of the individual masses of every particle/cell, of all types, in this group.
GroupMassType	N,6	$10^{10}\mathrm{M}_\odot/h$	Sum of the individual masses of every particle/cell, split into the six different types, in this group. Note: Wind phase cells are counted as gas (type 0) for GroupMassType.
GroupNsubs	N	-	Count of the total number of SUBFIND groups within this FoF group (int32).
GroupPos	N,3	ckpc/h	Spatial position within the periodic box of size 75000 ckpc/h of the maximum bound particle. Comoving coordinate.
GroupSFR	N	$\mathrm{M}_\odot$/yr	Sum of the individual star formation rates of all gas cells in this group.
GroupStarMetallicity	N	-	Mass-weighted average metallicity (Mz/Mtot, where Z = any element above He) of all star particles in this FOF group.
GroupVel	N,3	km/s/a	Velocity of the group, computed as the sum of the mass weighted velocities of all particles/cells in this group, of all types. The peculiar velocity is obtained by multiplying this value by $1/a$.
GroupWindMass	N	$10^{10}\mathrm{M}_\odot/h$	Sum of the individual masses of all wind phase gas cells (type 4, BirthTime $<= 0$) in this group.
Group_M_Crit200	N	$10^{10}\mathrm{M}_\odot/h$	Total mass of this group enclosed in a sphere whose mean density is 200 times the critical density of the Universe, at the time the halo is considered.
Group_M_Crit500	N	$10^{10}\mathrm{M}_\odot/h$	Likewise, but for 500 times the critical density of the Universe.
Group_M_Mean200	N	$10^{10}\mathrm{M}_\odot/h$	Likewise, but for 200 times the mean density of the Universe.
Group_M_TopHat200	N	$10^{10}\mathrm{M}_\odot/h$	Likewise, but for Δ_c times the critical density of the Universe, where Δ_c derives from the solution of the collapse of a spherical top-hat perturbation (fitting formula from Bryan and Norman (1998)). The subscript 200 can be ignored.
Group_R_Crit200	N	ckpc/h	Comoving radius of a sphere centered at the GroupPos of this Group whose mean density is 200 times the critical density of the Universe, at the time the halo is considered.
Group_R_Crit500	N	ckpc/h	Likewise, but for 500 times the critical density of the Universe.
Group_R_Mean200	N	ckpc/h	Likewise, but for 200 times the mean density of the Universe.
Group_R_TopHat200	N	ckpc/h	Likewise, but for Δ_c times the critical density of the Universe.

Table B.2: Description of all fields in the SUBFIND subhalo catalogs (Part I). All fields are float32 unless otherwise specified.

Field	Dimensions	Units	Description
SubhaloBHMass	N	$10^{10}\mathrm{M}_\odot/h$	Sum of the masses of all black holes in this subhalo.
SubhaloBHMdot	N	$\frac{10^{10}\mathrm{M}_\odot/h}{0.978\mathrm{Gyr}/h}$	Sum of the instantaneous accretion rates $\dot{M}$ of all black holes in this subhalo.
SubhaloCM	N,3	ckpc/h	Comoving center of mass of the Subhalo, computed as the sum of the mass weighted relative coordinates of all particles/cells in the Subhalo, of all types.
SubhaloGasMetallicity	N	-	Mass-weighted average metallicity (Mz/Mtot, where Z = any element above He) of the gas cells bound to this Subhalo, but restricted to cells within twice the stellar half mass radius.
SubhaloGasMetallicityHalfRad	N	-	Same as SubhaloGasMetallicity, but restricted to cells within the stellar half mass radius.
SubhaloGasMetallicityMaxRad	N	-	Same as SubhaloGasMetallicity, but restricted to cells within the radius of V_{max}.
SubhaloGasMetallicitySfr	N	-	Mass-weighted average metallicity (Mz/Mtot, where Z = any element above He) of the gas cells bound to this Subhalo, but restricted to cells which are star forming.
SubhaloGasMetallicitySfrWeighted	N	-	Same as SubhaloGasMetallicitySfr, but weighted by the cell star-formation rate rather than the cell mass.
SubhaloGrNr	N	-	Index into the Group table of the FOF host/parent of this Subhalo (int32).
SubhaloHalfmassRad	N	ckpc/h	Comoving radius containing half of the total mass (SubhaloMass) of this Subhalo.
SubhaloHalfmassRadType	N,6	ckpc/h	Comoving radius containing half of the mass of this Subhalo split by Type (SubhaloMassType).
SubhaloIDMostbound	N	-	The ID of the particle with the smallest binding energy (could be any type, int64).
SubhaloLen	N	-	Total number of member particle/cells in this Subhalo, of all types (int32).
SubhaloLenType	N,6	-	Total number of member particle/cells in this Subhalo, separated by type (int32).
SubhaloMass	N	$10^{10}\mathrm{M}_\odot/h$	Total mass of all member particle/cells which are bound to this Subhalo, of all types.
SubhaloMassInHalfRad	N	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all particles/cells within the stellar half mass radius.
SubhaloMassInHalfRadType	N,6	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all particles/cells (split by type) within the stellar half mass radius.
SubhaloMassInMaxRad	N	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all particles/cells within the radius of V_{max}.
SubhaloMassInMaxRadType	N,6	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all particles/cells (split by type) within the radius of V_{max}.
SubhaloMassInRad	N	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all particles/cells within twice the stellar half mass radius.
SubhaloMassInRadType	N,6	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all particles/cells (split by type) within twice the stellar half mass radius.

Table B.3: Description of all fields in the SUBFIND subhalo catalogs (Part II). All fields are float32 unless otherwise specified. Note that for all mass calculations by type, wind phase cells are counted as gas.

Field	Dimensions	Units	Description
SubhaloMassType	N,6	$10^{10}\mathrm{M}_\odot/h$	Total mass of all member particle/cells which are bound to this Subhalo, separated by type.
SubhaloParent	N	-	Index into the Subhalo table of the unique SUBFIND parent of this Subhalo (int32).
SubhaloPos	N,3	ckpc/h	Spatial position within the periodic box of size 75000 ckpc/h of the maximum bound particle. Comoving coordinate.
SubhaloSFR	N	$\mathrm{M}_\odot$/yr	Sum of the individual star formation rates of all gas cells in this subhalo.
SubhaloSFRinHalfRad	N	$\mathrm{M}_\odot$/yr	Same as SubhaloSFR, but restricted to cells within the stellar half mass radius.
SubhaloSFRinMaxRad	N	$\mathrm{M}_\odot$/yr	Same as SubhaloSFR, but restricted to cells within the radius of V_{max}.
SubhaloSFRinRad	N	$\mathrm{M}_\odot$/yr	Same as SubhaloSFR, but restricted to cells within twice the stellar half mass radius.
SubhaloSpin	N,3	(kpc/h)(km/s)	Total spin per axis, computed for each as the mass weighted sum of the relative coordinate times relative velocity of all member particles/cells.
SubhaloStarMetallicity	N	-	Mass-weighted average metallicity (Mz/Mtot, where Z = any element above He) of the star particles bound to this Subhalo, but restricted to stars within twice the stellar half mass radius.
SubhaloStarMetallicityHalfRad	N	-	Same as SubhaloStarMetallicity, but restricted to stars within the stellar half mass radius.
SubhaloStarMetallicityMaxRad	N	-	Same as SubhaloStarMetallicity, but restricted to stars within the radius of V_{max}.
SubhaloStellarPhotometrics	N,8	mag	Eight bands: U, B, V, K, g, r, i, z. Magnitudes based on the summed-up luminosities of all the stellar particles of the group. For details on the bands, see snapshot details.
SubhaloStellarPhotometrics MassInRad	N	$10^{10}\mathrm{M}_\odot/h$	Sum of the mass of the member stellar particles, but restricted to stars within the radius SubhaloStellarPhotometricsRad.
SubhaloStellarPhotometricsRad	N	ckpc/h	Radius at which the surface brightness profile (computed from all member stellar particles) drops below the limit of 20.7 mag arcsec^{-2} in the K band (in comoving units).
SubhaloVel	N,3	km/s	Peculiar velocity of the group, computed as the sum of the mass weighted velocities of all particles/cells in this group, of all types.
SubhaloVelDisp	N	km/s	One-dimensional velocity dispersion of all the member particles/cells (the 3D dispersion divided by $\sqrt{3}$).
SubhaloVmax	N	km/s	Maximum value of the spherically-averaged rotation curve.
SubhaloVmaxRad	N	kpc/h	Comoving radius of rotation curve maximum (where V_{max} is achieved).
SubhaloWindMass	N	$10^{10}\mathrm{M}_\odot/h$	Sum of masses of all wind-phase cells in this subhalo (with Type==4 and BirthTime<= 0).

Table B.4: Description of all fields in the Header group of the group catalog files. Each header field is an attribute.

Field	Type	Description
SimulationName	string	e.g. 'Illustris-1' or 'Illustris-2-Dark'
SnapshotNumber	int	snapshot number (should be consistent with filename)
Ngroups_ThisFile	int	Number of groups within this file chunk.
Nsubgroups_ThisFile	int	Number of subgroups within this file chunk.
Ngroups_Total	int	Total number of groups for this snapshot.
Nsubgroups_Total	int	Total number of subgroups for this snapshot.
NumFiles	int	Total number of file chunks the group catalog is split between.
Num_ThisFile	int	Index of this file chunk (should be consistent with the filename).
Time	float	Scale factor of the snapshot corresponding to this group catalog.
Redshift	float	Redshift of the snapshot corresponding to this group catalog.
BoxSize	float	Side-length of the periodic volume in code units.
FileOffsets_Snap	$[N_\mathrm{c}, 6]$ int array	The offset table (by type) for the snapshot files, giving the first particle index in each snap file chunk. Determines which files(s) a given offset+length will cover. A two-dimensional array, where the element (i, j) equals the cumulative sum (i.e. offset) of particles of type i in all snapshot file chunks prior to j.
FileOffsets_Group	$[N_\mathrm{c}]$ int array	The offset table for groups in the group catalog files. A one-dimensional array, where the i^{th} element equals the first group number in the i^{th} groupcat file chunk.
FileOffsets_Subhalo	$[N_\mathrm{c}]$ int array	The offset table for subhalos in the group catalog files. A one-dimensional array, where the i^{th} element equals the first subgroup number in the i^{th} groupcat file chunk.
FileOffsets_SubLink	$[N_\mathrm{c}]$ int array	The offset table for trees in the SubLink files. A one-dimensional array, where the i^{th} element equals the first tree number in the i^{th} SubLink file chunk.

Table B.5: Description of all fields in the Offsets group of the group catalog files. Note that all three LHaloTree or SubLink values equal -1 if that subhalo is not in the respective merger tree, which can occur if searching at a snapshot prior to $z = 0$. For the offsets, N_c indicates the number of file chunks (or pieces) over which that data product has been split.

Field	Dimensions	Description
Group_SnapByType	Ngroups_Total,6	The offset table for a given group number (by type), into the snapshot files. That is, the global particle index (across all snap file chunks) of the first particle of this group. A two-dimensional array, where the element (i, j) equals the cumulative sum (i.e. offset) of particles of type i in all groups prior to group number j.
Group_FuzzByType	Ngroups_Total,6	Offset into the "outer fuzz" (at the end of each snapshot file) for this group.
Subhalo_SnapByType	Nsubgroups_Total,6	The offset table for a given subhalo number (by type), into the snapshot files. That is, the global particle index (across all snap file chunks) of the first particle of this subhalo. A two-dimensional array, where the element (i, j) equals the cumulative sum (i.e. offset) of particles of type i in all subhalos prior to subhalo number j.
Subhalo_LHaloTreeFile	Nsubgroups_Total	The LHaloTree file number with the tree which contains this subhalo.
Subhalo_LHaloTreeNum	Nsubgroups_Total	The number of the tree within the above file within which this subhalo is located (e.g. TreeX).
Subhalo_LHaloTreeIndex	Nsubgroups_Total	The LHaloTree index within the above tree dataset at which this subhalo is located.
Subhalo_SublinkRowNum	Nsubgroups_Total	The SubLink global index of the location of this subhalo.
Subhalo_SublinkSubhaloID	Nsubgroups_Total	The SubLink ID of this subhalo.
Subhalo_SublinkLastProgenitorID	Nsubgroups_Total	The SubLink ID of the last progenitor of this tree (all the subhalos contained in the tree rooted in this subhalo are the ones with IDs between SubhaloID and LastProgenitorID).

Table B.6: Listing of all fields and their descriptions for the SubLink merger trees. Note that in addition to the tree fields, all subhalo fields are also present, copied exactly from the Subfind catalogs. The advantage is that they are ordered in the same order as the tree structure. See the group catalog description for their units and descriptions. The Group_M_Crit200, Group_M_Mean200, and Group_M_Tophat200 fields are also present, but are FoF group quantities, such that all subhalos in the same FOF group will have the same value for these three fields.

Field	Type	Description
SubhaloID	int64	Unique identifier of this subhalo, assigned in a "depth-first" fashion (Lemson and Virgo Consortium, 2006). This value is contiguous within a single tree.
SubhaloIDRaw	int64	Unique identifier of this subhalo in raw format (= SnapNum$\times 10^{12}$ + SubfindID).
LastProgenitorID	int64	The SubhaloID of the last progenitor of the tree rooted at this subhalo. Since the SubhaloIDs are assigned in a "depth-first" fashion, all the subhalos contained in the tree rooted at this subhalo are the ones with SubhaloIDs between (and including) the SubhaloID and LastProgenitorID of this subhalo. For subhalos with no progenitors, LastProgenitorID == SubhaloID.
MainLeafProgenitorID	int64	The SubhaloID of the last progenitor along the main branch, i.e. the earliest progenitor obtained by following the FirstProgenitorID pointer. For subhalos with no progenitors, MainLeafProgenitorID == SubhaloID.
RootDescendantID	int64	The SubhaloID of the latest subhalo that can be reached by following the DescendantID link, i.e. the root of the tree to which this subhalo belongs. For subhalos with no descendants, RootDescendantID == SubhaloID.
TreeID	int64	Unique identifier of the tree to which this subhalo belongs.
SnapNum	int16	The snapshot in which this subhalo is found.
FirstProgenitorID	int64	The SubhaloID of this subhalo's first progenitor. The first progenitor is the one with the "most massive history" behind it. For subhalos with no progenitors, FirstProgenitorID == -1.
NextProgenitorID	int64	The SubhaloID of the subhalo with the next most massive history which shares the same descendant as this subhalo. If there are no more subhalos sharing the same descendant, NextProgenitorID == -1.
DescendantID	int64	The SubhaloID of this subhalo's descendant. If this subhalo has no descendants, DescendantID == -1.
FirstSubhaloInFOFGroupID	int64	The SubhaloID of the first subhalo (i.e., the one with the most massive history) from the same FOF group.
NextSubhaloInFOFGroupID	int64	The SubhaloID of the next subhalo (ordered by their mass history) from the same FOF group. If there are no more subhalos in the same FOF group, NextSubhaloInFOFGroupID == -1.
NumParticles	uint32	Number of particles in the current subhalo which were used in the merger tree to determine descendants (e.g. DM-only or stars + star-forming gas).
Mass	float32	Mass of the current subhalo, including only the particles which were used in the merger tree to determine descendants (e.g. DM-only or stars + star-forming gas), in units of $10^{10}\mathrm{M}_\odot/h$.
MassHistory	float32	Sum of the Mass field of all progenitors along the main branch (De Lucia and Blaizot, 2007), in units of $10^{10}\mathrm{M}_\odot/h$.
SubfindID	int32	Index of this subhalo in the Subfind group catalog.

Table B.7: Listing of all fields in the LHaloTree merger trees. Note that in addition to the tree fields, the majority of subhalo fields are also present, copied exactly from the Subfind catalogs. The advantage is that they are ordered in the same order as the tree structure. See the group catalog description for their units and descriptions. The Group_M_Crit200, Group_M_Mean200, and Group_M_Tophat200 fields are also present, but since they are FoF group quantities, all subhalos from the same FOF group will have the same value for these three fields.

Field	Dimensions	Description
		Header Groups
Redshifts	{N_snap}	List of redshifts of the snapshots used to create this merger tree.
TotNsubhalos	{N_snap}	Equal to the number of Subfind groups in the group catalog, for each snapshot used to create this merger tree.
TreeNHalos	{N_halos}	The size of {N} for each TreeX group in this file, e.g. the total number of halos (across time) in that group.
FirstSnapshotNr	1	First snapshot number used to make these merger trees (should be 0).
LastSnapshotNr	1	Last snapshot number used to make these merger trees (should be 135).
SnapSkipFac	1	Snapshot stride when making these merger trees (should be 1).
NtreesPerFile	1	The size of {N_halos} for this file, can be used to calculate the offset to map a FoF group number to a TreeX group name (made to be roughly equal across chunks).
NhalosPerFile	1	The total number of tree members (subhalos) in this file. Equals the sum of all elements of TreeNHalos.
ParticleMass	1	The dark matter particle mass used to make these merger trees, in units of $10^{10} \mathrm{M}_\odot/h$.
		TreeX Groups
SubhaloNumber	(N)	The ID of this subhalo, unique within the full simulation for this snapshot. Indexes the Subfind group catalog at SnapNum.
Descendant	(N)	The index of the subhalo's descendant within the merger tree, if any (-1 otherwise). Indexes this TreeX group.
FirstProgenitor	(N)	The index of the subhalo's first progenitor within the merger tree, if any (-1 otherwise). The first progenitor is defined as the most massive one. (-1 if none) Indexes this TreeX group.
NextProgenitor	(N)	The index of the next subhalo from the same snapshot which shares the same descendant, if any (-1 if this is the last). Indexes this TreeX group.
FirstHaloInFOFGroup	(N)	The index of the main subhalo (i.e. the most massive one) from the same FOF group. Indexes this TreeX group.
NextHaloInFOFGroup	(N)	The index of the next subhalo from the same FOF group (-1 if this is the last). Indexes this TreeX group.
FileNr	(N)	File number in which the subhalo is found. Redundant, i.e. for a given [chunkNum] file, this array will be constant and equal to [chunkNum].
SnapNum	(N)	The snapshot in which this subhalo was found.

Appendix C: Supplementary Data Details

Table C.1: Details of the supplementary data catalog: Photometric Non-Parametric Stellar Morphologies. The four bands which replace band_name are: gSDSS, iSDSS, uSDSS, and hWFC3 (WFC3-IR/F160W). The four camera views are indexed 0, 1, 2, and 3.

Group Name	Units	Description
/Snapshot_135/SubfindID_cam0,1,2,3	-	The SUBFIND IDs these values correspond to (different for each camera view, but the same for all bands and fields). 10654,10618,10639,10620 entries.
/Snapshot_135/band_name/Gini_cam0,1,2,3	-	The G_ini coefficient, which measures the relative distribution of the galaxy pixel flux values.
/Snapshot_135/band_name/M20_cam0,1,2,3	-	M_{20}, the second-order moment of the brightest 20% of the galaxy's flux.
/Snapshot_135/band_name/C_cam0,1,2,3	-	The concentration parameter C.
/Snapshot_135/band_name/RP_cam0,1,2,3	kpc	The elliptical Petrosian radius r_P.
/Snapshot_135/band_name/RE_cam0,1,2,3	kpc	The elliptical half-light radius r_E.

Table C.2: Details of the supplementary data catalog: Stellar Circularities, Angular Momenta, and Axis Ratios. Note that, in addition to these values which are measured within $10R_E$, several fields are also computed including all stars in the subhalo, and are available as the "_allstars" datasets.

Group Name	Units	Description
/Snapshot_N/ SubfindID	-	The SUBFIND IDs these values correspond to (27345 entries).
/Snapshot_N/ SpecificAngMom	km/s × kpc	The specific angular momentum of the stars.
/Snapshot_N/ CircAbove07Frac	-	The fraction of stars with $\epsilon > 0.7$. This is a common definition of the disk stars - those with significant (positive) rotational support.
/Snapshot_N/ CircAbove07 MinusBelowNeg07Frac	-	The fraction of stars with $\epsilon > 0.7$ minus the fraction of stars with $\epsilon < -0.7$. This removes the contribution of the bulge to the disk, assuming the bulge is symmetric around $\epsilon = 0$.
/Snapshot_N/ CircTwiceBelow0Frac	-	The fraction of stars with $\epsilon < 0$, multiplied by two. This is another common way in the literature to define the bulge.
/Snapshot_N/ MassTensorEigenVals	kpc	Three numbers for each galaxy, which are the eigenvalues of the mass tensor of the stellar mass inside the stellar $2R_{1/2}$. This means that in a coordinate system that is aligned with the eigenvectors (principal axes), the component i equals $M_i \equiv \sqrt{\sum_j m_j r_{j,i}^2} / \sqrt{\sum_j m_j}$, where j enumerates over stellar particles inside that radius, $r_{j,i}$ is the distance of stellar particle j in the i axis from the most bound particle of the galaxy, and m_j is its mass, and $i \in (1, 2, 3)$. They are sorted such that $M_1 < M_2 < M_3$. Example use: $M_1/\sqrt{M_2 M_3}$ can represent the flatness of the galaxy.
/Snapshot_N/ ReducedMass TensorEigenVals	-	Similar to the above, except less weight is given to further away particles. The orientation of the system is the same, but the quantity measured for each axis is instead $M_i \equiv \sqrt{\sum_j m_j r_{j,i}^2 / R_j^2} / \sqrt{\sum_j m_j}$, where $R_j \equiv \sum_i r_{j,i}^2$ is the distance of star j from the centre of the galaxy.

Appendix D: API Examples and Reference

To be explicit by way of example, the following are absolute URLs for the Illustris API covering some of its functionality, where the type of the request should be clear from the preceding documentation.

- `http://www.illustris-project.org/api/Illustris-2/`
- `http://www.illustris-project.org/api/Illustris-2/snapshots/68/`
- `http://www.illustris-project.org/api/Illustris-1/snapshots/135/subhalos/73664/`
- `http://www.illustris-project.org/api/Illustris-1/snapshots/135/subhalos/73664/stellar_mocks/broadband.fits`
- `http://www.illustris-project.org/api/Illustris-1/snapshots/135/subhalos/73664/stellar_mocks/sed.txt`
- `http://www.illustris-project.org/api/Illustris-1/snapshots/80/halos/523312/cutout.hdf5?dm=Coordinates&gas=all`
- `http://www.illustris-project.org/api/Illustris-3/snapshots/135/subhalos?mass__gt=10.0&mass__lt=20.0`
- `http://www.illustris-project.org/api/Illustris-2/snapshots/68/subhalos/50000/sublink/full.hdf5`
- `http://www.illustris-project.org/api/Illustris-2/snapshots/68/subhalos/50000/sublink/mpb.json`
- `http://www.illustris-project.org/api/Illustris-1/files/groupcat-135.5.hdf5`
- `http://www.illustris-project.org/api/Illustris-2/files/snapshot-135.10.hdf5`
- `http://www.illustris-project.org/api/Illustris-2/files/snapshot-135.10.hdf5?dm=all`
- `http://www.illustris-project.org/api/Illustris-3/files/sublink.2.hdf5`

In the online documentation we provide a complete getting started guide for the web-based API, as well as a cookbook of common tasks, in Python, IDL, and Matlab. Here we include just four examples taken from that documentation, and only in Python, to give a flavor of the approach. The task numbers are taken from the online version.

Task 0: First, we define a helper function, to make the HTTP response, and check for errors. If the response is JSON, automatically parse it. If the response is binary data, automatically save it to a file.

```
>>> def get(path, params=None):
>>>     # make HTTP GET request to path
>>>     headers = {"api-key":"INSERT_API_KEY_HERE"}
>>>     r = requests.get(path, params=params, headers=headers)
>>>
>>>     # raise exception if response code is not HTTP SUCCESS (200)
>>>     r.raise_for_status()
>>>
>>>     if r.headers['content-type'] == 'application/json':
>>>         return r.json() # parse json responses automatically
>>>
>>>     if 'content-disposition' in r.headers:
>>>         filename = r.headers['content-disposition'].split("filename=")[1]
>>>         with open(filename, 'wb') as f:
>>>             f.write(r.content)
>>>         return filename # return the filename string
```

Task 1: For Illustris-1 at $z = 0$, get all the fields available for the subhalo with id=0 and print its total mass and stellar half mass radius.

```
>>> url = "http://www.illustris-project.org/api/Illustris-1/snapshots/135/subhalos/0/"
>>> r = get(url)
>>> r['mass']
22174.8

>>> r['halfmassrad_stars']
12.395
```

Task 2: For Illustris-1 at $z = 2$, search for all subhalos with total mass $10^{11.9}\mathrm{M}_\odot < M < 10^{12.1}\mathrm{M}_\odot$, print the number returned, and the SUBFIND IDs of the first five results.

```
>>> # first convert log solar masses into group catalog units
>>> mass_min = 10**11.9 / 1e10 * 0.704
>>> mass_max = 10**12.1 / 1e10 * 0.704
>>>
>>> params = {'mass__gt':mass_min, 'mass__lt':mass_max}
>>>
>>> # make the request
>>> url = "http://www.illustris-project.org/api/Illustris-1/snapshots/z=2/subhalos/"
>>> subhalos = get(url, params)
>>> subhalos['count']
550

>>> ids = [ subhalos['results'][i]['id'] for i in range(5) ]
>>> ids
[1, 1352, 5525, 6574, 12718]
```

Task 8: For Illustris-1 at $z = 2$, for five specific SUBFIND IDs (from above: 1, 1352, 5525, 6574, 12718), locate the $z = 0$ descendant of each by using the API to walk down the SUBLINK descendant links.

```
>>> ids = [1, 1352, 5525, 6574, 12718]
>>> z0_descendant_ids = [-1]*len(ids)
>>>
>>> for i,id in enumerate(ids):
>>>     start_url = "http://www.illustris-project.org/api/Illustris-1/snapshots/z=2/subhalos/"
>>>     start_url += str(id)
>>>     sub = get(start_url)
>>>
>>>     while sub['desc_sfid'] != -1:
>>>         # request the full subhalo details of the descendant by following the sublink URL
>>>         sub = get(sub['related']['sublink_descendant'])
>>>         if sub['snap'] == 135:
>>>             z0_descendant_ids[i] = sub['id']
>>>
>>>     if z0_descendant_ids[i] >= 0: # note: possible that descendant branch did not reach z=0
>>>         print 'Descendant of ' + str(id) + ' at z=0 is ' + str(z0_descendant_ids[i])

Descendant of 1 at z=0 is 30465
Descendant of 1352 at z=0 is 41396
Descendant of 5525 at z=0 is 99148
Descendant of 6574 at z=0 is 51811
Descendant of 12718 at z=0 is 194303
```

Task 11: Download the entire Illustris-1 $z = 0$ snapshot including *only the positions, masses, and metallicities of stars* (in the form of 512 HDF5 files). In this example, since we only need these three fields for stars only, we can reduce the download and storage size from ∼1.5 TB to ∼17 GB.

```
>>> base_url = "http://www.illustris-project.org/api/Illustris-1/"
>>> sim_metadata = get(base_url)
>>> params = {'stars':'Coordinates,Masses,GFM_Metallicity'}
>>>
>>> for i in range(sim_metadata['num_files_snapshot']):
>>>     file_url = base_url + "files/snapshot-135." + str(i) + ".hdf5"
>>>     saved_filename = get(file_url, params)
>>>     print saved_filename
```

Table D.1: API Endpoint Descriptions and Reference (I): simulation and snapshot meta-data, subhalos and halos, merger trees.

Endpoint	Description	Return Type
/api/	list all simulations currently accessible to the user	json,api (?format=)
/api/{sim_name}/	list metadata (including list of all snapshots+redshifts) for {sim_name}	json,api (?format=)
/api/{sim_name}/ snapshots/	list all snapshots which exist for this simulation	json,api (?format=)
/api/{sim_name}/ snapshots/{num}/	list metadata for snapshot {num} of simulation {sim_name}	json,api (?format=)
/api/{sim_name}/ snapshots/z={redshift}/	redirect to the snapshot which exists closest to {redshift} (with a maximum allowed error of 0.1 in redshift)	json,api (?format=)
define [base] = /api/{sim_name}/snapshots/{num} or [base] = /api/{sim_name}/snapshots/z={redshift}		
(after selection of a particular simulation and snapshot)		
Subfind Subhalos		
[base]/subhalos/	paginated list of all subhalos for this snapshot of this run	json,api (?format=)
[base]/subhalos/ ?{search_query}	execute {search_query} over all subhalos, return those satisfying the search with basic fields and links to /subhalos/{id}	json,api (?format=)
[base]/subhalos/ {id}	list available data fields and links to all queries possible on SUBFIND subhalo {id}	json,api (?format=)
[base]/subhalos/ {id}/info.json	extract all group catalog fields for subhalo {id}	json (.ext)
[base]/subhalos/ {id}/cutout.hdf5	return snapshot cutout of subhalo {id}, all particle types and fields	HDF5 (.ext)
[base]/subhalos/ {id}/cutout.hdf5 ?{cutout_query}	return snapshot cutout of subhalo {id} corresponding to the {cutout_query}	HDF5 (.ext)
FoF Halos		
[base]/halos/{halo_id}/	list what we know about this FoF halo, in particular the 'child_subhalos'	json,api (?format=)
[base]/halos/{halo_id}/ info.json	extract all group catalog fields for halo {halo_id}	json (.ext)
[base]/halos/{halo_id}/ cutout.hdf5	return snapshot cutout of halo {halo_id}, all particle types and fields	HDF5 (.ext)
[base]/halos/{halo_id}/ cutout.hdf5?{cutout_query}	return snapshot cutout of halo {halo_id} corresponding to the {cutout_query}	HDF5 (.ext)
Merger Trees		
[base]/subhalos/{id}/ lhalotree/full.hdf5	retrieve full tree (flat HDF5 format or hierchical/nested JSON format)	HDF5,json (.ext)
[base]/subhalos/{id}/ lhalotree/mpb.hdf5	retrieve only main progenitor branch (towards higher redshift for this subhalo)	HDF5,json (.ext)
[base]/subhalos/{id}/ sublink/full.hdf5	same as above for 'lhalotree' but for sublink	HDF5,json (.ext)
[base]/subhalos/{id}/ sublink/mpb.hdf5	same as above for 'lhalotree' but for sublink	HDF5,json (.ext)

Table D.2: API Endpoint Descriptions and Reference (II): supplementary data catalogs, file downloads.

Endpoint	Description	Return Type
supplementary data: stellar mocks		
[base]/subhalos/{id}/stellar_mocks/broadband.fits	download raw broadband fits file for subhalo {id}	FITS (.ext)
[base]/subhalos/{id}/stellar_mocks/broadband.hdf5?view={view}	download subset of broadband fits file for subhalo {id}: all 36 bands for view number {view}	HDF5 (.ext)
[base]/subhalos/{id}/stellar_mocks/broadband.hdf5?band=band	download subset of broadband fits file for subhalo {id}: all 4 views for band {band} (1-indexed number, or name)	HDF5 (.ext)
[base]/subhalos/{id}/stellar_mocks/image.png	download stellar mock png 2D image (subhalo particles only)	PNG (.ext)
[base]/subhalos/{id}/stellar_mocks/image_fof.png	download stellar mock png 2D image (all group particles)	PNG (.ext)
[base]/subhalos/{id}/stellar_mocks/image_gz.png	download stellar mock png 2D image ('galaxy zoo' image w/ realistic noise and background)	PNG (.ext)
[base]/subhalos/{id}/stellar_mocks/sed.txt	download stellar mock integrated 1D SED for subhalo {id}	txt,json (.ext)
direct file downloads		
define [base] = /api/sim_name/files		
[base]/	list of each 'files' type available for this simulation (excluding those attached to specific snapshots)	json,api (?format=)
[base]/snapshot-{num}/	list of all the actual file chunks to download snapshot {num}	json,api (?format=)
[base]/snapshot-{num}.{chunknum}.hdf5	download chunk {chunknum} of snapshot {num}	HDF5 (.ext)
[base]/snapshot-{num}.{chunknum}.hdf5?{cutout_query}	download only {cutout_query} of chunk {chunknum} of snapshot {num}	HDF5 (.ext)
[base]/groupcat-{num}/	list of all the actual file chunks to download group catalog (fof/subfind) for snapshot {num}	json,api (?format=)
[base]/groupcat-{num}.{chunknum}.hdf5	download chunk {chunknum} of group catalog for snapshot {num}	HDF5 (.ext)
[base]/lhalotree/	list of all the actual file chunks to download LHaloTree merger tree for this simulation	json,api (?format=)
[base]/lhalotree.{chunknum}.hdf5	download chunk {chunknum} of LHaloTree merger tree for this simulation	HDF5 (.ext)
[base]/sublink/	list of all the actual file chunks to download SubLink merger tree for this simulation	json,api (?format=)
[base]/sublink.{chunknum}.hdf5	download chunk {chunknum} of SubLink merger tree for this simulation	HDF5 (.ext)