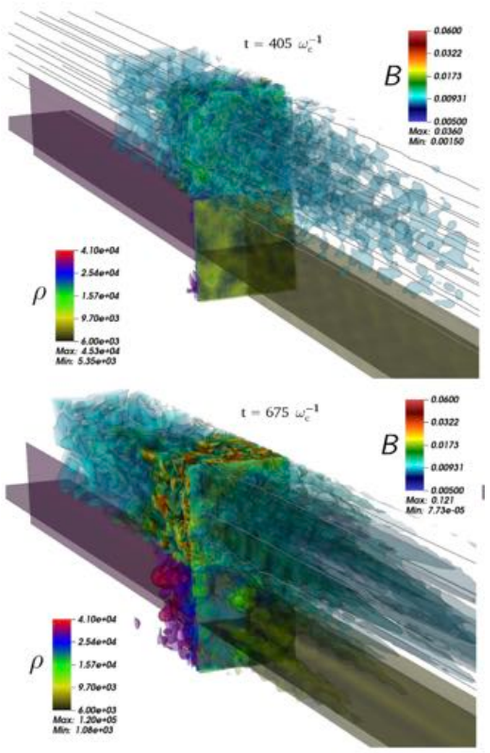
~~Le point de vue de l'Action Spécifique Numérique~~

L'ASN est encore jeune et le point de vue de l'ASN sur cette question n'est pas encore formalisé

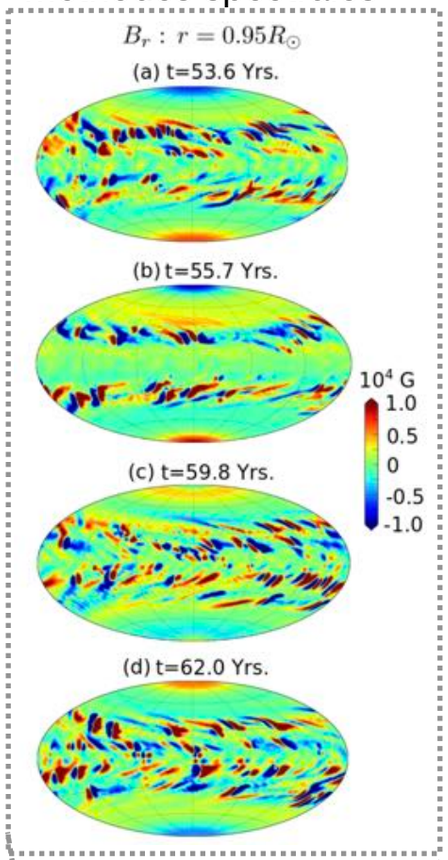
Yohan Dubois
Institut d'Astrophysique de Paris

Accélération rayons cosmiques u~c_{light} dans des chocs. Code particle-in-cell



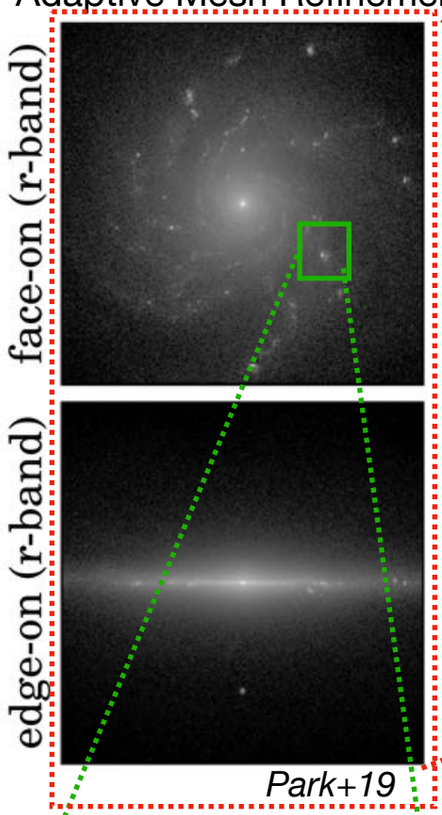
van Marle+19

Evolution champ magnétique solaire Méthodes spectrales



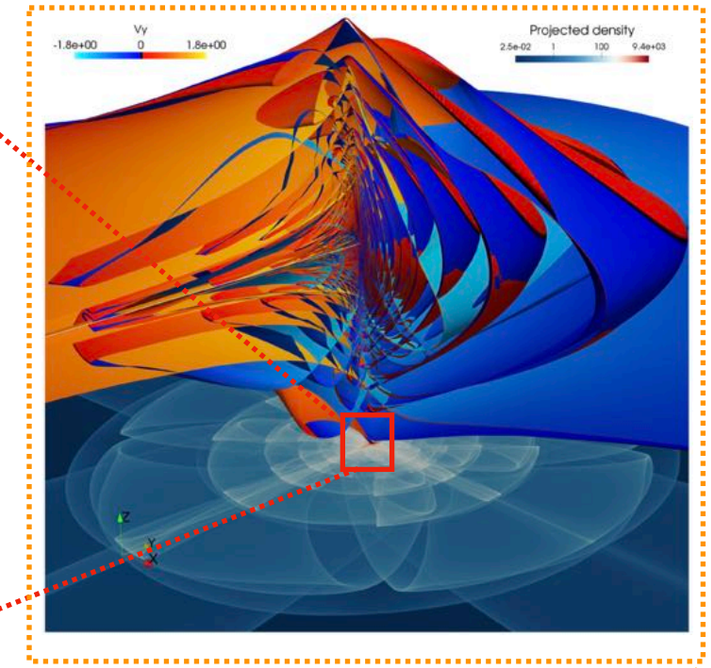
Kumar+19

Evolution des galaxies Adaptive Mesh Refinement



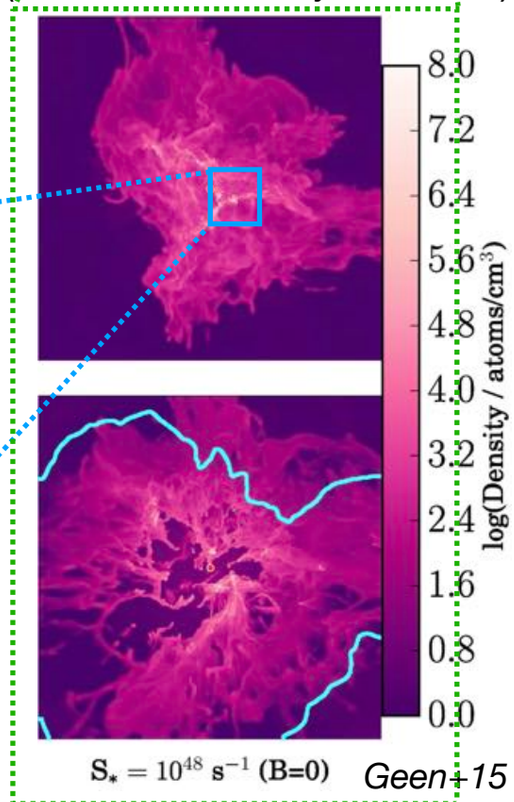
Park+19

Effondrement gravitationnel halo matière noire Vlasov-Poisson 6D



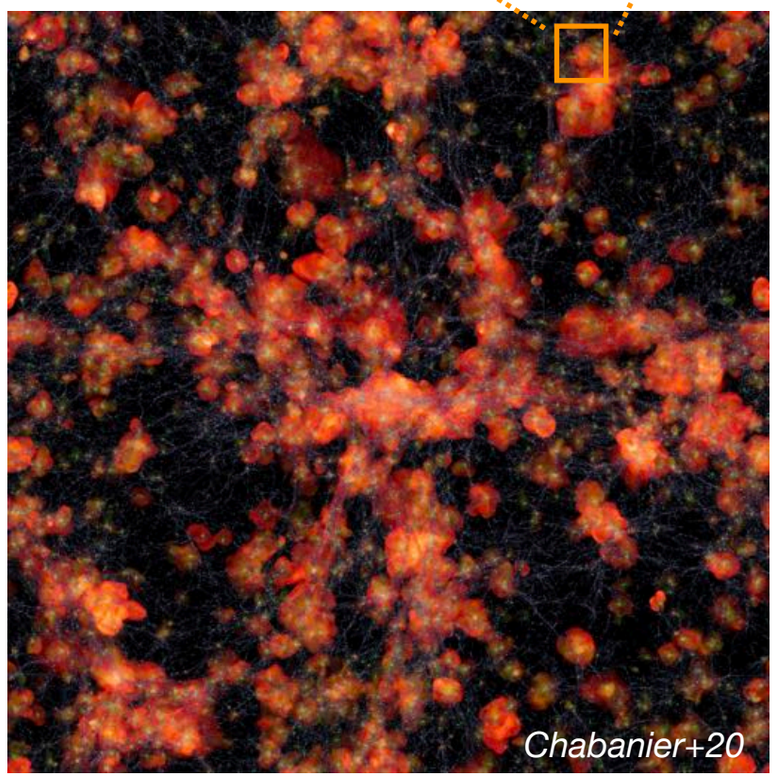
Sousbie+16

Evolution nuage à formation d'étoiles Adaptive Mesh Refinement (MHD+Transfert rayonnement)



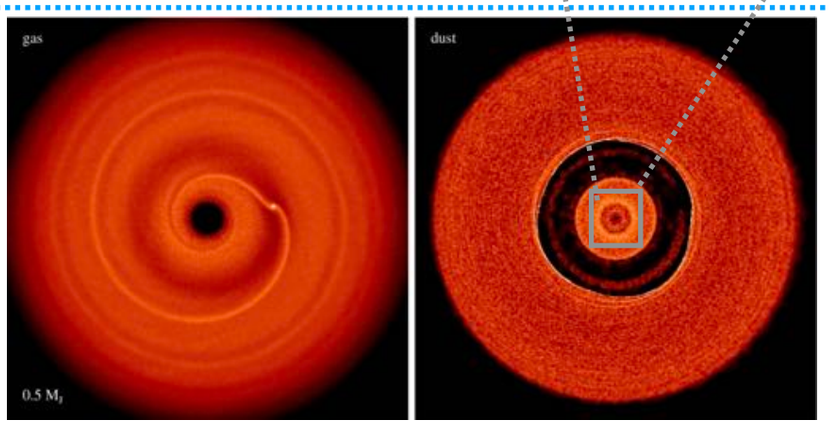
Geen+15

Formation de la toile cosmique et galaxies Adaptive Mesh Refinement



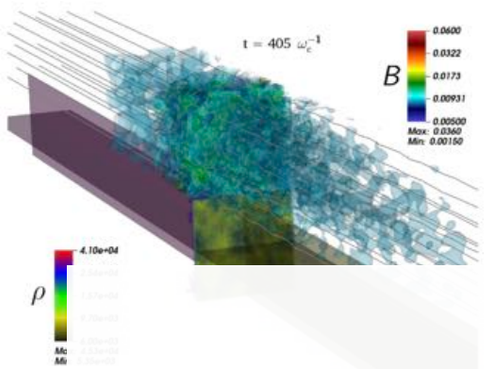
Chabanier+20

Formation planètes dans disques proto-stellaires Smoothed Particles Hydrodynamics

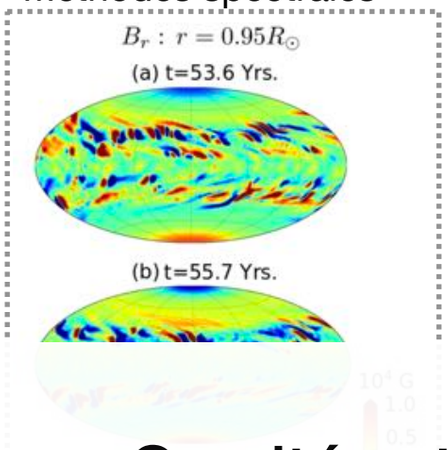


Dipierro+16

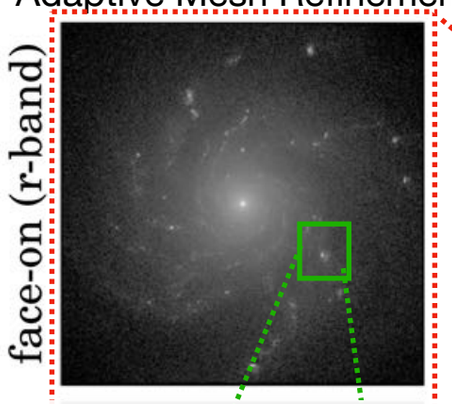
Accélération rayons cosmiques u~c_{light} dans des chocs. Code particle-in-cell



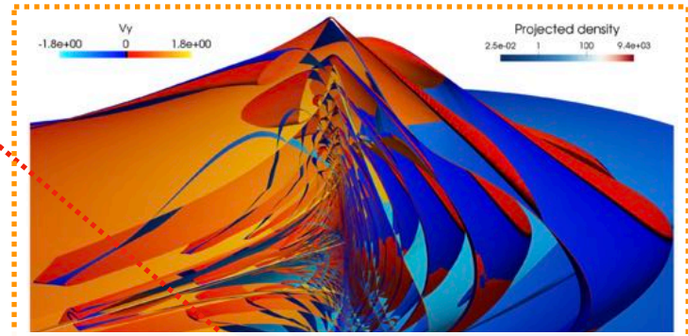
Evolution champ magnétique solaire Méthodes spectrales



Evolution des galaxies Adaptive Mesh Refinement



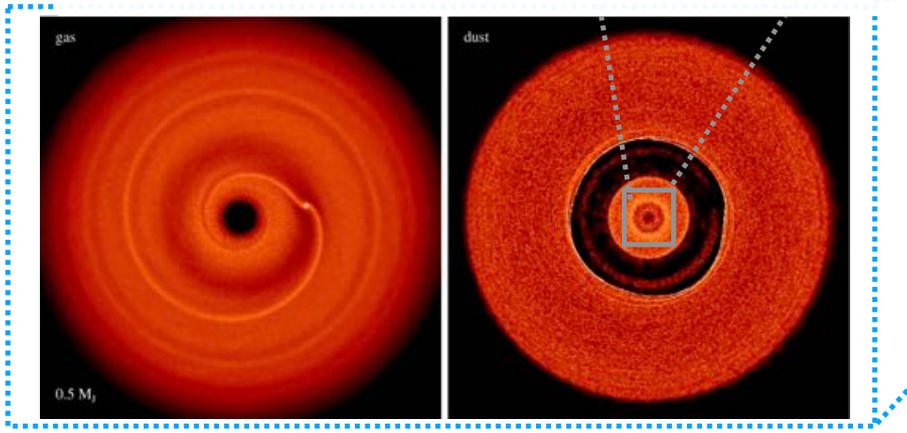
Effondrement gravitationnel halo matière noire Vlasov-Poisson 6D



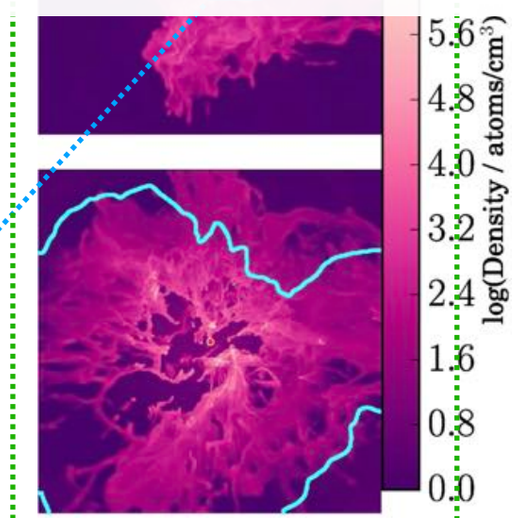
Gravité est un phénomène structurant
-> Problèmes multi-échelles -> structure des données complexe

Problèmes multi-physiques (chimie, rayonnement, champ magnétique, etc.)
-> gourmands en mémoire

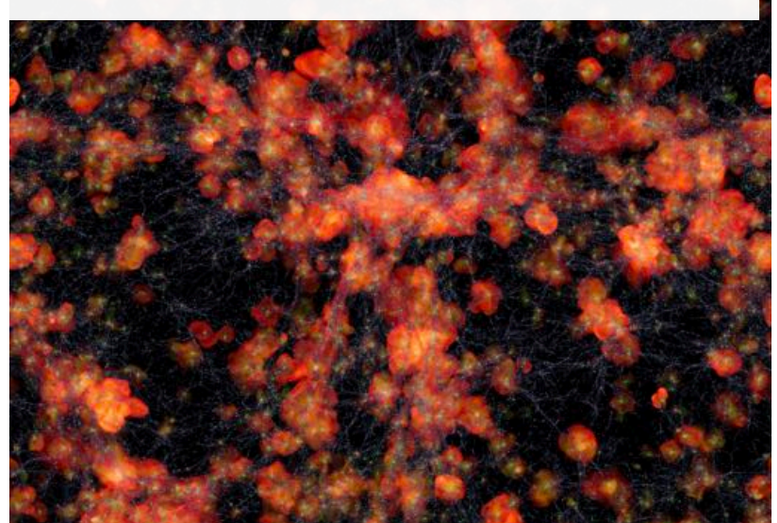
Problèmes multi-espèces (combinaison grilles et particules)
-> localité des données en mémoire



Dipierro+16



Geen+15



Chabanier+20

Prospective INSU-AA 2019

Série de 6 enquêtes :

- Utilisateurs (63 réponses)
- Laboratoires (17 réponses/23)
- Mésocentres (24/46)
- Centres d'expertise régionaux (CER) (5/7)
- Pôles thématiques nationaux (PTN) (3/5)
- Une infrastructure de recherche :
  Centre de Données astronomiques de Strasbourg (CDS)

Complétées par :

- Base de données INSU ANO5 (données)
- Retour informel du président CT4 Astrophysique GENCI
- Retours d'une sélection de missions types (Euclid, SKA, etc.)
- Retour colloque AstroSim 2019*

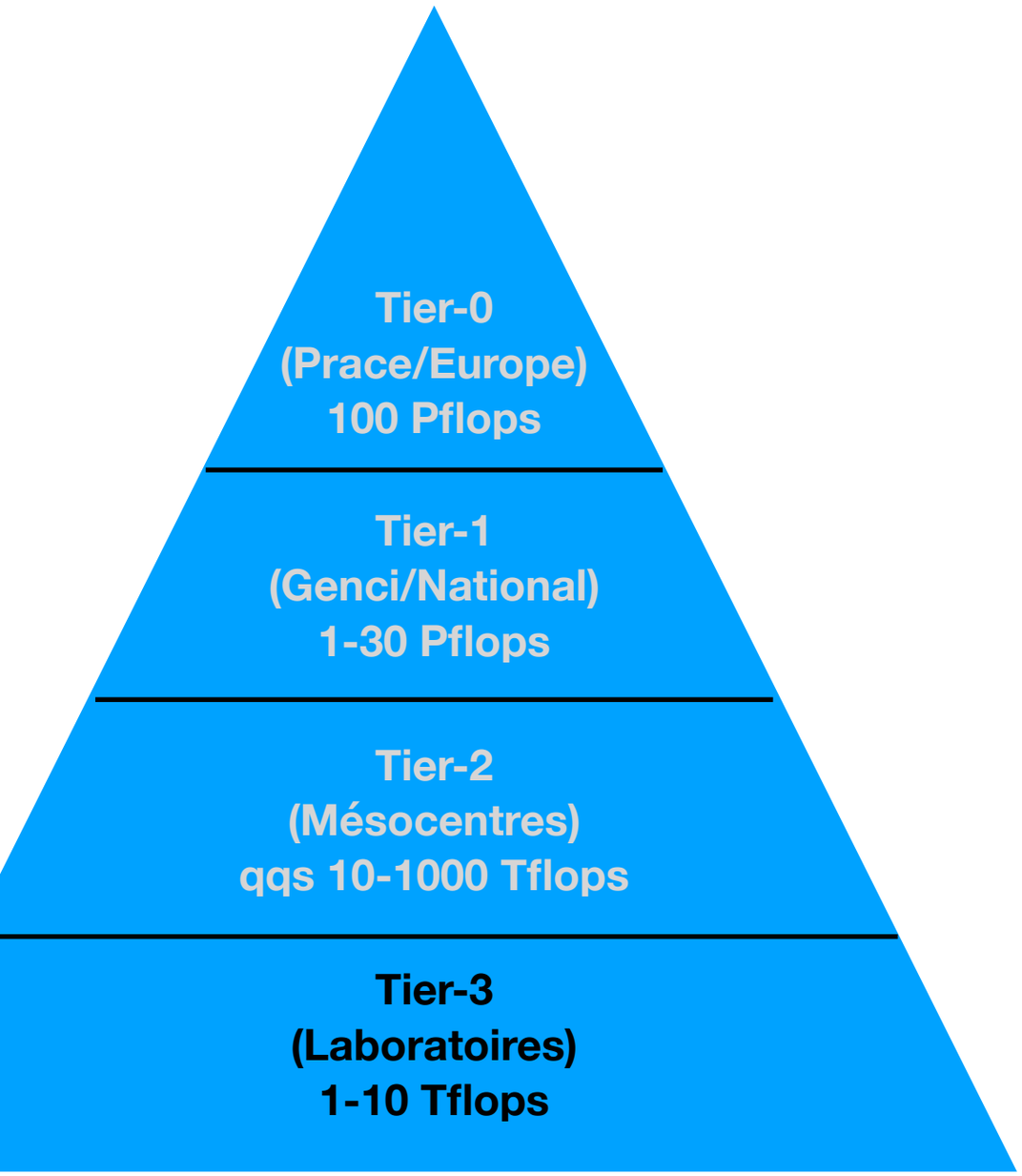
ANO5 : Action Nationale pour l'Observation (Centres de traitement, d'archivage et de diffusion de données)

CT4 : Comité Thématique Astrophysique et Géophysique

GENCI : Grand Equipement National de Calcul Intensif

*** https://astrosimconf.sciencesconf.org/data/pages/Synthe_seAstroSim.pdf**

Laboratoires



Total : 12 000 coeurs **et 18.5 PB stockage**
Moyenne labo : 900 coeurs **et 1.2 PB stockage**

Quelques moyens de calcul importants (CPU) avec une machine à ~1000 coeurs+ à accès ouvert (à coloriage simulations mais pas uniquement), et beaucoup de petites machines entre 10 et 200 coeurs

<u>Financement non pérenne</u> : maintien par sources multiples (ERC, ANR, Labo, Région, CNES)

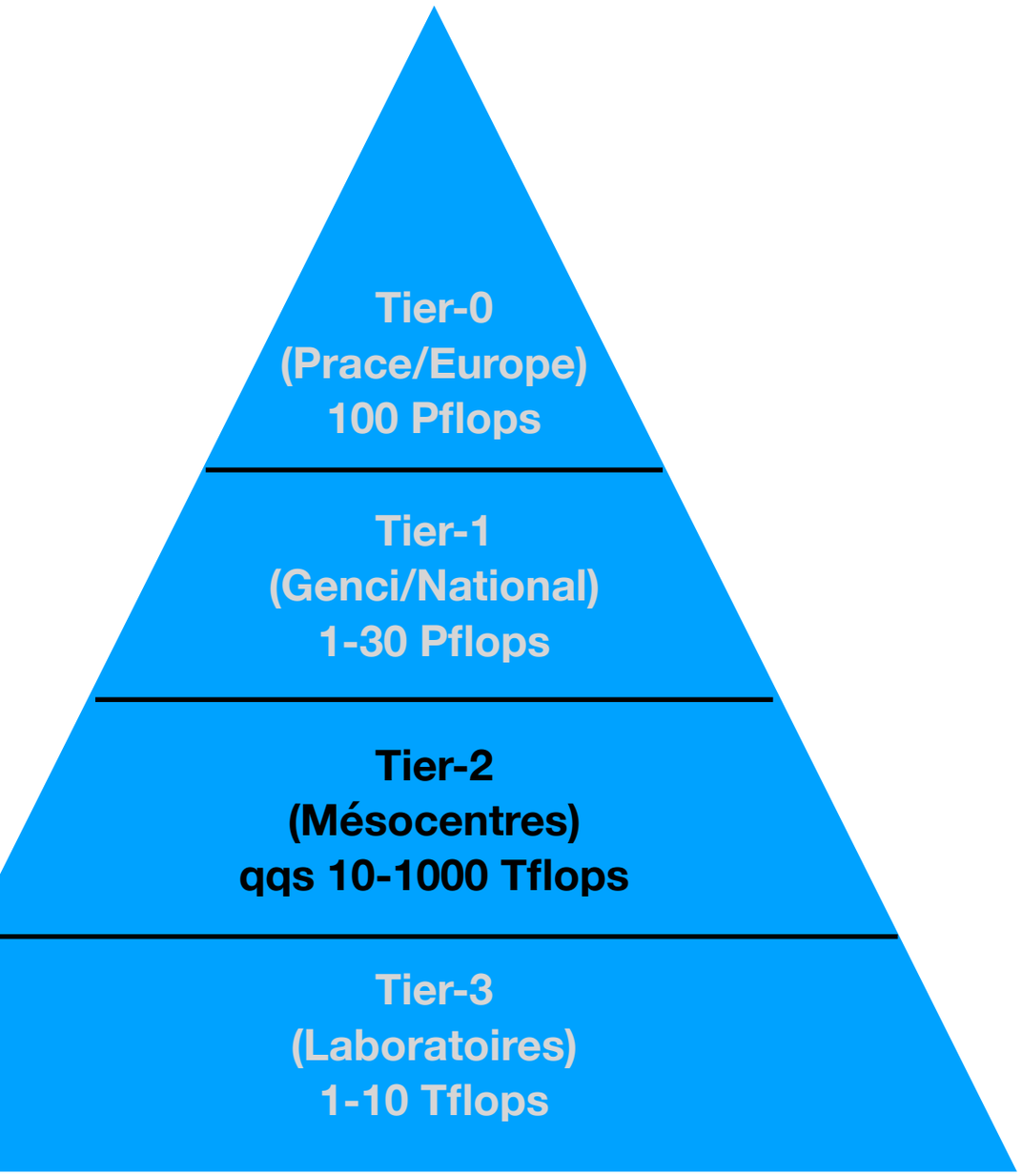
Besoins humains ITA développement logiciel forts (simulations, observations)

<u>Utilisation</u> :

- Etapes de développement des algorithmes (Prototypage, scaling, développement de codes, pipelines)
- Réduction/analyse de données pour petits volumes (observées ou simulées) <100 TB

en 2019

Mésocentres



Total de ~10+ Pflops, ~200 000 coeurs, **40+ PB de stockage (comparable au Tier-1)**

Très forte hétérogénéité sur le territoire

<u>Financement</u> : Forte impulsion (ANR Equip@Meso 10.5 M€ 2011), maintien par sources multiples (CPER, FEDER, ERC, ANR, Labo, Région, utilisateurs)

<u>Utilisation</u> :

- **Importante par les simulations**
- Moins par les observations
  (problèmes : installation des pipelines, stockage pérenne)

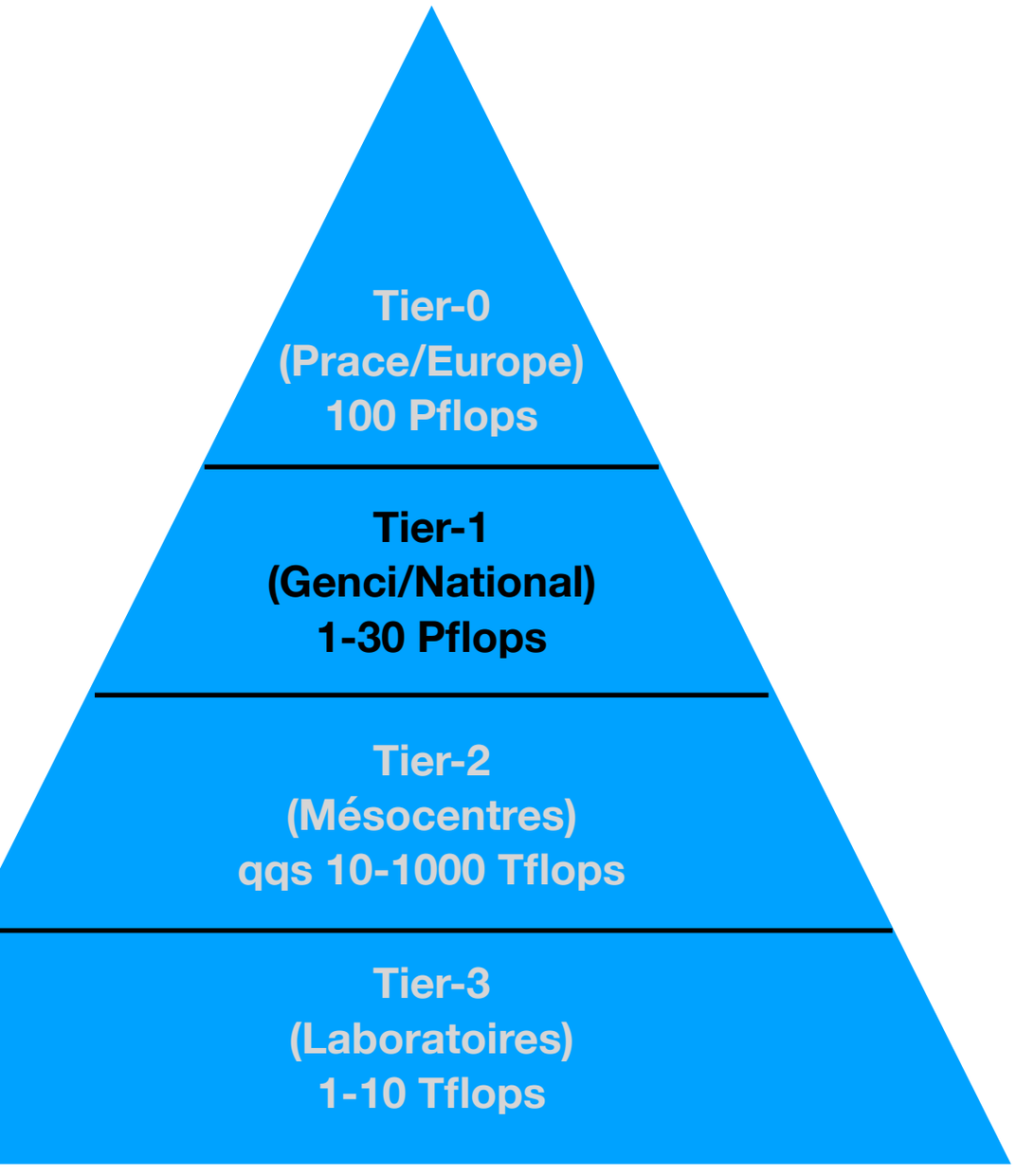
<u>Difficultés</u> :

- Ressources humaines accompagnement utilisateurs mais pas satisfaisant dans certains cas
- Utilisateurs pas toujours conscients possibilité accompagnent au déploiement codes
- Quid d'installation de matériel financé par ERC ?

en 2019

CPER : Contrat de Plan État-Région
FEDER : Fonds Européen de Développement Régional

Centres calculs nationaux



en 2019

3 Centres appels à projets GENCI : CINES, IDRIS, TGCC
14 Pflops CPU + 14 Pflops CPUGPU (Jean Zay@IDRIS)
350 000 coeurs CPU + 1 000 cartes GPU

Astro gros consommateur et fort taux succès :
120 Mheures distribuées (valeur : 4.2 M€/an),
70 demandes/an, entre 1 et 10 Mh

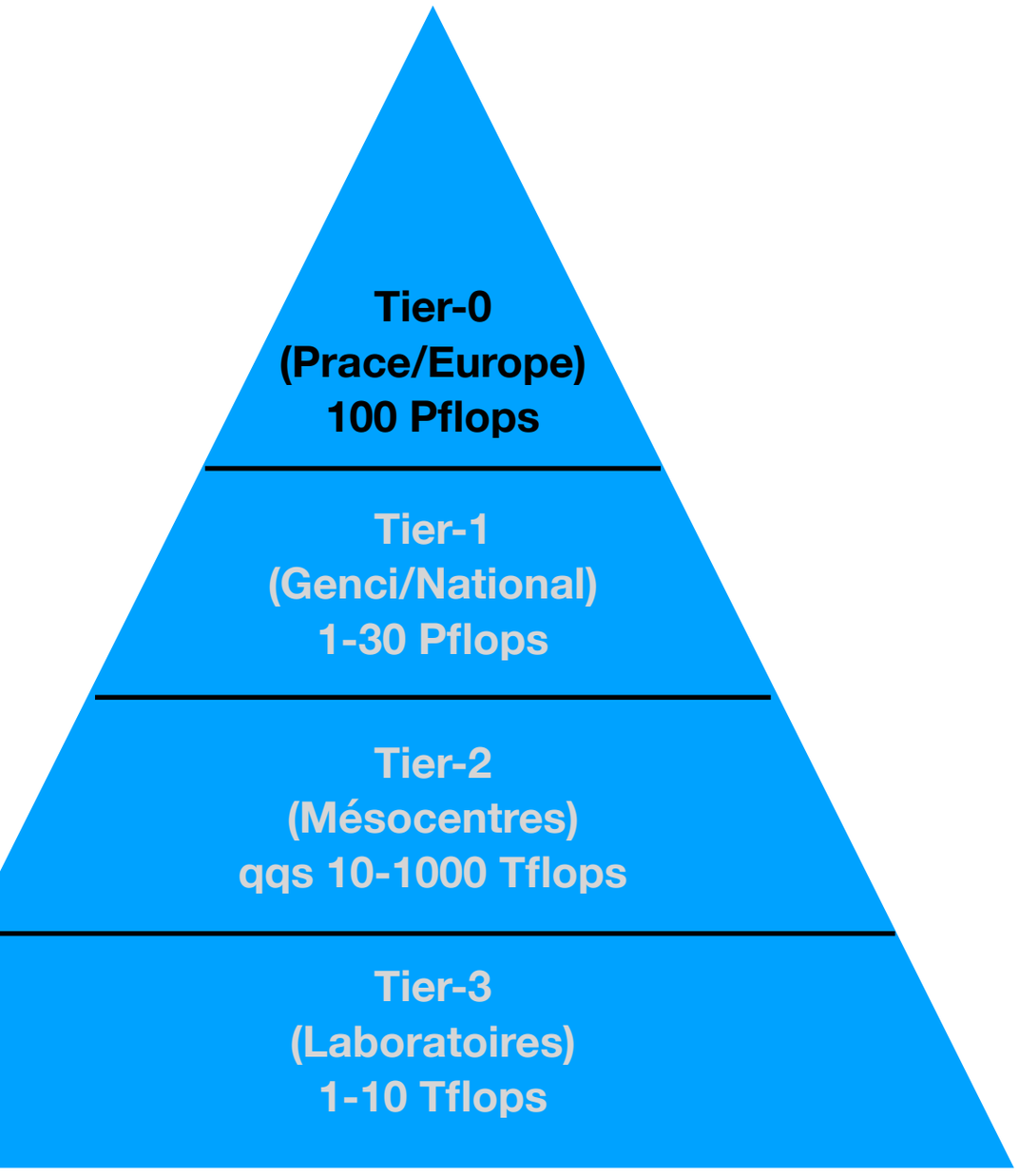
Utilisation :
- 95% simulations numériques, 5% analyse de données
- Politique GENCI : la réduction de données n'est pas acceptée, mais l'analyse oui

Difficultés :
- **Peu de stockage et de possibilités d'analyse massive de données (moyens calcul >> stockage)**
- **Politique de conservation des données pas toujours « heureuse » (suppression brutale des fichiers du scratch après 1-2 mois, limite sur le nombre et la taille des fichiers du store)**
- Utilisateurs pas toujours conscients possibilité accompagnent au déploiement et à l'optimisation codes

Equilibre thématique des Programmes Nationaux
Communauté RAMSES (35% du temps CT4A)

CINES : Centre Informatique National de l'Enseignement Supérieur (Ministère)
IDRIS : Institut du Développement et des Ressources en Informatique Scientifique (CNRS)
TGCC : Très Grand Centre de Calcul (CEA)

Centres calculs européens et internationaux



Succès de AA-France sur les 2 derniers appels (5 AA acceptés sur 6) : ~80 Mh/an

Grosses demandes entre 15 et 68 Mh sur CPU ou GPU

Seulement 18% des sondés ont participé à une demande PRACE, mais ~50% s'y projettent à l'avenir

Quelques très grosses demandes hors Europe (jusqu'à 100 Mh, récemment obtenu 350Mh)

en 2019

PRACE : Partnership for Advanced Computing in Europe

Problème du traitement de données

Problèmes du traitement des données massives issues des grands relevés observationnels
LSST (2022): raw 30TB/jour, utile 140PB en 10 ans
CTA (2025) : raw 100TB/jour, utile 25PB/an
SKA (2030) : raw **10EB**/jour, utile 600PB/an

En 2019,
Les grosses simulations produisent quelques 100TB : transfert en local des données encore possible
Quelques simulations (Euclid Flagship 16000^3, CODA-II) doivent avoir un plan de gestion des données : impossibilité de conserver « toutes » les données brutes -> quoi conserver ? comment naviguer rapidement dans les données ? comment croiser entre elles les différentes propriétés statistiques des objets simulés (IA) ?

Les simulations ont une expérience de produire et de manipuler de gros volumes de données.
Une simulation flagship atteint maintenant facilement 1PB de données utiles.

CT4(Astro) A13 GENCI : 9 demandes 50-100 TB; 4 Demandes >100TB dont 2 >1PB
(Les scratchs des machines GENCI ~2-5PB)

Centre de données et d'analyse d'instruments/grands projets

Partage d'expertise à opérer

- **Cascade de traitement complexe à travers un continuum d'infrastructures**
  transfert > stockage > réduction > analyse > diffusion des données > archivage/pérennisation

- Où stocker et réduire les données ?
  - Stratégie et vision globale versus Solutions ponctuelles
    - Intégration dans l'existant (Mésocentres, CC-IN2P3, GENCI, CC-IN2P3INSU, CC-IN2P3AA)
    - Création de CC (à un projet, CC-INSU)
    - Organisation au niveau international pour les grands projets

- Difficultés de migration des pipelines sur de nouvelles architectures (accompagnement par les centres ? ITA dédié ? formation aux solutions type conteneur?)
- Globalement besoin de formation et accompagnement sur les thématiques « Data Computing (HDA, HPC, AI), Archiving » : la communauté ne semble pas prête

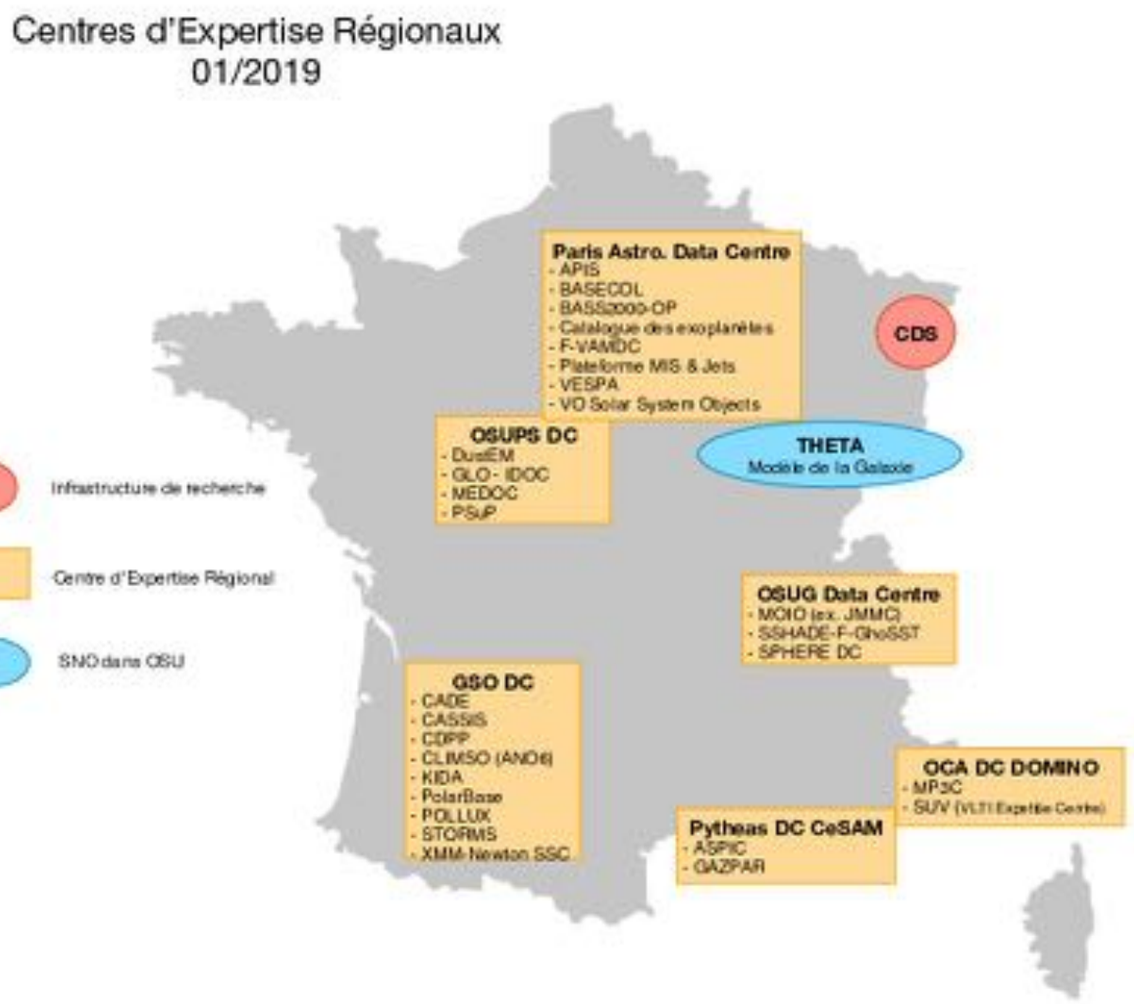
CC : Centre de Calcul

HDA : High-end Data Analysis

AI : Artificial Intelligence

Centres d’Expertise Régionaux (CER)

assurent les moyens pour le développement, la maintenance et la pérennisation des services de l’ANO5 reconnus par l‘INSU



6 CER, 1 IR, 29 SNO

IR : Infrastructure de Recherche

Pourquoi pas un CER uniquement simulations ?

L’infrastructure de recherche CDS

- Inscrite par le MESR sur la feuille de route nationale des infrastructures de recherche
- Pilier du réseau de bases de données françaises
- Collaboration avec et support aux autres bases de données et services français
- Rôle important dans l’ASOV et l’IVOA

Cela aurait-il un sens d’utiliser le CDS pour mettre à disposition les catalogues issus de simulations ?

ASOV : Action Spécifique Observatoire Virtuel
IVOA : International Virtual Observatory Alliance

Bilan Prospective INSU-AA 2019

- Complémentarité des 4 Tiers permet une montée en gamme et en technicité des applications

- Traitement de données & simulations présents au Tier-3/2
  - Précarité du financement des machines (multiples guichets, machines hors garantie)
  - Pression accrue par la présence montante des données (quelle intégration dans l'existant ?)
  - Problématique cohabitation simulations versus données

- **HP**C (Tier-1/0) massivement du fait des simulations

- Mur de l'Exascale hybride (GPU) aux étages Tier-1/0 : communauté HPC pas prête
  - Fort besoin structuration communautaire
    Initiatives : Ecoles & Ateliers (70 participants) AstroSim ; Code grille adaptative communautaire (w/ Maison de la Simulation) ; Proposition service observation RAMSES ; Contrat progrès IDRIS/GPU RAMSES

- Recommandations :
  - Action Spécifique Simulations : structuration communautaire et ExaScale hybride
  - SO Code Simulations (ANO spécifique simus?) : structuration communautaire et reconnaissance des métiers du HPC
  - Recrutements ingénieurs développement code

Bilan (2/2)

- Comité ANO5 joue pleinement son rôle
- 6 CER opérationnels et dotés de comité de pilotage
- ASOV essentiel pour la fédération des CER

- Moyens calcul, stockage et humains existants dans les laboratoires et Mésocentres mais précaires et de volumétrie variable
  - Volonté instances nationales de fédérer mais pas de visibilité sur la méthode
  - Attention ! les chercheurs apprécient fortement l'étage Tier-3 : doit être maintenu dans les bonnes proportions

- Besoin d'une réflexion globale et d'une stratégie nationale sur le stockage et la pérennité des données
  - Réflexion sur la migration éventuelle du stockage des CER vers les centres de calculs
  - CDS : étude de migration du backup vers le centre de calcul de l'U. de Strasbourg
  - Ces regroupements ne doivent pas se faire au détriment de l'accessibilité et de la fiabilité des données (avec étude de coût)

Prospective INSU(AA/OA/TS/SIC)
Défi transverse 17 : convergence HPC/HPDA

- **Recommandations** :

1. **Lancer une Action Nationale INSU Transverse sur le HPC** visant à développer une stratégie nationale face aux enjeux de l'exascale.
2. **Renforcer les compétences en ingénierie des codes** et faire émerger un pôle d'expertises INSU
3. **Développer des formations en HPC <u>mais aussi en IA adaptées aux besoins de l'INSU</u>.** En lien avec les experts IA pour attirer l'expertise IA vers les domaines de recherches INSU.
4. **Développer des collaborations** autour de la préparation à l'exascale
5. <u>Soutenir le développement de **plateformes d'interconnexion des données et des services,** intégrant les différents niveaux, nationaux, régionaux et les OSU.</u>
6. <u>Soutenir formellement les échanges et interactions entre les **communautés d'observations et les communautés modélisation et simulation**, en organisant un ou des ateliers dédiés</u>

Action Spécifique Numérique (ASN)

ASN HPC et HPDA (composition équilibrée des communautés simus/données)

L'ASN traite du Numérique au sens large, mais doit se concentrer sur les aspects traitement des masses de données, performances et portabilités des **codes**. Les *codes* sont aussi bien des codes de simulations numériques que des codes de traitement/analyse de données.

Objectifs de l'ASN :

i) Préparer les communautés numériques aux défis à venir (ExaScale, BigData, etc.) afin d'y apporter des solutions méthodologiques, techniques et humaines
ii) Favoriser la formation des communautés aux outils numériques de pointe

La dotation de l'ASN va se partager entre des financements distribués sous forme d'appel d'offre, et sous forme de soutien à des actions d'animation scientifique et méthodologique à l'initiative de l'ASN. Un aspect essentiel de cette ASN sera d'animer la thématique HPC/HPDA par des colloques réguliers invitant des numériciens AA, mais aussi des intervenants extérieurs, qui permettront de donner de la visibilité aux pratiques et expertises numériques disponibles.

L'ASN n'a pas pour but de se concentrer sur les aspects Open Science, c'est le rôle de l'ASOV, mais aura bien évidemment à l'esprit ses recommandations qu'elle relaiera.

Action Spécifique Numérique (ASN)
Mandat

➡<u>5 grandes thématiques</u> :

- Convergence des communautés HPC/HPDA
- Optimisation de la performance, de la qualité des méthodes, softs et hardware (Développement de nouveaux outils)
- Adaptations des outils et des moyens humains à disposition et modes d'utilisation des moyens de calcul (Amélioration de l'utilisation des moyens existants)
- Liens avec d'autres communautés (autres domaines disciplinaires et industrie ?)
- Impact sociétal et technologique des recherches dans le domaine HPC/HPDA

➡<u>Eléments clés</u> :

- Animer la convergence HPC/HPDA : identifier les méthodes communes
- Soutenir l'adaptation des codes aux nouvelles architectures matérielles
- Veille technologique sur les architectures et les nouvelles méthodes
- Identifier les besoins en numérique en AA
- Conseil pour l'achat de matériel (guichets de financement et proposants)
- Conseil pour l'accès aux ressources matérielles
- Veille sur les offres de formation sur le numérique
- Orientation vers les outils et experts adéquats
- Explorer les liens méthodologiques avec l'industrie

Action Spécifique Numérique (ASN)

Actualité

Conférence 12-16 Décembre 2022 (ENS Lyon)
https://asnum2022.sciencesconf.org/

Atelier commun ASOV-ASN
https://indico.in2p3.fr/event/28071

- Appel d'offre en 2023
- Liste de diffusion
- Mise en place de webinaires

L’analyse des simulations « flagship » n’est plus le fait d’une personne

The Horizon Run 5 Cosmological Hydrodynamical Simulation: Probing Galaxy Formation from Kilo- to Gigaparsec Scales

Jaehyun Lee[1,12], Jihye Shin[2,12], Owain N. Snaith[3], Yonghwi Kim[1], C. Gareth Few[4,5], Julien Devriendt[6], Yohan Dubois[7], Leah M. Cox[4], Sungwook E. Hong[2,8], Oh-Kyoung Kwon[9], Chan Park[10], Christophe Pichon[1,7], Juhan Kim[11], Brad K. Gibson[4], and Changbom Park[1]

The EAGLE project: simulating the evolution and assembly of galaxies and their environments

Joop Schaye,[1★] Robert A. Crain,[1] Richard G. Bower,[2] Michelle Furlong,[2] Matthieu Schaller,[2] Tom Theuns,[2,3] Claudio Dalla Vecchia,[4,5] Carlos S. Frenk,[2] I. G. McCarthy,[6] John C. Helly,[2] Adrian Jenkins,[2] Y. M. Rosas-Guevara,[2] Simon D. M. White,[7] Maarten Baes,[8] C. M. Booth,[1,9] Peter Camps,[8] Julio F. Navarro,[10] Yan Qu,[2] Alireza Rahmati,[7] Till Sawala,[2] Peter A. Thomas[11] and James Trayford[2]

Cosmic Dawn II (CoDa II): a new radiation-hydrodynamics simulation of the self-consistent coupling of galaxy formation and reionization

Pierre Ocvirk,[1★] Dominique Aubert,[1] Jenny G. Sorce,[1,2,3] Paul R. Shapiro,[4] Nicolas Deparis,[1] Taha Dawoodbhoy,[4] Joseph Lewis,[1] Romain Teyssier,[5] Gustavo Yepes,[6,7] Stefan Gottlöber,[3] Kyungjin Ahn,[8] Ilian T. Iliev[9] and Yehuda Hoffman[10]

The Obelisk simulation: Galaxies contribute more than AGN to H I reionization of protoclusters

Maxime Trebitsch[1,2,3], Yohan Dubois[1], Marta Volonteri[1], Hugo Pfister[1,4,5], Corentin Cadiou[1,6], Harley Katz[7,★], Joakim Rosdahl[8], Taysun Kimm[9], Christophe Pichon[1,10], Ricarda S. Beckmann[1], Julien Devriendt[7,8], and Adrianne Slyz[7]

Formation of compact galaxies in the Extreme-Horizon simulation

S. Chabanier[1,2], F. Bournaud[2,1], Y. Dubois[3], S. Codis[3,4], D. Chapon[1], D. Elbaz[2], C. Pichon[3,4,5], O. Bressand[6,7], J. Devriendt[8], R. Gavazzi[3], K. Kraljic[9], T. Kimm[10], C. Laigle[3], J.-B. Lekien[6,7], G. Martin[11,12], N. Palanque-Delabrouille[1], S. Peirani[13], P.-F. Piserchia[6,7], A. Slyz[14], M. Trebitsch[15,16], and C. Yèche[1]

Introducing the NewHorizon simulation: Galaxy properties with resolved internal dynamics across cosmic time

Yohan Dubois[1], Ricarda Beckmann[1], Frédéric Bournaud[2,3], Hoseung Choi[4], Julien Devriendt[5], Ryan Jackson[6], Sugata Kaviraj[6], Taysun Kimm[4], Katarina Kraljic[7,8], Clotilde Laigle[1], Garreth Martin[9,10], Min-Jung Park[4], Sébastien Peirani[11,1], Christophe Pichon[1,12,13], Marta Volonteri[1], and Sukyoung K. Yi[4]

Simulating galaxy formation with the IllustrisTNG model

Annalisa Pillepich,[1,2★] Volker Springel,[3,4] Dylan Nelson,[5★] Shy Genel,[6,7] Jill Naiman,[2] Rüdiger Pakmor,[3] Lars Hernquist,[2] Paul Torrey,[8] Mark Vogelsberger,[8†] Rainer Weinberger[3] and Federico Marinacci[8]

Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe

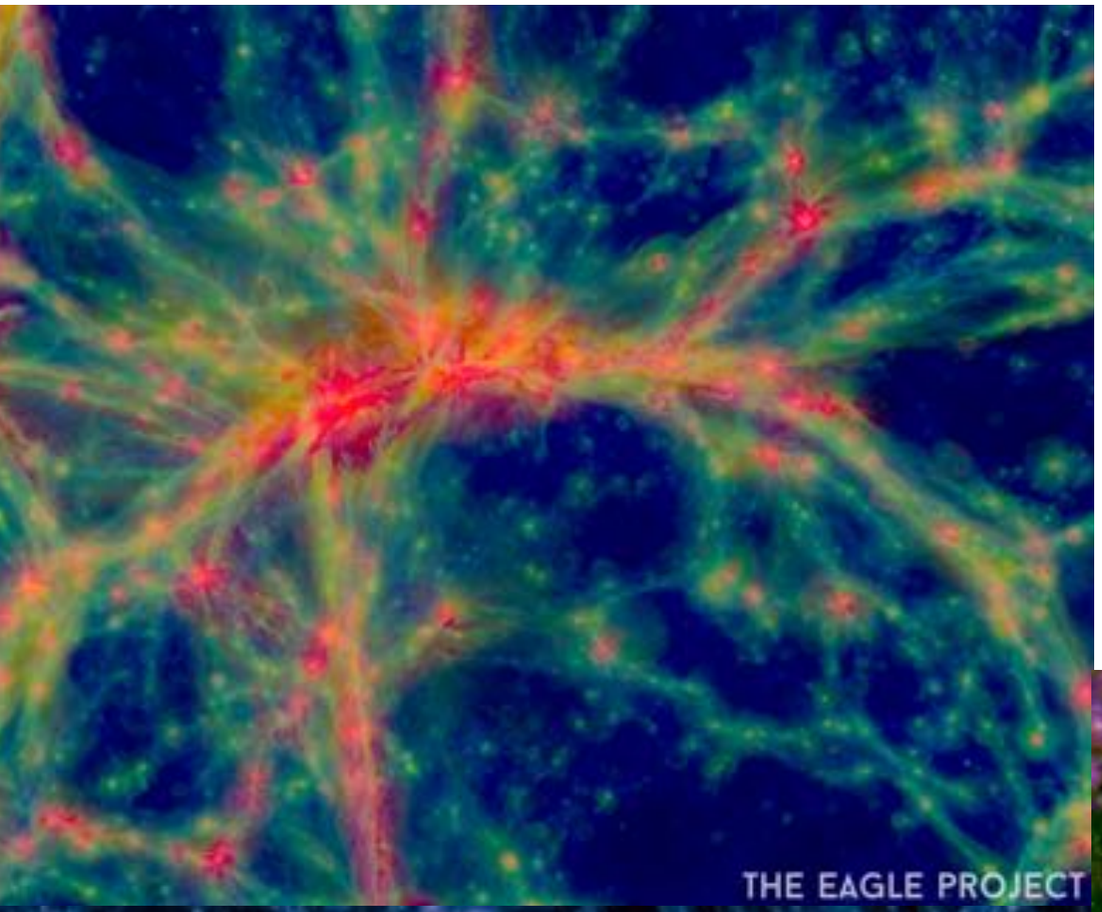
Mark Vogelsberger,[1★] Shy Genel,[2] Volker Springel,[3,4] Paul Torrey,[2] Debora Sijacki,[5] Dandan Xu,[3] Greg Snyder,[6] Dylan Nelson[2] and Lars Hernquist[2]

The SPHINX cosmological simulations of the first billion years: the impact of binary stars on reionization

Joakim Rosdahl,[1★] Harley Katz,[2,3] Jérémy Blaizot,[1] Taysun Kimm,[3,4] Léo Michel-Dansac,[1] Thibault Garel,[1] Martin Haehnelt,[3] Pierre Ocvirk[5] and Romain Teyssier[6]

L’analyse et l’exploitation scientifique de telles simulations ne se fait pas par une seule personne (builders de ~10 personnes) et l’exploitation scientifique se fait au-delà du consortium initial

Quelle durée de vie pour ces simulations ?



« ~~Avec la loi de Moore, il suffit d'attendre 2-3 ans pour refaire la simulation pour presque rien~~ »

Ces 3 simulations ont été publiées en 2014-2015 (réalisées en 2012-2013)

8-10 ans plus tard, on a toujours des papiers qui continuent d'en faire l'exploitation scientifique

La gestion des données se pose très tôt

Horizon-AGN (Dubois+14) ~ Illustris (Vogelsberger+14) ~ Eagle (Schaye+15) ~ 100-300TB
Différents types de snapshots
ex. Horizon-AGN : on conserve 66 snapshots bruts tous les ~150 Myr (l'essentiel du volume des données), le lightcone complet, mais 1000 snapshots (~10Myr) avec seulement les étoiles, 20 000 snapshots (0.5Myr) pour les trous noirs pour avoir de la haute fréquence temporelle.

CODA (Ocvirk+2016) ~ 2PB (pas conservable sur le long terme) -> réduit à 200TB

Euclid Flagship V2 (Stadel, Potter, Teyssier 2018, 4.10^{12} particules) 800 TB pour le produit final
->1 ou 2 snapshots complets + lightcone sont conservés et tout le reste ce sont des catalogues de halos (avec extraction à la volée)

Flagship V2

Des exemples vertueux

EAGLE Public Data Release

Welcome to the EAGLE public data release.

Below we provide access to the EAGLE public database, which contains galaxy properties (such as masses, star formation rates, luminosities and metallicities), merger histories and images for more than 1,000,000 simulated galaxies extracted from multiple simulations of various box sizes, numerical resolutions and physical models. Additionally, we also provide access to the particle data for each of these simulations.

Access to both the database and particle data require only a single account, which you can register for below.

EAGLE galaxy database

If you already have an account, follow this link to access the data or you can register on the following webpage.

Documentation can be found on the database itself or in the galaxy catalogue release paper.

The python examples from the release paper are available here:

- To connect to the database directly via Python here
- Galaxy stellar mass function example here.

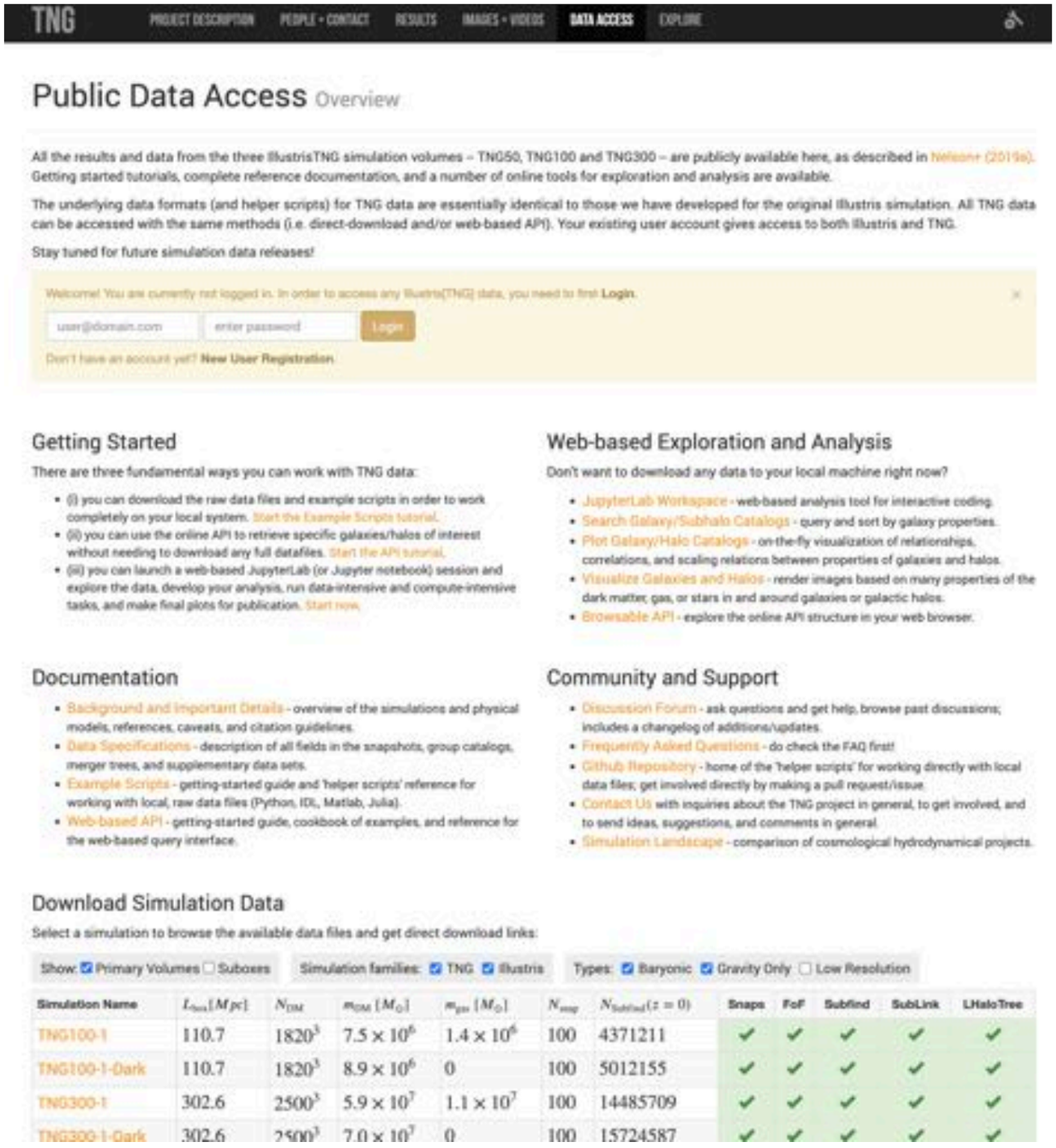
EAGLE particle data

If you already have an account, the particle data can be accessed from this link.

The documentation accompanying the particle data can be downloaded in the data release paper.

The Python examples from the release paper are available below:

- Reading a dataset (Section 4.1) here.
- Reading the snapshot header (Section 4.2) here.
- Reading dark matter mass (Section 4.3) here.
- Plotting the rotation curve of a galaxy (Section 4.4) here.
- Plotting the temperature-density relation for a galaxy (Section 4.5) here.
- The *read_eagle()* routine (Section 4.6) is available via git repository here.
- The example using *read_eagle()* to create the temperature-density relation for a galaxy (Section 4.6.1) can be found here.



Public Data Access Overview

All the results and data from the three IllustrisTNG simulation volumes – TNG50, TNG100 and TNG300 – are publicly available here, as described in Nelson+ (2019a). Getting started tutorials, complete reference documentation, and a number of online tools for exploration and analysis are available.

The underlying data formats (and helper scripts) for TNG data are essentially identical to those we have developed for the original Illustris simulation. All TNG data can be accessed with the same methods (i.e. direct-download and/or web-based API). Your existing user account gives access to both Illustris and TNG.

Stay tuned for future simulation data releases!

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

user@domain.com | enter password | Login

Don't have an account yet? **New User Registration**.

Getting Started

There are three fundamental ways you can work with TNG data:

- (i) you can download the raw data files and example scripts in order to work completely on your local system. Start the Example Scripts tutorial.
- (ii) you can use the online API to retrieve specific galaxies/halos of interest without needing to download any full datafiles. Start the API tutorial.
- (iii) you can launch a web-based JupyterLab (or Jupyter notebook) session and explore the data, develop your analysis, run data-intensive and compute-intensive tasks, and make final plots for publication. Start now.

Web-based Exploration and Analysis

Don't want to download any data to your local machine right now?

- JupyterLab Workspace - web-based analysis tool for interactive coding.
- Search Galaxy/Subhalo Catalogs - query and sort by galaxy properties.
- Plot Galaxy/Halo Catalogs - on-the-fly visualization of relationships, correlations, and scaling relations between properties of galaxies and halos.
- Visualize Galaxies and Halos - render images based on many properties of the dark matter, gas, or stars in and around galaxies or galactic halos.
- Browsable API - explore the online API structure in your web browser.

Documentation

- Background and Important Details - overview of the simulations and physical models, references, caveats, and citation guidelines.
- Data Specifications - description of all fields in the snapshots, group catalogs, merger trees, and supplementary data sets.
- Example Scripts - getting-started guide and 'helper scripts' reference for working with local, raw data files (Python, IDL, Matlab, Julia).
- Web-based API - getting-started guide, cookbook of examples, and reference for the web-based query interface.

Community and Support

- Discussion Forum - ask questions and get help, browse past discussions; includes a changelog of additions/updates.
- Frequently Asked Questions - do check the FAQ first!
- Github Repository - home of the 'helper scripts' for working directly with local data files; get involved directly by making a pull request/issue.
- Contact Us with inquiries about the TNG project in general, to get involved, and to send ideas, suggestions, and comments in general.
- Simulation Landscape - comparison of cosmological hydrodynamical projects.

Download Simulation Data

Select a simulation to browse the available data files and get direct download links:

Show: ☑ Primary Volumes ☐ Suboxes | Simulation families: ☑ TNG ☑ Illustris | Types: ☑ Baryonic ☑ Gravity Only ☐ Low Resolution

Simulation Name	$L_{box}[Mpc]$	N_{DM}	$m_{DM}\,[M_\odot]$	$m_{gas}\,[M_\odot]$	N_{snap}	$N_{Subfind}(z=0)$	Snaps	FoF	Subfind	SubLink	LHaloTree
TNG100-1	110.7	1820^3	7.5×10^6	1.4×10^6	100	4371211	✓	✓	✓	✓	✓
TNG100-1-Dark	110.7	1820^3	8.9×10^6	0	100	5012155	✓	✓	✓	✓	✓
TNG300-1	302.6	2500^3	5.9×10^7	1.1×10^7	100	14485709	✓	✓	✓	✓	✓
TNG300-1-Dark	302.6	2500^3	7.0×10^7	0	100	15724587	✓	✓	✓	✓	✓

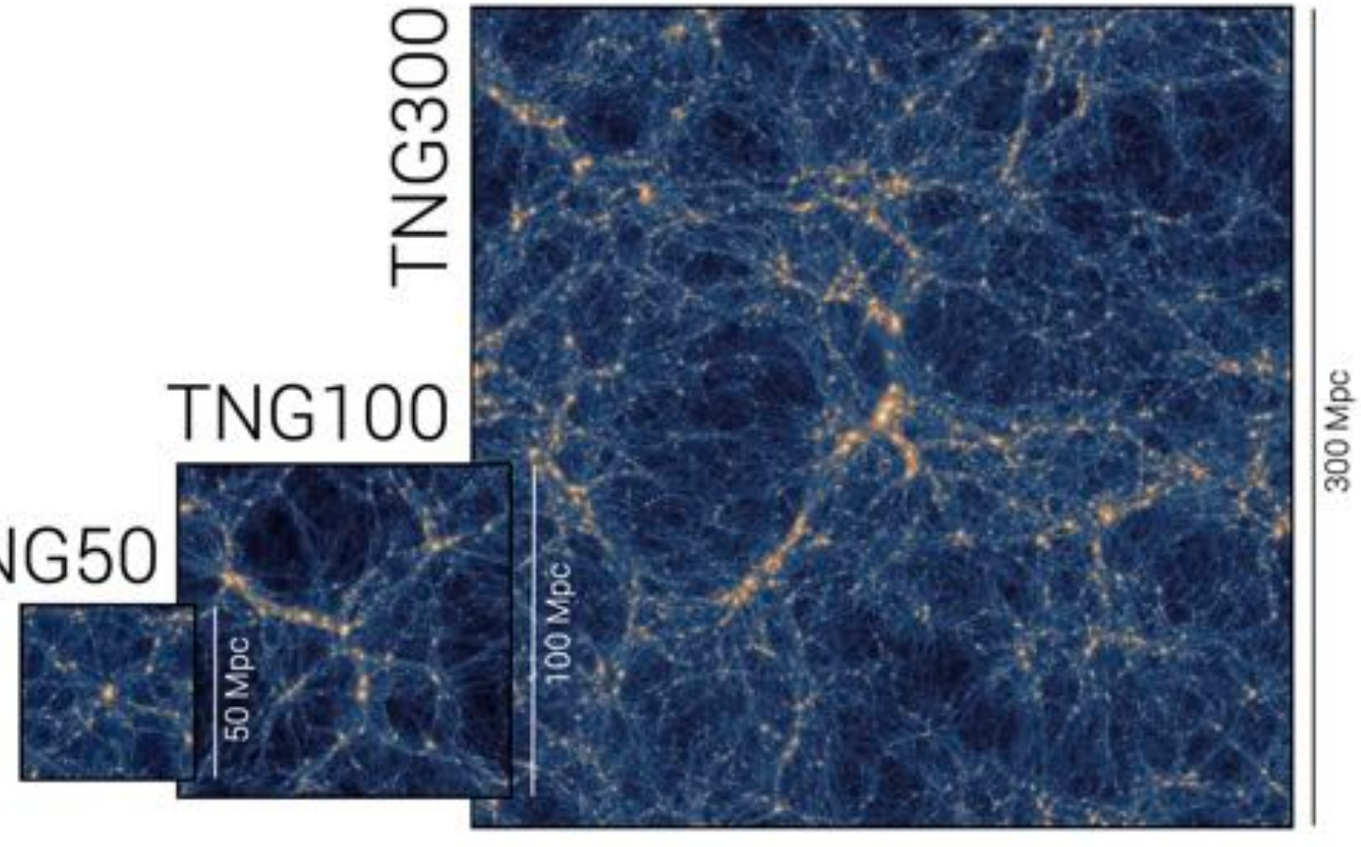
IllustrisTNG

Download Simulation Data

Select a simulation to browse the available data files and get direct download links:

Show: ☑ Primary Volumes ☐ Suboxes Simulation families: ☑ TNG ☑ Illustris Types: ☑ Baryonic ☑ Gravity Only ☐ Low Resolution

Simulation Name	$L_{\rm box}\,[Mpc]$	$N_{\rm DM}$	$m_{\rm DM}\,[M_\odot]$	$m_{\rm gas}\,[M_\odot]$	$N_{\rm snap}$	$N_{\rm Subfind}(z=0)$	Snaps	FoF	Subfind	SubLink	LHaloTree
TNG100-1	110.7	1820^3	7.5×10^6	1.4×10^6	100	4371211	✓	✓	✓	✓	✓
TNG100-1-Dark	110.7	1820^3	8.9×10^6	0	100	5012155	✓	✓	✓	✓	✓
TNG300-1	302.6	2500^3	5.9×10^7	1.1×10^7	100	14485709	✓	✓	✓	✓	✓
TNG300-1-Dark	302.6	2500^3	7.0×10^7	0	100	15724587	✓	✓	✓	✓	✓



The public release of IllustrisTNG (hereafter, TNG) follows upon and further develops tools and ideas pioneered in the original Illustris data release. We offer 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 the IDL, Python, and Matlab languages) are provided.

Nelson+19

IllustrisTNG

Appendix A: Simulation Data Details

Table A.1 Details on the file organization for all twenty TNG runs, both baryonic and dark-matter only. We include the number of file chunks, the average size of a full snapshot and the corresponding group catalog, and an estimate of the total data volume of the simulation.

Run	Alternate Name	Total $N_{\rm DM}$	$N_{\rm chunks}$	Full Snapshot Size	Avg Groupcat Size	Total Data Volume
L35n270TNG	TNG50-4	19,683,000	11	5.2 GB	20 MB	0.6 TB
L35n270TNG_DM	TNG50-4-Dark	19,683,000	4	1.2 GB	10 MB	0.1 TB
L35n540TNG	TNG50-3	157,464,000	11	44 GB	130 MB	7.5 TB
L35n540TNG_DM	TNG50-3-Dark	157,464,000	4	9.4 GB	50 MB	0.6 TB
L35n1080TNG	TNG50-2	1,259,712,000	128	350 GB	860 MB	18 TB
L35n1080TNG_DM	TNG50-2-Dark	1,259,712,000	85	76 GB	350 MB	4.5 TB
L35n2160TNG	TNG50-1	10,077,696,000	680	2.7TB	7.2 GB	~320 TB
L35n2160TNG_DM	TNG50-1-Dark	10,077,696,000	128	600 GB	2.3 GB	36 TB
L75n455TNG	TNG100-3	94,196,375	8	27 GB	110 MB	1.5 TB
L75n455TNG_DM	TNG100-3-Dark	94,196,375	4	5.7 GB	40 MB	0.4 TB
L75n910TNG	TNG100-2	753,571,000	56	215 GB	650 MB	14 TB
L75n910TNG_DM	TNG100-2-Dark	753,571,000	8	45 GB	260 MB	2.8 TB
L75n1820TNG	TNG100-1	6,028,568,000	448	1.7 TB	4.3 GB	128 TB
L75n1820TNG_DM	TNG100-1-Dark	6,028,568,000	64	360 GB	1.7 GB	22 TB
L205n625TNG	TNG300-3	244,140,625	16	63 GB	340 MB	4 TB
L205n625TNG_DM	TNG300-3-Dark	244,140,625	4	15 GB	130 MB	1 TB
L205n1250TNG	TNG300-2	1,953,125,000	100	512 GB	2.2 GB	31 TB
L205n1250TNG_DM	TNG300-2-Dark	1,953,125,000	25	117 GB	810 MB	7.2 TB
L205n2500TNG	TNG300-1	15,625,000,000	600	4.1 TB	14 GB	235 TB
L205n2500TNG_DM	TNG300-1-Dark	15,625,000,000	75	930 GB	5.2 GB	57 TB

The full snapshots of TNG50-1, TNG100-1, and especially those of TNG300-1, are sufficiently large that it may be prohibitive for most users to acquire or store a large number. We note that transferring ~1.5 TB (the size of one full TNG100-1 snapshot) at a reasonably achievable 10 MB/s will take roughly 48 hours, increasing to roughly five days for a ~4.1 TB full snapshot of TNG300-1. As a result, projects requiring access to full simulation datasets, or extensive post-processing computations beyond what are being made publicly available, may benefit from closer interaction with members of the TNG collaboration.

Nelson+19

Illustris (not TNG) 3.5 ans après sa mise sous forme publique

1.7TB/jour ~ 160Mb/s* !

7.1 Usage of the Illustris Public Data Release

Since its release, the original Illustris public data release has seen widespread adoption and use. To date, in the three and a half years since launch, 2122 new users have registered and made a total of 269 million API requests, including 2.7 million 'mock FITS' file downloads. For the flagship Illustris-1 run, a total of 1390 full snapshots, 6650 group catalogs, and 180 merger trees have been downloaded. 26 million subhalo 'cutouts' of particle-level data, and 3.1 million Sub-Link merger tree extractions have been requested. The total data transfer for this simulation to date is ≈ 2.15 PB. Roughly 3100 subbox snapshots of Illustris-1 have been downloaded. The next most accessed simulation is Illustris-3, likely because it is included in the getting started tutorials as an easy, lightweight alternative to Illustris-1. Since launch, there has been a nearly constant number of ~ 100 – 120 active users, based on activity within the last 30 days.

To date, 163 publications have directly resulted from, or included analysis results from, the Illustris simulation. While early papers were written largely by the collaboration itself, recent papers typically do not involve members of the Illustris team, representing widespread public use of the data release. Of the 10 most recent papers published on Illustris, only one was from the team. Given the significantly expanded scope of TNG with respect to Illustris, as well as the relatively more robust and reliable physical model and outcomes, we expect that uptake and usage will be similarly broad.

*Connexion IAP 1Gb/s partagé avec (fourni par) l'Observatoire de Paris…

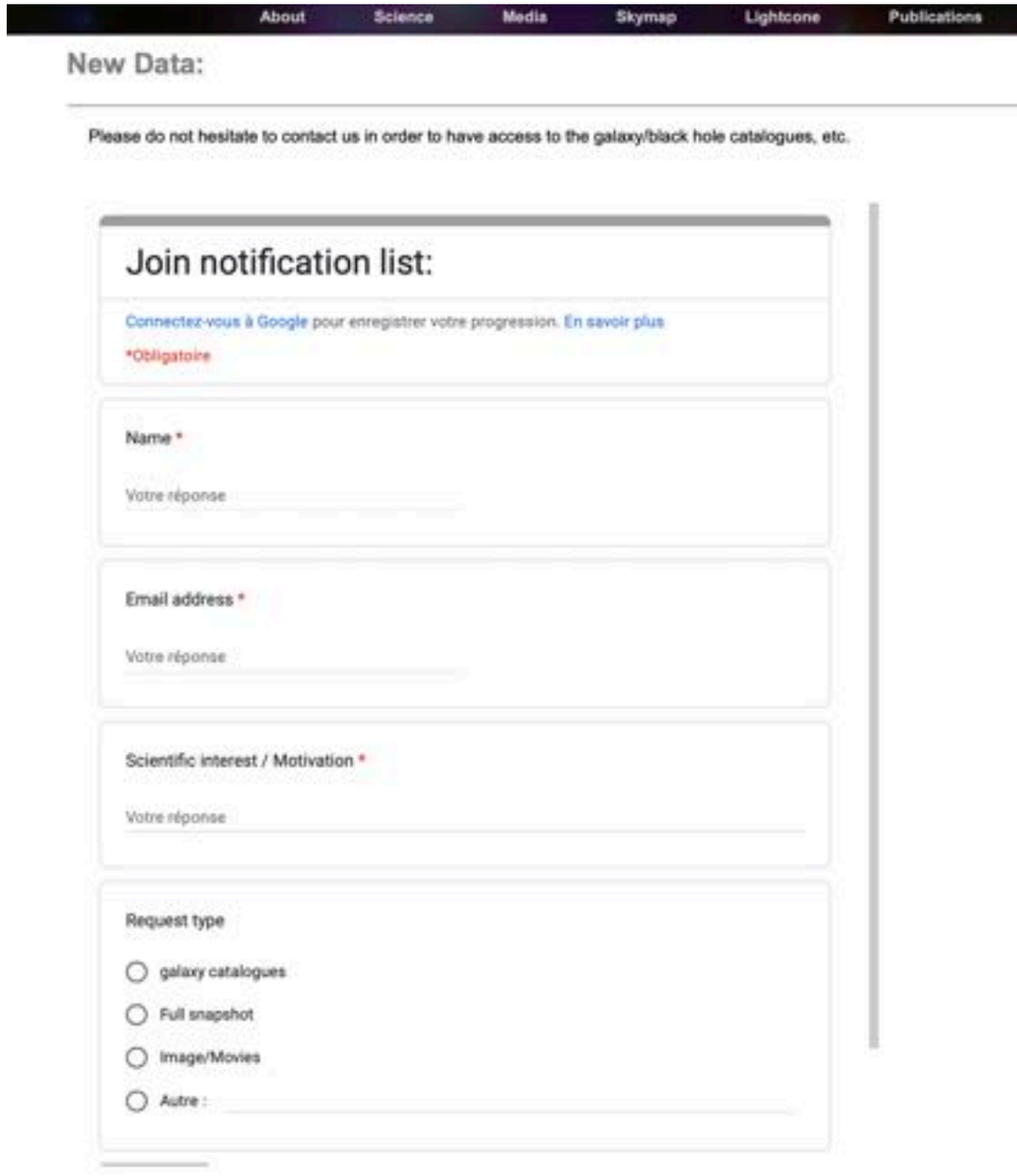
Nelson+19

Un exemple moins vertueux (le mien)

Politique de diffusion des données Horizon-AGN :
« contactez-nous on vous fournira les données »
-> En pratique ça ne marche pas. Beaucoup de temps perdu.
-> Au final on ouvre un compte sur la machine en local (IAP), on fournit les outils d'analyse, on accompagne, et on collabore (ou pas)

Très sous-optimal. Cela limite la diffusion de la simulation et son exploitation.

Pourquoi ne pas avoir fait autrement ? pas les compétences et pas l'énergie pour faire un data release complet.



Une reflexion sur les aspects humains

Discussion avec Dylan Nelson :

Illustris : **4 mois de travail** à temps plein pendant sa thèse après avoir développé une expertise scientifique sur la simulation pour son propre travail scientifique. Dylan avait déjà une expertise sur les bases de données, et le développement web… Probablement **~1 an est une durée plus réaliste** si la personne ne maîtrise pas ces aspects.

Doctorat à CfA : 6 ans !!!

TNG : **1 mois de travail** à temps plein -> **rentabilisation du travail** fourni sur Illustris (même code, même données, même type de simulation)

- En pratique le travail de construction des catalogues et mise à disposition des données repose souvent sur 1 personne
- Dans le système français (3 ans de doctorat, 2-3 ans pour un postdoc) est-on prêt à permettre à un doctorant/post-doctorant de dépenser une fraction significative de son travail pour faire ce type de data release ?
- Quelle reconnaissance en terme de recrutement quand il s'agit de mettre en compétition des profils de chercheurs ?
  - ➡ CNRS : Section 17 versus CID 55 ?
  - ➡ CNAP : comment faire recruter des profils de ce type sur une tâche de service ANO5 ?

Conclusion

Opportunités :

- Augmenter la visibilité et l'impact scientifique de ses travaux sur le long terme
- Faciliter le travail d'inter-comparaison (à condition que la construction des catalogues soit équivalente…)
- Permettre la reproductibilité des résultats

Risques :

- Travail « ingrat » (loin de nos expertises) qui repose souvent, mais pas toujours, sur de jeunes épaules
- Manque d'accompagnement
- Quelle valorisation de ce travail dans les recrutements ? (CID55 maintenant)
- Tension sur les ressources matérielles (réseau, noeuds mis à disposition du reste du monde)