DMP du projet "LOcal Clusters And supercLuster In sZ: Adding Thermal, kInetic and relativistic cOrrectioNs"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Plan Details

Plan title	DMP du projet "LOcal Clusters And supercLuster In sZ: Adding Thermal, kInetic and relativistic cOrrectioNs"
Plan purpose/scope	The primary data consist of simulated cubes of the local Universe and of simulated cubes around clusters of galaxies. These simulation data can be divided in two sorts: N-body for the dark matter (DM) only component and hydrodynamical for the DM and baryonic components. The simulations differ by their sizes and by the resolution of their mass particles. The simulated data are obtained by running two different, numerical methods Ramses [Teyssier 2002] & OpenGadget3 [Springel 2005, Beck et al. 2016] based on different numerical techniques (AMR vs. SPH). The codes will be run on supercomputers: SuperMUC-NG supercomputer in Munich and Joliot-Curie supercomputer in France. The data provenance will be fully described in the science articles that will accompany the presentation and the analysis of the simulations. A shorter description will accompany the actual data on the repositories.
Fields of science and technology (from OECD classification)	Physical sciences
Language	eng
Creation date	2022-08-24
Last modification date	2023-04-19
Identifier	10.48326/idoc.dmp.localization
Identifier type	DOI
License	**Name** Etalab Open License 2.0 **URL** http://spdx.org/licenses/etalab-2.0.json
Associated documents (publications,reports, patents, experimental plan...), website	• LOcal Clusters And supercLuster In sZ: Adding Thermal, kInetic and relativistic cOrrectioNs : https://localization.ias.universite-paris-saclay.fr/

Project Details

Project title	LOcal Clusters And supercLuster In sZ: Adding Thermal, kInetic and relativistic cOrrectioNs
Acronym	LOCALIZATION
Abstract	Galaxy clusters offer unique opportunities to test the cosmological model. This relies on the use of statistical assessments of their properties from comparison of simulations and observations. To overcome the limitation of statistical approaches, one would ideally need an actual reproduction of the Universe. Constrained simulations now permit building a high fidelity clone of the local Large Scale Structure (LSS) of the Universe within 400Mpc valid down to cluster scales. LOCALIZATION will produce and use simulated clones of local clusters and superclusters to fully understand the building up of these cosmic laboratories. We will study their physical and dynamical properties via cosmic flows, X-rays and Sunyaev-Zeldovich signals and we will derive the cosmological signatures of the local LSS on the cosmic background at large scales. LOCALIZATION will pave the way from *precision* cosmology from averaged statistics towards *accurate understanding* of the cosmological growth of structures.
Funding	• Agence Nationale de la Recherche : ANR-21-CE31-0019
Start date	2021-12-01
End date	2024-11-30
Partners	• Ludwig-Maximilians-Universität München https://www.lmu.de/en/about-lmu/index.html

Research outputs :

1. OG3 Sims : numerical, particle data : binary (Dataset)
2. Subfind Postprocessing : numerical, galaxy catalogs, meta data : binary + ASCII (Dataset)
3. RAMSES Sims : numerical, grid and particle data : binary (Dataset)
4. Rockstar postprocessing : numerical, halo and galaxy catalogs : hdf5+ASCII (Dataset)

Contributors

Name	Affiliation	Roles
Aghanim Nabila - https://orcid.org/0000-0002-6688-8992	CNRS	• DMP manager • Personne contact pour les données (Rockstar, Subfind , RAMSES Sims, OG3 Sims) • Project coordinator
Poulleau Gilles		

DMP du projet "LOcal Clusters And supercLuster In sZ: Adding Thermal, kInetic and relativistic cOrrectioNs"

1. Data description and collection or re-use of existing data

1a. How will new data be collected or produced and/or how will existing data be re-used?

The primary data consist of simulated cubes of the local Universe and of simulated cubes around clusters of galaxies. These simulation data can be divided in two sorts: N-body for the dark matter (DM) only component and hydrodynamical for the DM and baryonic components. The simulations differ by their sizes and by the resolution of their mass particles.

The simulated data are obtained by running two different, numerical methods Ramses [Teyssier 2002] & OpenGadget3 [Springel 2005, Beck et al. 2016] based on different numerical techniques (AMR vs. SPH). The codes will be run on supercomputers: SuperMUC-NG supercomputer in Munich and Joliot-Curie supercomputer in France.

The data provenance will be fully described in the science articles that will accompany the presentation and the analysis of the simulations. A shorter description will accompany the actual data on the repositories.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

For the full 500 Mpc/h box, the DM simulations are produced either with 2048^3 particles or 3072^3 particles. The hydrodynamic simulations are produced with 768^3 particles, 1536^3 particles, and 3072^3 particles.

In addition, a zoom simulation in a 20 Mpc/h diameter sphere centered on the Virgo cluster is produced with 8192^3 particles effective in DM. Counterpart hydrodynamic simulations of the Virgo cluster are produced for different initial mass functions with 2048^3 particles effective (minimum cell size 3820pc/h) and with 8192^3 particles effective (minimum cell size 250pc/h).

Name of Data	Data Type	Data Format	Data size
OG3 Sims	numerical, particle data	binary	Hundreds of TB
Subfind Postprocessing	numerical, galaxy catalogs, meta data	binary + ASCII	Hundreds of GB
RAMSES Sims	numerical, grid and particle data	binary	Hundreds of TB
Rockstar postprocessing	numerical, halo and galaxy catalogs	hdf5+ASCII	Hundreds of GB

---

2. Documentation and data quality

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata will contain standardized catalogs of halos, galaxies and their properties, simulations will be described in YML files. Data will be uploaded into a database served by the cosmo sim web portal at LRZ. Specialized catalogs will be made available on the project webpage for download.

2b. What data quality control measures will be used?

Simulation data and catalogs will be tested against observational data to guarantee scientific usability and correctness. Data from different simulations obtained with various resolutions and codes (RAMSES and OpenGadget3) will be cross-checked to estimate systematic errors due to numerical methods and uncertainties in the underlying physical models.

---

3. Storage and backup during the research process

3a. How will data and metadata be stored and backed up during the research?

LOCALIZATION data will be stored at the HPC facilities where they are produced. The HPC facilities also provide means for long term storage (backup) of the plain simulation data. The metadata will be stored additionally at local computing facilities of Universite Paris Saclay and LMU, as well as on the cosmological web portal.

3b. How will data security and protection of sensitive data be taken care during the research

LOCALIZATION data are not sensitive ones.

The data disseminated by IDOC-DATA allow the study of the properties of objects in the universe (stars, planets, galaxies, interstellar medium for example), excluding the planet earth.

Therefore, concerns on this subject related to the data themselves are not relevant because the study subjects are objects with no relationship to people.

Moreover, for public data, access to them through IDOC-DATA is made without registration of identifiers.

As recommended in the GDPR (**General Data Protection Regulation** - European Union), logs of anonymous connections are kept for one year for security and server protection reasons. Then this information is the subject of statistical extracts allowing the visualization of the uses of IDOC-DATA in a global way.

Finally, after this one-year period, these logs are deleted.

One of the IDOC-DATA staff is clearly identified with the function of security officer.

---

4. Legal and ethical requirements, code of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

The LOCALIZATION project does not rely on personal data.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Data published by LOCALIZATION will be freely available to everyone. Intellectual rights will be acknowledged by citing the according LOCALIZATION publications.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

This does not apply to LOCALIZATION

---

5. Data sharing and long-term preservation

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

The LOCALIZATION data will be shared both/either via the nationally-recognised data center IDOC-DATA at Institut d'Astrophysique Spatiale in Paris Saclay and via the web portal for cosmological simulations developed at Ludwig Maximilian University. The latter offers the possibility to interactively explore the datacubes.

The data will be first analyzed and used by the LOCALIZATION collaboration during the course of the project, i.e. from Dec. 2021 to Nov. 2024.

During this curation period, data may be made available to other teams for external projects upon request if these projects do not impact the core science goals of LOCALIZATION.

After Nov. 2024 the data will progressively then be made public.

We foresee that a full year after completion of the project will be necessary to finalize the science projects involving students and young researchers. The envisaged date for public release will thus be Dec. 2025.

Note : The scientific team has all expertise in disseminating data.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

All post-processed data are foreseen to be made public.

The LOCALIZATION data will be shared both/either via the nationally-recognised data center IDOC-DATA at Institut d'Astrophysique Spatiale in Paris Saclay and via the web portal for cosmological simulations developed at Ludwig Maximilian University. The latter offers the possibility to interactively explore the datacubes.

5c. What methods or software tools are needed to access and use data?

Web browser or python code available on GitHub and with free license. Data will be shared via both a repository or handled directly for specific requests. Data will be also shared via the "Data Science Storage (DSS)" service at LRZ.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

One DOI will be generated for each set of data.

---

6. Data management responsibilities and resources

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The LOCALIZATION management will collectively act as the data steward.

Currently this support will be under the responsability of Nabila Aghanim (IAS Paris Saclay) and Klaus Dolag (LMU München)

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

This will be handled by IDOC-DATA and the existing cosmo sim data portal operated by the LMU.

Regarding IDOC-DATA, the platform has been actively involved since the beginning in processes that aim to improve the infrastructure supporting the reuse of scholarly data. Therefore, when a diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that now refer to as the FAIR Data Principles, IDOC-DATA makes every effort to comply with these recommendations.