This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957047.

H2020-LC-SC3-EE-2020-1/LC-SC3-B4E-6-2020

Big data for buildings

Building Information aGGregation, harmonization and analytics platform

Project Nº 957047

D4.3 - Public BIGG Data

Responsible:	Stoyan Danov (CIMNE)
Document Reference:	D4.3
Dissemination Level:	Public
Version:	Final
Date:	23/11/2023

Executive Summary

The present report is an accompanying document to the deliverable D4.3 - Public BIGG Data denominated in the Grant Agreement as Open Research Data Pilot type of deliverable. The content of the deliverable itself is uploaded in the public GitHub repository of the BIGG project: https://github.com/biggproject.

The document presents a high-level overview of the BIGG project repository's content and includes screenshots and links to the different sections contained in it.

The repository will be kept alive on GitHub with public access to all information that is not sensitive according to the stipulated in the BIGG Data and Ethics Management Plan (D1.3) and the BIGG Consortium Agreement.

The responsibility for the BIGG project repository is of the project coordinator.

Table of Contents

Table of Figures

Table of Acronyms and Definitions

Acronym	Definition
AI	Artificial Intelligence
JSON	JavaScript Object Notation
RDF	Resource Description Framework
RML	RDF Mapping Language
UML	Unified Modelling Language
Turtle	Terse RDF Triple Language
WP	Work Package

Contributors Table

DOCUMENT SECTION	AUTHOR(S)	CONTRIBUTOR(S) to results	REVIEWER(S)
All sections	Stoyan Danov (CIMNE)	All developers of the BIGG project repository in GitHub	María Pérez (Inetum)

I. INTRODUCTION

This document presents a high-level overview of the BIGG project public repository which contains the open-source software, technical documentation and data generated during the project.

I.1. Organization of the document

This report is organized as follows:

- Section II describes the content of the BIGG project public repository.
- Section III provides conclusions and future work.

I.2. Scope and audience

This document is important for all project participants and users of the BIGG tools. It provides an overview and references to the public BIGG project repository containing open-source software, technical documentation and data generated in the project.

II. BIGG PROJECT PUBLIC REPOSITORY

The public data of the BIGG project is uploaded in a public GitHub repository containing the BIGG ontology, data harmonization tools, software libraries and tools for data analytics, technical documentation, and data generated during the project.

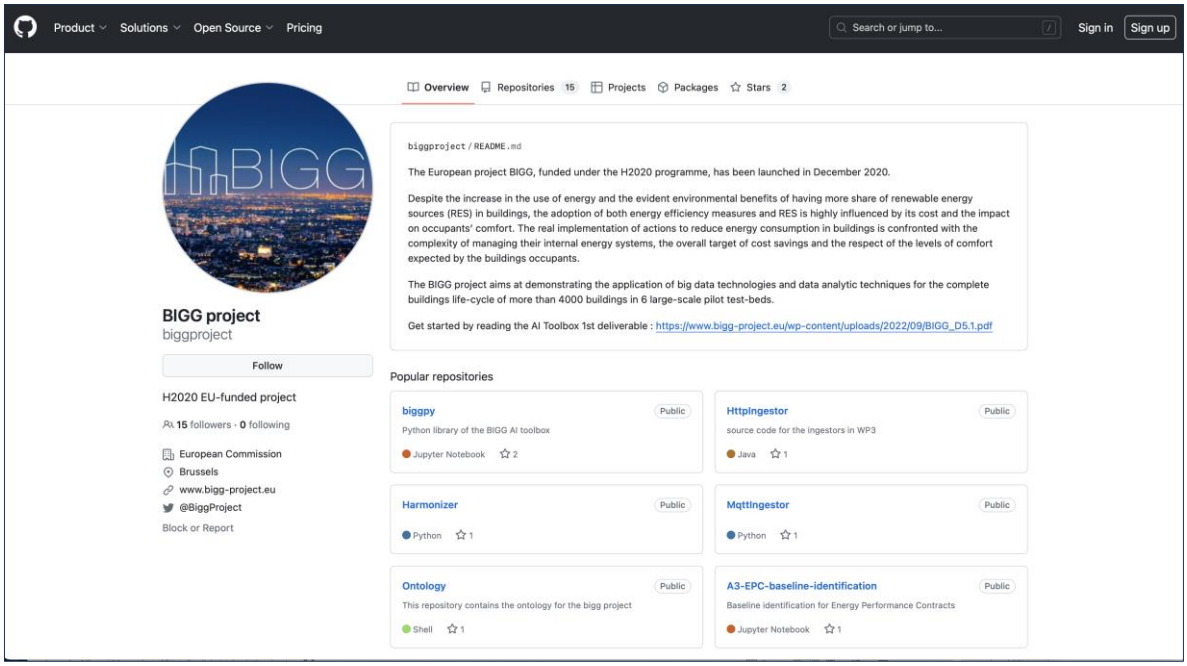


Figure 1 - BIGG project public repository (**https://github.com/biggproject**)

The information is structured into sub-repositories that are briefly described next.

II.1. Ontology

This repository contains the BIGG project ontology and the technical documentation. The information includes the full description of the BIGG and BIGGstd in RDF Turtle, diagrams, extensions, and enumerations.

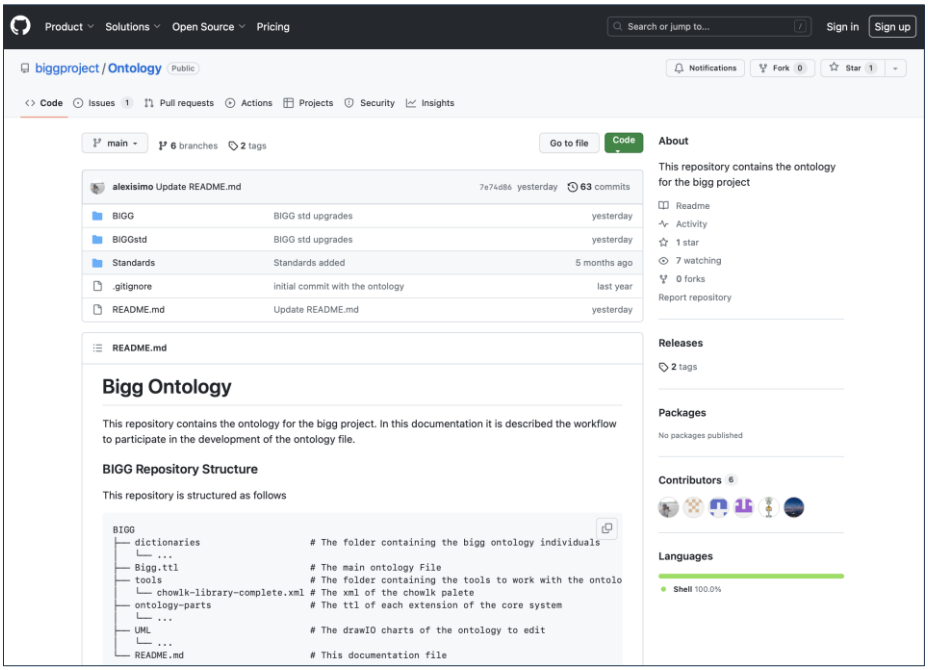
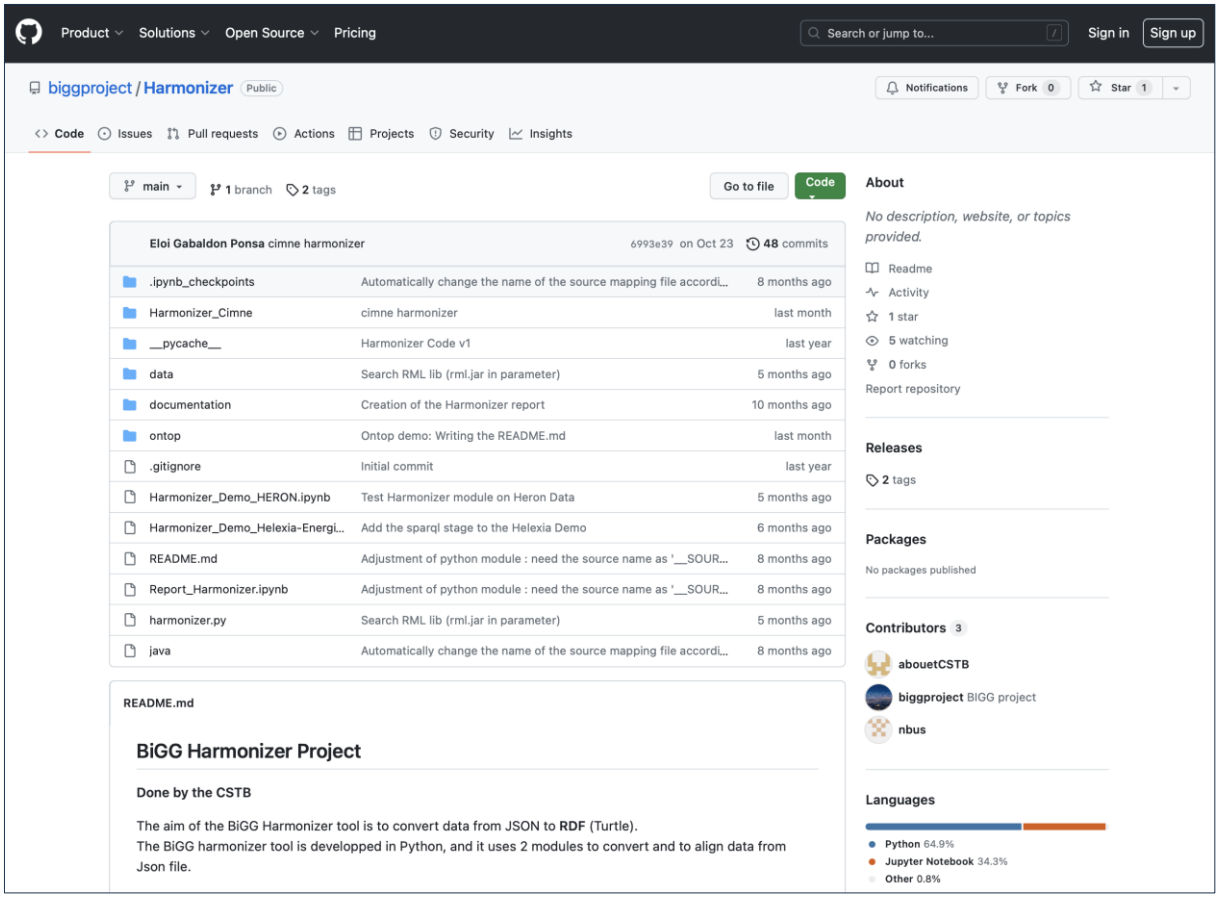


Figure 2 - BIGG project ontology repository (**https://github.com/biggproject/Ontology**)

II.2. Harmonizer

The repository contains the source code and the documentation of the BIGG harmonizer tools. It includes the BIGG Harmonizer tool is developed by the CSTB in Python and aims to convert data from JSON to RDF (Turtle) and produces RML mapping for this. It includes also the developed by CIMNE custom Python library that transforms heterogeneous data formats into the harmonised RDF format following the BIGG Ontology.

The repository contains instructions, examples, descriptions of data sources and mapping files.



II.3. biggr

This repository contains an open-source R package for the statistical analysis of building's data within the framework of BIGG project. It represents the part of the AI toolbox that allows the clustering, classification and modelling of building time series and its metadata. The package is prepared to take its input following the BIGG Ontology and the authors recommend its usage when elaborating data analytics pipelines with biggr.

The package contains instructions on the installation and all dependent R libraries, test files and examples.

II.4. biggpy

This repository contains an open-source Python library which provides all the tools necessary for building machine learning pipelines within the AI Toolbox and the related business cases and use cases of the BIGG project. For the language-agnostic documentation, please refer to: https://github.com/biggproject/biggdocs.

The repository contains usage examples as Jupyter notebooks.

II.5. biggdocs

This repository contains the language-agnostic documentation for the use of the R and Python libraries implemented on the framework of the BIGG AI Toolbox. The documentation includes the following modules:

- Data preparation
- Data transformation
- Modelling
- Reinforced learning

The repository contains also references to the data model and data type definitions for use with the AI Toolbox.

II.6. A1 - Energy benchmarking of buildings

This is an application for energy benchmarking of buildings. This is a tutorial pipeline that uses BIGG Ontology harmonised data as input to generate results for the energy benchmarking of buildings. The algorithm to perform this task was developed during the implementation of Business Case 1 of the project.

The benchmarking of buildings is made using a set of multi-dimensional Key Performance Indicators (KPIs) related with energy use, cost, and emissions. It can be performed from a longitudinal perspective, when compare KPIs of one building across its own timeline, and from a cross-sectional point of view, when compare results among similar buildings.

The repository contains software installation requirements, methods & diagrams, description of input and output data, instructions, and examples.

II.7. A2 - Energy Efficiency Measures assessment

This is an application pipeline for assessment of Energy Efficiency Measures. This pipeline uses Measurement & Verification (M&V) techniques to assess Energy Efficiency Measures (EEMs) applied in buildings. The pipeline is based on the BIGG Ontology harmonised data and uses the package biggr to generate the results.

The repository contains software installation requirements, methods & diagrams, description of input and output data, instructions, and examples.

II.8. A3 - Energy Performance Contract baseline identification

This is an application pipeline for Energy Performance Contract baseline identification. It is goals is to track the savings realised by energy conservation measures undertaken by an ESCO on a daily/weekly/monthly basis (M&V). The process to estimate savings involves the identification of a baseline model for the electricity consumption of the building before the retrofit period, e.g. before new equipment is installed. This model will be used to predict the electricity consumption after the renovation, i.e. post-retrofit, to know what the consumption would have been without retrofit. The real consumption post-retrofit and the predicted one are then compared to estimate savings.

The repository contains information about the necessary data input, instructions on its use, and examples.

II.9. A4 - Occupancy pattern detection

This is an application pipeline for detection of building occupancy pattern. The goal of this business case is to forecast the occupancy of a building/zone based on movement sensor data. The occupancy forecasts or the occupancy model produced can be used for example by edge devices to control the HVAC equipment of the building in a smart and autonomous way or to improve other models.

The repository contains information about the necessary data input, instructions on its use, and examples.

II.10. A6 - Gas Demand Response

The goal of this application is to develop a demand response (DR) scheme exploiting gas flexibility in space heating for residential complex. Gas providers can avoid additional costs or CO2 emissions by letting customers play an active role in DR, thereby restoring the balance between supply and demand. The flexible assets used in this project for providing DR are gas-based domestic hot water boilers used for residential hot water and heating and the objective is to is to meet a targeted gas consumption level.

The repository contains information on implementation, configuration, model training, and examples.

II.11. Ingestor skeleton

This is a HTTP based ingestor skeleton project, it uses Spring Boot REST. The ingestor also uses Spring Kafka, all properties are configurable using the present Spring Kafka properties. The goal of this project is to provide an implementation of a HTTP ingestor that allows for a custom implementation.

Provides specifications, methods, properties and information on configuration.

II.12. MQTT ingestor

This repository contains a Python code implementation of MQTT ingestor for the BIGG Reference Architecture. The repository provides specifications, configuration information and examples.

II.13. HTTP ingestor

This is a HTTP based ingestor that uses Spring Boot REST for the BIGG Reference Architecture. The ingestor also uses Spring Kafka, all properties are configurable using the present Spring Kafka properties. The repository provides specifications, configuration information and examples.

II.14. BC1-2-3 ingestors and harmonizers

This repository contains the ingestors and harmoniser tools for the heterogeneous data from proprietary and public data sources used for business cases 1 - 3, including mapping files and data sources description.

III. CONCLUSIONS

All BIGG project's public data has been made available in a public GitHub repository:

https://github.com/biggproject

The repository contains the information and data that is not sensitive according to the stipulated in the BIGG Data and Ethics Management Plan (D1.3) and the BIGG Consortium Agreement and includes the open documentation of the BIGG Ontology and the open-source software for the BIGG AI Toolbox, the latter available with a MIT licence allowing copying, distribution, and use of the software without limitations only requiring preservation of copyright and license notices.