NRG4CAST
FP7-2012-NMP-ENV-ENERGY-ICT-EeB
Contract no.: 600074
www.nrg4cast.org

NRG4CAST

Deliverable D2.1

Initial data processing and specific data adapters

Editor:	Ioannis Chamodrakas, SingularLogic
Author(s):	Ioannis Chamodrakas, SingularLogic; Kostas Kalaboukas, SingularLogic; Simon Tzanakis, SingularLogic
Reviewers	Maja Škrjanc (JSI)
Deliverable Nature:	Prototype (P)
Dissemination Level: (Confidentiality)[1]	Public (PU)
Contractual Delivery Date:	May 2013
Actual Delivery Date:	May 2013
Suggested Readers:	Data integration software developers, Energy monitoring/forecasting software developers
Version:	1.0
Keywords:	Data acquisition, Data adapters, Data pre-processing

---

Disclaimer

This document contains material, which is the copyright of certain NRG4CAST consortium parties, and may not be reproduced or copied without permission.

All NRG4CAST consortium parties have agreed to full publication of this document.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the NRG4CAST consortium as a whole, nor a certain party of the NRG4CAST consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Executive Summary

The purpose of this deliverable report is to present the prototype, which was developed with respect to the initial data processing and implementation of specific data adapters for all available data resources of the NTUA and the Miren-FIR-CSI Piemonte pilot cases in the context of the NRG4CAST project.

The data adapters were developed as software components within the NRG4CAST data gathering infrastructure which was described in D1.5, i.e. the Open Grid Software Architecture – Data Access and Integration platform.

The available data resources for the NTUA and the MFC pilot cases integrated within the data gathering infrastructure through the development of the data adapters are the following:

NTUA Campus pilot case

NTUA Weather data (temperature, rainfall, humidity, pressure, wind direction)

EUROSTAT Gas and Electricity Prices Database

NTUA Total Gas and Electricity Consumption

NTUA Hydraulics Building Electricity Consumption Sensor

NTUA Lampadario Building Electricity Consumption Sensor

MFC pilot case

CSI Weather data (temperature, rainfall, humidity, pressure, wind direction)

Energy 3D Cadastre for the City of Turin

Energy Performance Certificates of Buildings in Piemonte Region

Heating Installations certification data in Piemonte Region

The developed data adapters may be easily modified and /or extended to integrate further data resources of each pilot case.

The data resources are exposed through a Web Service interface for data cleaning, data fusion (D2.2), data distribution (D2.4) and data stream integration (D6.3).

.

Table of Contents

Abbreviations

CSV	Comma Separated Values
DBF	dBase Filename extension
EUROSTAT	Statistical Office of the European Union
GZIP	GNU zip compressed file format
HTS	Hierarchical Time Series
JAVA	JAVA programming language
OGSA-DAI	Open Grid Software Architecture – Data Access and Integration
TSV	Tab Separated Values
WSDL	Web Service Description Language

1 Introduction

Deliverable 2.1 aims at implementing initial data adapters, which include methods and services for data acquisition from the data resources of the NRG4CAST pilot cases. The available data resources for each pilot case are different in type, format, content and quality.

Initially, the data resources have been pre-processed to detect and correct basic problems such as indicating missing values or performing required transformations so that the data cleaning and data fusion methods will function properly.

The data adapters have been created as software components of the NRG4CAST data gathering infrastructure which was described in Deliverable 1.5, i.e. the OGSA-DAI platform. The data resources were integrated as indexed file or relational resources within the NRG4CAST OGSA-DAI platform according to their different format and type.

Further, Data Consolidation Web Services have been implemented hiding low-level details and complexities of the OGSA-DAI platform in order to expose the integrated data resources in a simple and effective way.

The developed data adapters may be easily modified and /or extended to integrate further data resources of each pilot case.

The remainder of this deliverable is structured as follows: Section 2 describes the initial processing of the data resources. Section 3 presents the different data adapters created for all available data resources as well as how the integrated data resources are exposed through a Web Service interface. Finally, Section 4 summarizes the contents of the deliverable.

2   Initial data processing

This section reports initial processing of all available data resources for the NTUA and the MFC pilot cases.

2.1   NTUA University Campus pilot case

2.1.1   NTUA Weather data

NTUA Weather data are accessible through the Hydrological Observatory of Athens web site which provides access to the measurements of the meteorological station sited inside the National Technical University campus in Zografou (http://hoa.ntua.gr/stations/d/13/). The data are provided as HTS files in plain text TSV format.

The main issue that has been identified with respect to the specific data resources, which include pressure, humidity, rainfall, temperature and wind direction measurements, is the existence of missing values which have been appropriately indicated so that data cleaning and fusion methods will function properly.

2.1.2   EUROSTAT Electricity and Gas Prices 2003-12

EUROSTAT Electricity and Gas Prices are accessible through the EUROSTAT web site (http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?dir=data&sort=1&sort=2&start=n). The data are provided as compressed (gzipped) plain text CSV files.

No issue has been identified concerning the EUROSTAT prices data.

2.1.3   NTUA Total Gas / Electricity consumption

NTUA Total Gas / Electricity consumption data are provided as static Excel spread sheets which are periodically updated by responsible NTUA staff.

No issue has been identified concerning the NTUA Total Gas / Electricity consumption data.

2.1.4   NTUA Existing Sensor Data for Electricity Consumption (Hydraulics / Lampadario Buildings)

NTUA electricity consumption sensor measurement data are provided as CSV files. The raw measurement values are multiplied by a constant factor which is a characteristic of the specific sensor hardware.

2.2   MFC pilot case

2.2.1   CSI Weather data

CSI Weather data are initially provided as a CSV file.

No issue has been identified concerning the CSI Weather data.

2.2.2   Energy 3D Cadastre for the City of Turin (ENERCAD)

The Energy 3D Cadastre for the City of Turin is initially provided as a dBase file (.dbf).

No issue has been identified concerning the ENERCAD3D data.

2.2.3   Energy Performance Certificates of Buildings in Piemonte Region (SICEE)

The SICEE Energy Performance building certificates data are initially provided as a CSV file.

No issue has been identified concerning the SICEE data.

2.2.4   Heating Installations certification data in Piemonte Region (SIGIT)

The SIGIT heating installations certification data are initially provided as a CSV file.

No issue has been identified concerning the SICEE data.

2.2.4   Heating Installations certification data in Piemonte Region (SIGIT)

3   Data adapters

This section reports the development of the initial data adapters for all available data resources of the NTUA and the MFC pilot cases

3.1   NTUA University Campus pilot case

3.1.1   NTUA Weather data

The NTUA Weather data are integrated as indexed file data resources within the NRG4CAST OGSA-DAI platform. In order to create the appropriate Lucene indexes a set of special indexing Java classes has been developed (gr.singularlogic.nrg4cast.ogsadai.activity.indexedfile.NTUAMeteo*) which are provided in the enclosed source code compressed file (singularlogic.nrg4cast.data.adapters.zip) [1].

In order to provide access to the NTUA Weather data a set of Web service operations are exposed which are described in the WSDL file provided online at:

 http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the NTUA Weather data is located at the Integration Tier of the NRG4CAST toolkit architecture (see Deliverable 1.3 – Data Consolidation Services) [2] in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through the RESTful services (Data Access Tier) it exposes and returns the results to the client in a form which is described in the WSDL document provided above. The source code of the NTUA Weather data Web service operations is provided in the enclosed compressed file [1].

3.1.2   EUROSTAT Electricity and Gas Prices 2003-12

The EUROSTAT Electricity and Gas Prices data are integrated as indexed file data resources within the NRG4CAST OGSA-DAI platform. In order to create the appropriate Lucene indexes a special indexing Java class has been developed (gr.singularlogic.nrg4cast.ogsadai.activity.indexedfile.PricesIndexWriter), provided in the enclosed compressed file of the source codes [1]. Since the provided data are in a gzipped compressed format, the indexing class is responsible to first decompress the files into plain text CSV format before proceeding with indexing.

In order to provide access to the EUROSTAT prices data a set of Web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the EUROSTAT Electricity and Gas Prices data is located at the Integration Tier of the NRG4CAST toolkit architecture in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through its exposed RESTful services (Data Access Tier) and returns the results to the client in a form which is described in the WSDL document provided above. The source code of the EUROSTAT Electricity and Gas Prices Web service operations is provided in the enclosed compressed file [1].

3.1.3   NTUA Total Gas / Electricity consumption

The NTUA Total Gas / Electricity consumption data are integrated as indexed file data resources within the NRG4CAST OGSA-DAI platform. In order to create the appropriate Lucene indexes two special indexing Java classes have been developed (gr.singularlogic.nrg4cast.ogsadai.activity.indexedfile.NTUATotal*) provided in the enclosed compressed file of the source codes [1]. Since the data are provided as excel spreadsheets, the indexing class is responsible to first extract the appropriate Excel sheets in plain text CSV format before proceeding with indexing. The extraction of the spread sheets in CSV format is performed through the use of the Apache POI library [3].

In order to provide access to the NTUA Total Gas / Electricity consumption data a set of Web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the NTUA Total Gas / Electricity consumption data is located at the Integration Tier of the NRG4CAST toolkit architecture in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through its exposed RESTful services (Data Access Tier) and returns the results to the client in a form which is described in the WSDL document provided above. The source code of NTUA Total Gas / Electricity consumption Web service operations is provided in the enclosed compressed file [1].

3.1.4 NTUA Existing Sensor Data for Electricity Consumption (Hydraulics / Lampadario Buildings)

The NTUA sensor data are integrated as indexed file data resources within the NRG4CAST OGSA-DAI platform. In order to create the appropriate Lucene indexes a special indexing Java class has been developed (gr.singularlogic.nrg4cast.ogsadai.activity.indexedfile.NTUAExistingSensorDataIndexWriter) provided in the enclosed compressed file of the source codes [1].

 In order to provide access to the NTUA Total Gas / Electricity consumption data a set of Web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the NTUA sensor data is located at the Integration Tier of the NRG4CAST toolkit architecture in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through its exposed RESTful services (Data Access Tier) and returns the results to the client in a form which is described in the WSDL document provided above. The source code of NTUA Sensor Data Web service operations is provided in the enclosed compressed file [1].

3.2 MFC pilot case

3.2.1 CSI Weather data

The CSI Weather data are integrated as an indexed file data resource within the NRG4CAST OGSA-DAI platform. In order to create the appropriate Lucene indexes a special indexing Java class has been developed (gr.singularlogic.nrg4cast.ogsadai.activity.indexedfile.CSIWeatherIndexWriter) which is provided in the enclosed compressed file [1].

In order to provide access to the CSI Weather data a set of Web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the CSI Weather data is located at the Integration Tier of the NRG4CAST toolkit architecture (see Deliverable 1.3 – Data Consolidation Services) [2] in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through the RESTful services (Data Access Tier) it exposes and returns the results to the client in a form which is described in the WSDL document provided above. The source code of the CSI Weather data Web service operations is provided in the enclosed compressed file [1].

3.2.2 Energy 3D Cadastre for the City of Turin (ENERCAD)

The ENERCAD data are integrated as a relational data resource within the NRG4CAST OGSA-DAI platform. For this purpose, an open source JDBC driver for dBase files has been utilized [4].

In order to provide access to the ENERCAD data a set of web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the ENERCAD data is located at the Integration Tier of the NRG4CAST toolkit architecture (see Deliverable 1.3 – Data Consolidation Services) [2] in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through the RESTful services (Data Access Tier) it exposes and returns the results to the client in a form which is described in the WSDL document provided above. The source code of the ENERCAD data web service operations is provided in the enclosed compressed file [1].

3.2.3   Energy Performance Certificates of Buildings in Piemonte Region (SICEE)

The SICEE Energy Performance building certificates data are integrated as a relational data resource within the NRG4CAST OGSA-DAI platform. For this purpose, an open source JDBC driver has been utilized [4]

In order to provide access to the SICEE data a set of web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the SICEE data is located at the Integration Tier of the NRG4CAST toolkit architecture (see Deliverable 1.3 – Data Consolidation Services) [2] in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through the RESTful services (Data Access Tier) it exposes and returns the results to the client in a form which is described in the WSDL document provided above. The source code of the SICEE data web service operations is provided in the enclosed compressed file [1].

3.2.4   Heating Installations certification data in Piemonte Region (SIGIT)

The SIGIT Heating Installations certification data are integrated as a relational data resource within the NRG4CAST OGSA-DAI platform. For this purpose, an open source JDBC driver has been utilized [4]

In order to provide access to the SIGIT data a set of web service operations are exposed which are described in the WSDL file provided online at:

http://83.212.123.209:8080/NRG4CASTServices/services/Nrg4CastServicesPort?wsdl.

The Web service providing access to the SIGIT data is located at the Integration Tier of the NRG4CAST toolkit architecture (see Deliverable 1.3 – Data Consolidation Services) [2] in order to hide low-level details of the OGSA-DAI platform. Specifically the Web service submits an appropriate workflow to the OGSA-DAI platform through the RESTful services (Data Access Tier) it exposes and returns the results to the client in a form which is described in the WSDL document provided above. The source code of the SIGIT data web service operations is provided in the enclosed compressed file [1].

4   Summary

This deliverable reports work performed with regard to initial data processing and the creation of specific adapters for all the data resources of the NTUA Campus and the MFC pilot cases which are available at this point of time.

The developed software components are available at https://www.dropbox.com/s/jwfdt6c4dtjtftq/singularlogic.nrg4cast.data.adapters.zip and are implemented in a way that enables easy modification and adaptability (easy integration of further data resources) if there will be the need to so, in the further steps of the project.

References

[1]   https://www.dropbox.com/s/jwfdt6c4dtjtftq/singularlogic.nrg4cast.data.adapters.zip

[2]   NRG4CAST Deliverable 1.3.

[3]   Apache POI - the Java API for Microsoft Documents, documentation available at: http://poi.apache.org/.

[4]   CsvJdbc: a Java database driver for reading comma-separated-value and DBF files, documentation available at: http://csvjdbc.sourceforge.net/.