

---

# **SciGRID\_gas: The raw CONS data set**

*Release 1.0*

**J.E. Sandoval, J.C. Diettrich, A. Pluta, W. Medjroubi**

**Jun 14, 2021**



# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Project information . . . . .	7
1.2	Background . . . . .	8
1.3	Project goal . . . . .	9
1.4	Document overview . . . . .	10
<b>2</b>	<b>Data structure</b>	<b>11</b>
2.1	Data structure description . . . . .	11
2.1.1	Terminology . . . . .	11
2.2	Summary . . . . .	16
<b>3</b>	<b>Data sources</b>	<b>17</b>
3.1	Non-OSM data . . . . .	18
3.2	Gas consumers data set . . . . .	20
3.2.1	Input data sets for the CONS data set . . . . .	20
3.2.2	Incorporating the CONS data into the SciGRID_gas data frame . . . . .	22
3.2.3	Copyright and disclaimer for the INET data set . . . . .	23
3.2.4	Summary CONS data . . . . .	24
3.3	Data summary . . . . .	26
3.4	Summary . . . . .	26
<b>4</b>	<b>Conclusion</b>	<b>27</b>
<b>5</b>	<b>Appendix</b>	<b>29</b>
5.1	Glossary . . . . .	29
5.2	Unit conversions . . . . .	31
5.3	Attribute <i>exact</i> . . . . .	31
5.4	Location name alterations . . . . .	32
5.5	Country name abbreviations . . . . .	32
5.6	Acknowledgement . . . . .	33
	<b>Bibliography</b>	<b>35</b>



*How to cite*

J.E. Sandoval, J.C. Dietrich, A. Pluta, W. Medrjoubi  
SciGRID\_gas: The raw CONS data set  
German Aerospace Center (DLR), Institute for Networked Energy Systems  
Germany  
doi: 10.5281/zenodo.4922515

*Impressum*

German Aerospace Center (DLR), Institute for Networked Energy Systems  
Carl-von-Ossietzky-Str. 15  
26129 Oldenburg  
Germany  
Tel.: +49 (441) 999 060





## LIST OF FIGURES

2.1	Data structure for the SciGRID_gas data set. . . . .	12
3.1	Overview map of the CONS data set for Europe. . . . .	25



## LIST OF TABLES

3.1	CONS component summary . . . . .	24
5.1	Dataset abbreviations . . . . .	29
5.2	Glossary . . . . .	30
5.3	Unit conversions . . . . .	31
5.4	Unit conversions . . . . .	31
5.5	Country codes . . . . .	32

### Summary

The goal of SciGRID\_gas is to develop methods to create an automated process that can generate a gas transmission network data set for Europe. Gas transmission networks are fundamental for simulations by the gas transmission modelling community, to derive major dynamic characteristics. Such simulations have a large scope of application, for example, they can be used to perform case scenarios, to model the gas consumption, to minimize leakages and to optimize overall gas distribution strategies. The focus of SciGRID\_gas will be on the European transmission gas network, but the principal methods will also be applicable to other geographic regions.

Data required for such models are the gas facilities, such as compressor stations, LNG terminals, pipelines, etc. One needs to know their locations, in addition to a large range of attributes, such as pipeline diameter and capacity, compressor capacity, configuration, and consumption etc. Most of this data is not freely available. However throughout the SciGRID\_gas project it was determined, that data can be found and grouped into two fundamental different groups: a) OSM data, and b) non-OSM data. The OSM data consists of geo-referenced facility data that is stored in the OpenStreetMap (OSM) data base, and is freely available. However, the OSM data set currently contains hardly any other information than the location of the facilities. The Non-OSM data set can fill some of those gaps, by supplying information such as regional gas demand, pipeline diameter, compressor capacity and more. Part of the SciGRID\_gas project is to mine and collate such data, and combine it with the OSM data set.

Here, this document describes one of the non-OSM data sets, called the “CONS” data set (consumers), which was generated by Sandoval :cite:’Sandoval\_COMS\_2021’ in a Masters project as part of the SciGRID\_gas project. This document explains the origin and structure of this single data sets.

In this document, the chapter “Introduction” will supply some background information on the SciGRID\_gas project, followed by the chapter “Data structure”, that gives a detailed description of the data structure that is being used in the SciGRID\_gas project. Chapter “Data sources” describes the CONS data set.

The appendix contains a glossary, references, location name alterations convention and finishes with the table of country abbreviation.

## INTRODUCTION

**SciGRID\_gas** is a three-year project funded by the German Federal Ministry for Economic Affairs and Energy [BMWi20] within the funding of the 6. Energieforschungsprogramm der Bundesregierung [BMWi11].

The goal of SciGRID\_gas is to develop methods to generate and provide an open-source gas network data set and corresponding code. Gas transmission network data sets are fundamental for the simulations of the gas transmission within a network. Such simulations have a large scope of application, for example, they can be used to preform case scenarios, to model the gas consumption, to detect leaks and to optimize overall gas distribution strategies. The focus of SciGRID\_gas will be the generation of a data set for the European Gas Transmission Network, but the principal methods will also be applicable to other geographic regions.

Both the resulting method code and the derived data will be published free of charge under appropriate open-source licenses in the course of the project. This transparent data policy shall also help new potential actors in gas transmission modelling, which currently do not possess reliable data of the European Gas Transmission Network. It is further planned to create an interface to SciGRID\_power [MMK16] or heat transmission networks. Simulations on coupled networks are of major importance to the realization of the German *Energiewende*. They will help to understand mutual influences between energy networks, increase their general performance and minimize possible outages to name just a few applications.

This project was initiated, and is managed and conducted by the German Aerospace Center (DLR), Institute for Networked Energy Systems.

### 1.1 Project information

- **Project title:** Open Source Reference Model of European Gas Transport Networks for Scientific Studies on Sector Coupling (*Offenes Referenzmodell europäischer Gastransportnetze für wissenschaftliche Untersuchungen zur Sektorkopplung*)
- **Acronym:** SciGRID\_gas (Scientific GRID gas)
- **Funding period:** January 2018 - July 2021
- **Funding agency:** Federal Ministry for Economic Affairs and Energy (*Bundesministerium für Wirtschaft und Energie*), Germany
- **Funding code:** Funding Code: 03ET4063
- **Project partner:** German Aerospace Center (DLR), Institute for Networked Energy Systems.



Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages

## 1.2 Background

As of today, only limited data of the facilities of the European Gas Transmission Networks is publicly available, even for non-commercial research and related purposes. The lack of such data renders attempts to verify, compare and validate high resolution energy system models, if not impossible. The main reason for such sparse gas facility data is often the unwillingness of transmission system operators (TSOs) to release such commercially sensitive data. Regulations by EU and other lawmakers are forcing the TSOs to release some data. However, such data is sparse and too often not clearly understandable for non-commercial users, such as scientists.

Hence, details of the gas transmission network facilities and their properties are currently only integrated in in-house gas transmission models which are not publicly available. Thus, assumptions, simplifications and the degree of abstraction involved in such models are unknown and often undocumented. However, for scientific research those data sets and assumptions are needed, and consequently the learning curve in the construction of public available network models is rather low. In addition, the commercial sensitivity also hampers any (scientific) discussion on the underlying modelling approaches, procedures and simulation optimization results. At the same time, the outputs of energy system models take an important role in the decision-making process concerning future sustainable technologies and energy strategies. Recent examples of such strategies are the ones under debate and discussion for the Energiewende [BundesregierungDeutschland20] in Germany.

In this framework, the SciGRID\_gas project initiated by the German Aerospace Center (DLR), Institute for Networked Energy Systems (Oldenburg, Germany) aims to build an open source model of the European Gas Transmission network. Releasing SciGRID\_gas as open-source is an attempt to make reliable data on the gas transmission network available. Appropriate (open) licenses attached to gas transmission network data ensures that established models and their assumptions can be published, discussed and validated in a well-defined and self-consistent manner. In addition to the gas transmission network data, the Python software developed for building the model SciGRID\_gas will be published under the GPLv3 license.

The main purpose of the SciGRID\_gas project is to provide freely available and well-documented data on the European gas transmission network. Further, with the documentation and the Python code, users should be able to generate the data on their own computers.

The input data itself is based on data available from [openstreetmap.org](http://openstreetmap.org) (OSM) under the Open Database License (ODbL) as well as Non-OSM data gathered from different sources, such as Wikipedia pages, fact sheets from TSOs or even newspaper articles.

The main workload of this project is to:

- retrieve the OSM and Non-OSM data sets for the gas infrastructure
- merge all available data sets

- build a gas transmission component data set
- generate missing data using heuristic methods
- document the process and the output.

The first step of the project was to collate a Non-OSM data set by searching the web for metadata that will be useful for the project. This included information, such as pipelines, compressors, LNG terminals, and their attributes, such as diameters, capacities etc.

Here a single data source will be introduced, the **CONS** data set (CONS stands for the estimated gas consumer data set).

This multi-stage release will allow us to easily and effectively incorporate feedback from potential users during the lifetime of the project. Those releases can be downloaded from the SciGRID\_gas webpage with documentation, and can be seen as a snapshot of the current research project state.

Further information on the project can be found on the SciGRID\_gas web page: <https://www.gas.scigrid.de/pages/imprint.html>.

The web page is maintained throughout the project lifetime, and will contain information on:

- General project information
- Contact details
- Presentations
- Bug/data fixes
- Data, code and documentation releases
- Publications.

As part of the SciGRID\_gas webpage, one can also sign up to the SciGRID\_gas newsletter by sending an email to [news.gas-subscribe@scigrid.de](mailto:news.gas-subscribe@scigrid.de)

## 1.3 Project goal

The overall goals of the SciGRID\_gas project are:

- **Data output:** Creation of customisable gas transmission network data sets.
- **Open source:** Any one can download the data, make changes to it, pass it on to others, or even use it in commercial projects, as long as the SciGRID\_gas project is mentioned as the original source of the data (CC by).
- **Application:** The outcome of the project can be used for a variety of scientific applications (e.g. sector coupling, entry-exit models etc.).
- **Transparency:** The Python code, the documentation and the data (that can be passed on under copyright licences) is supplied.
- **Extendability:** Every user can extend the software code to their needs. However, we would encourage users to update and maintain the original git-repository and documentation for others.
- **Feedback:** Through constant data releases, it is hoped that the output data set will improve in quality and quantity by constantly incorporating feedback from the research community.

## 1.4 Document overview

This is an overview of the SciGRID\_gas documentation, as this will help the user to better understand the overall project, its aims and the steps that were taken to obtain/model the data set.

SciGRID\_gas has been coded in Python, and hence, with that came the overall data structure that was selected for the project. As this is the most fundamental aspect for anyone wanting to use the data and the code, it is described first. Under **Data Structure** we go into the definitions of the terms *Components*, *Elements*, and *Attributes*. We also give an overview on the internal workings of the SciGRID\_gas source code.

A fundamental building block for the SciGRID\_gas project is the data itself. Overall, we have classified the data into two groups: OSM and non-OSM data. The chapter **Data Sources** contains background information on NO data set only. Information is supplied, such as how the data was collated and how it was implemented into the SciGRID\_gas data sets structure. In addition, an overview of the extent of the data will also be given for each data source, e.g. the number of elements or the list of attributes that the data contains.

The document also contains the chapter **Appendix** that contains sub-sections, such as *Glossary*, *References* etc.

## DATA STRUCTURE

A well designed and documented data structure is fundamental in any large-scale project. Good data structure in combination with tools, based on algorithms, improve the performance of any project output.

This structure needs to represent the gas flow facilities as good as possible. Hence, it needs to include components, such as pipelines, compressors etc. A finite number of components have been identified that are required as building blocks of a gas network. In addition, each component will contain attributes, such as pipeline diameter, maximal operating pressure, maximal capacity, number of turbines etc.

It is anticipated that the adopted data structure can be implemented in different types of gas flow models and will be used by the research community for topics, such as sector coupling or identifying gas transmission bottlenecks.

Within the SciGRID\_gas project, the structure of the data model is part of classes defined within the Python code. Alterations may occur over the duration of the project, but it is envisaged that those will be small, and that compatibility will be assured.

The goal of this section is to describe in details the data structure that has been adopted and implemented into the Python code. This will be important in understanding other aspects of this document, such as exporting the data into CSV files or generating missing values.

### 2.1 Data structure description

This section contains information on the SciGRID\_gas data structure, the format, and the code that can be used to import publicly available data into the project, so that it can be used in subsequent steps. Paramount for an understanding of the data structure is a good understanding of the terminology used throughout this section and the document in general. Hence, terminology will be introduced in the following sub-section.

#### 2.1.1 Terminology

Throughout this document certain terms will be used, which will be described below and have been summarized in Figure 2.1.

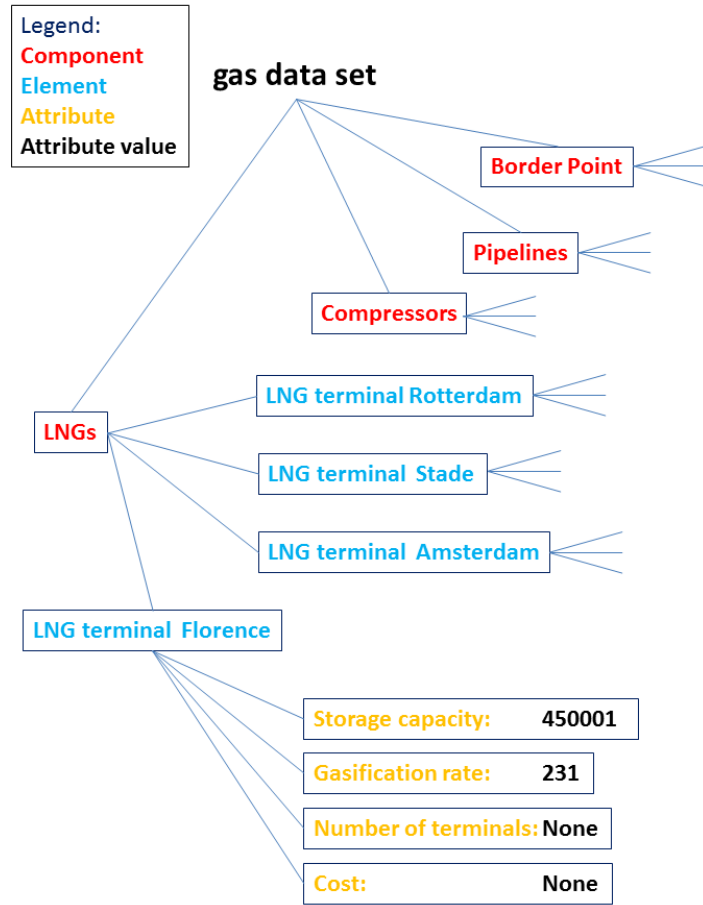


Figure 2.1: Data structure for the SciGRID\_gas data set.

## Gas transmission network

The term “gas transmission network” describes the physical gas transmission grid. This does not include the distribution of gas through gas distribution companies, but includes the long distance transmission of gas from producer countries to consumer countries, as carried out by the Transmission System Operators (TSO) [Wik20g]. In addition, throughout this document, the terms “transportation” and “transmission” are seen as interchangeable, and hence, will both be used describing the same.

## Gas component data set

The term “gas component data set” is used for all raw data of objects/facilities that have been loaded using SciGRID\_gas tools into a Python environment. Gas component data sets are used as input into our SciGRID\_gas project. Several data sources can be loaded as gas component data sets, and then combined into a single gas component data set. However, not all elements (e.g. compressors) must be connected to pipelines. Hence, such a data set is referred to as a “gas component data set”.

## Gas network data set

A “gas component data set” can be converted into a “gas network data set”, by connecting all non-pipeline elements to nodes and all nodes are connected to pipelines, and as part of the process all network islands have been connected or removed, resulting in a single network. Therefore, the network contains nodes and edges which are coherently connected, and all objects with the exception of pipelines are associated with nodes in this network, whereas pipelines are associated with edges.

## Component

There are several component types in a gas transmission network, such as compressors, LNG terminals, or pipelines. In Figure 2.1 they are coloured red. Hence, whenever the word “component” is mentioned, it refers to one of these components. There are roughly a dozen different components that will form a gas network data set. They will be briefly explained below.

## Element

The term “element” refers to individual facilities, e.g. the LNG terminal in Rotterdam, or the compressor in Radeland. In Figure 2.1 they are coloured blue. The first one is an element of the component LNG terminals, whereas the second one is an element of the component compressors. Hence, many elements make up a component. However, all elements are referring to different facilities by default. This means in a single network, one cannot have two elements of a component describing the same facility. The structure of elements is described below.

## Attribute

“Attribute” is a term that is being used for the individual parameters that are associated with the elements. Examples of this term are gas “pipeline diameter”, “maximum capacity”, “max gas pipeline pressure”, to name just a few and in Figure 2.1 they are coloured yellow. Overall there will be several hundred attributes in the SciGRID\_gas project. However, the same attributes can occur in more than one component, e.g. “max flow capacity” exist for pipelines and also for compressors. Throughout the project, we have tried to keep the units of such attributes the same, so that there is no unit conversion required.

### Attribute value

Each attribute has a value, most likely a number or a string. In [Figure 2.1](#) they are coloured black. While booleans (*True/False*) are also allowed, more likely a “1” will stand for *True* and “0” for *False*. However, not all attribute values are given. Therefore, a no value for attribute values needs to be specified. In the SciGRID\_gas Python code it is *None*.

The [Figure 2.1](#) depicts the relationships between the terms “gas data set”, “component”, “element”, “attribute”, and “attribute value”. As can be seen, a single gas data set consists of several components. On the next level, each component contains several elements. Further, each element has several attributes, where each attribute has a single or several values. The heuristic processes described in this document at a later stage will fill all missing values with heuristically generated values.

### Gas component types

A gas transmission network consists of different components, such as pipelines, compressors etc. For the SciGRID\_gas project a hand-full of components have been implemented, and will be described here briefly:

- *Nodes*: In a gas network, gas flows from one point to another point, which are given through their coordinates. All elements of all other components (such as compressor stations and power plants) have an associated node, which allows for the geo-referencing of each element. Overall the term *Nodes* will be used throughout this document, as it aligns with graph theory aspects.
- *PipeLines*: *PipeLines* allow for the transmission of the gas from one node to another. *PipeLines* are georeferenced by an ordered list of nodes.
- *PipeSegments*: *PipeSegments* are almost identical to *PipeLines*. However, are only allowed to connect two nodes. Hence, any *PipeLines* element (with 3 or more nodes) can easily be converted into multiple *PipeSegments* elements.
- *Compressors*: *Compressors* represent compressor stations, which increases the pressure of the gas, and hence, allows the gas to flow from one node to another node. A gas compressor station contains several gas compressors units (turbines).
- *LNGs*: *LNGs* is the acronym for the LNG terminals and LNG storages, which there are several in Europe, as some gas gets transported to Europe via ships.
- *Storages*: *Storages* are a further network component. Surplus gas can be stored underground (e.g. in old gas fields or salt caverns), and used during low supply or high demand periods.
- *Consumers*: *Consumers* is the term used for gas users, which can be households, industry and commercial. This data set will be generated through a Master project, and excludes power plants.
- *PowerPlants*: *PowerPlants* is the term used for gas use by power plants only.
- *Productions*: These can be wells inside a country where gas is pumped out of the ground. Most of the gas used in Europe comes from outside of the EU. However, there are several smaller gas production sites scattered throughout Europe.
- *BorderPoints*: *BorderPoints* are facilities at borders between countries, which are mainly used to meter the gas flow from one country to another.

## Element structure

As described above, elements are describing individual facilities, such as compressors or LNG terminals. However, the overall structure of those elements is the same for all elements of all components, and is described as follows:

- *id*: A string that is the ID of the element, and must be unique.
- *name*: A string that is the name of the facility, such as “Compressor Radeland”. In most cases this is not supplied.
- *source\_id*: A list of strings that are the data sources of the element. As several elements from different sources could have been combined into a single element, one might need to know the original data sources.
- *node\_id*: The ID of a geo-referenced node to which an element of the network is associated to. For a compressor, this will be just a single *node\_id*. However, for a gas pipeline this entry would be a list of at least two *node\_id* values: the starts node id and the end node id.
- *lat*: The latitude value of an element. For elements of type *PipeLines* and *PipeSegments*, *lat* is a list of latitude values. Throughout the SciGRID\_gas project the projection World Geodetic system 1984 (epsg:4326) will be used.
- *long*: The longitude, analogue to lat.
- *country\_code*: A string indicating the 2-digit ISO country code (Alpha-2 code, see [Chapter 5.5](#) for list of countries and their codes) of the associated node of elements or list of nodes in case of *PipeLines* or *PipeSegments*.
- *comment*: An arbitrary comment that is associated with the element. In most cases this is not supplied.
- *tags*: This dictionary is reserved for OpenStreetMap data. It contains all associated key:value-pairs of an OpenStreetMap item.

In addition, there are three further groups of attributes to each element, which have been coded as Python “dictionaries”, named:

- *param*
- *method*
- *uncertainty*.

The structure within each dictionary is the same. The dictionary *param* (short for “parameter”) contains a list of attributes and their values. This list of attributes will be different for each component. For the component *PipeLines* they might be pipeline diameter, max pipeline pressure, and max pipeline capacity. For the component *Compressors* they might be, a number of turbines, overall turbine power, energy source of turbine or other.

The other two attribute dictionaries are *method* and *uncertainty*. Each of those two dictionaries contains exactly the same list of attributes as the *param* dictionary. However, their attribute values reflect the name of the dictionary. E.g. the attributes in the dictionary *method* contain the information on the method used to derive the attribute value that is stored in the *param* dictionary. Here methods of value generation can include heuristic methods names (in form of strings) that have been implemented in the SciGRID\_gas project. However, if attribute values are not being generated by the SciGRID\_gas project, but originate from one of the input data sources, then the attribute values in the *method* dictionary is set to “raw”.

See example below, for an *LNGs* element with the following entries:

- “make\_Attrib(const)”: the attributes *end\_year*, and *is\_H\_gas* have been set to a constant value
- “raw”, indicating that the two attributes *max\_cap\_store2pipe\_M\_m3\_per\_d* and *start\_year* contain original values
- “Lasso(max\_cap\_store2pipe\_M\_m3\_per\_d)”, here for the attribute *median\_cap\_store2pipe\_M\_m3\_per\_d* a method was used that is based on the lasso method and uses the attribute *max\_cap\_store2pipe\_M\_m3\_per\_d* as input.

Similar is the content of the *uncertainty* dictionary. It contains information on the uncertainty of the attributes from the *param* dictionary of that component. Again, all attributes listed in the *param* dictionary are also present in the *uncertainty* dictionary. The attribute values here reflect the uncertainty of the attribute. Here, it is assumed that attributes with a method of “raw” have an uncertainty of zero. Only for those attributes, which were generated during heuristic SciGRID\_gas methods an uncertainty larger than zero will be specified.

In addition, there are two special attributes in the *param* dictionary of the component *PipeSegments*: *path\_lat* and *path\_long*. These contain the waypoints between the start and the end node. They are separated in latitude and longitude, and are ordered in such a way, that they are following from the start node and go towards the end node.

## 2.2 Summary

The SciGRID\_gas software is designed to construct a gas transmission network data set from different open and non-open source gas component data sets. The gas transmission data set needs to be available and stored in a precise and predefined way, which was described in this section. We have identified several *component*-types of a gas transmission network grid, like pipelines, compressor stations, LNG-terminals etc. Each specific facility that falls under such a component is considered an *element* of that component. Each element is described by a list of *attributes* and correspondent *attribute values*, including information on the uncertainty of the attribute value and the way the attribute value was generated.

## DATA SOURCES

Original data sets describing gas transmission networks are the property of the transmission system operators (TSOs) and are generally not freely available in the form and depth that is required for modelling purposes. The major reason for the difficulty of obtaining of such data is that most of the gas network infrastructure, namely pipelines, is buried underground. Thus, a pipeline diameter is hard to estimate locally. In addition, almost all of the data is commercially sensitive.

Nevertheless, some data is made available by gas transmission network operators, through different channels. E.g. information on the size and number of compressors could be made public through a press release, as part of a refurbishment. An example is given below (<https://www.maz-online.de/Lokales/Teltow-Flaeming/Neue-Verdichterstation-entsteht-in-Radeland>):

“Die Eugal-Pipeline dient dazu, Gas aus der neuen Ostseepipeline Nord Stream 2 bis zur tschechischen Grenze zu leiten. 275 Kilometer von ihr verlaufen in Brandenburg. Grundsätzlich soll die neue Leitung parallel zur bestehenden Opal-Pipeline gebaut werden.”

In addition, some information can be found on company web pages, (<https://www.open-grid-europe.com/cps/rde/SID-752BB6B5-E0A975F2/oge-internet-preview/hs.xsl/NewsDetail.htm?rdeLocaleAttr=en&newsId=50190C3B-E14F-4685-9E64-E40EEAB57A28>):

“Open Grid Europe (OGE) is investing roughly EUR 150 million at its compressor station in Werne to improve the security and flexibility of energy supply for North Rhine-Westphalia and Germany. The upgrade of the station, which is one of the hubs of the pipeline network, will allow gas flows to be switched (reversed) from north to south and south to north. In addition, OGE is preparing the station for the upcoming transition from L- to H-gas. Through this fitness programme, the station’s transmission capacity will increase by about 500,000 to 6.5 million m<sup>3</sup>/h, which is equivalent to the annual consumption of more than 2,100 single-family homes. The project, which is due for completion at the end of 2018, is fully on track.”

However, there is a public drive to gather such data and subsequently make it available. The major platform through which this is occurring is the Open Street Map database [Hel18]. OSM is a geo-referenced database through which people can supply geo-referenced information on all man-made and natural structures, ranging from mountains to buildings. To achieve this, people throughout the world wander the globe and geo-reference everything that they can find. This also includes gas-pipeline markers, compressor stations or LNG terminals. However, the major problem remains that one cannot measure or estimate the diameter of the underground pipelines, or the number and size of the compressor turbines, as compressors are within buildings, which are fenced off. Hence, such information is hardly supplied to the OSM platform.

For the reasons mentioned above, the available data can be separated into two different groups:

- OSM data: Data can be found in the OSM data base. OSM data is well geo-referenced, but contains little meta-information (information on the facility attributes, such as pipeline diameter or pipeline capacity). OSM data is very helpful to obtain accurate routes of pipelines.
- Non-OSM data: Non-OSM data have in general lower geographical accuracy but contain a lot of meta-information. Unfortunately, such information is only known for a few facilities. One exception to this rule

are shapefiles from TSOs. They are rare, but well geo-referenced. However, the resolution of the meta information can vary from TSO to TSO.

The following section will introduce non-OSM data sets, and at a later stage, this will be followed by a section on the OSM data.

### 3.1 Non-OSM data

Non-OSM data includes data from internet research, TSO press releases, TSO transparency platform, TSO public data, national open-source gas network data sets<sup>1</sup> etc.

Some of the TSO information had to be made available due to EU-regulations. Other information has been made public as part of a company's self-presentation and advertisement. The information used by the SciGRID\_gas project focuses on:

- the quality of the data
- the format of the data
- the level of representation of the data
- and the copyright restrictions on the data.

In addition, each data source is unique. Source specific tools need to be developed, so that all data sources can be made accessible for the SciGRID\_gas project.

A significant portion of the project was spent on finding non-OSM data sets. Further data sources might be available, but unknown to the authors. If the authors are made aware of additional sources, the project will try to incorporate those, as this would only increase the depth of the data available and increase the applicability of the gas network data set and model.

Non-OSM data sources are very specific, addressing only certain aspects of the entire gas infrastructure. E.g. the GIE [GasIEurope20] data set supplies information on the daily gas flow in and out of gas storages in LNG terminals. However, they fall short on specifying the fundamental information of the actual physical location. Other data sets, such as the LKD [KKS+17] data set is quite detailed in respect of pipelines, compressors and consumptions, however, only available for Germany.

Hence, the main task is to look closely at each data source, distil which data attribute values can be used, how it can be downloaded and incorporated into the SciGRID\_gas model, and identify the copyright restrictions on the data source.

Due to copyright regulations, there are roughly two groups of data:

- Non-copyright restrictive data (N-CRRD): Here the copyright does not restrict the download, use and distribution of the data.
- Copyright restrictive data (CRRD): Here the data can be downloaded and used internally, but not re-distributed to others.

The following is a list of the data sources that will be used throughout the project and an identification into which group of copyright restriction they fall:

- **OSM** (<https://www.openstreetmap.org>) (N-CRRD)
- **GB** (<https://www.nationalgridgas.com/land-and-assets/network-route-maps>) (CRRD)
- **NO** (<https://www.npd.no/en/about-us/information-services/available-data/map-services/>) (N-CRRD)
- **LKD** (<https://tu-dresden.de/bu/wirtschaft/ee2/forschung/projekte/lkd-eu>) (N-CRRD)
- **ENTSOG** (<https://transparency.entsog.eu/>) (CRRD)

---

<sup>1</sup> An entire gas network data set is only available from the UK, see <https://www.nationalgridgas.com/land-and-assets/network-route-maps>.

- **EMAP** ([https://www.entsog.eu/sites/default/files/2020-01/ENTSOG\\_CAP\\_2019\\_A0\\_1189x841\\_FULL\\_401.pdf](https://www.entsog.eu/sites/default/files/2020-01/ENTSOG_CAP_2019_A0_1189x841_FULL_401.pdf)) (CRRD)
- **GIE** (<https://www.gie.eu/>) (N-CRRD)
- **GSE** (<https://www.gie.eu/index.php/gie-publications/databases/storage-database>) (N-CRRD)
- **IGU** (<https://www.igu.org/>) (CRRD)
- **INET** (see [DPM20e]) (N-CRRD)
- **CONS** (see [San21]) (N-CRRD).

Each data set and source comes with different copyright regulations. The copyright can be rather non-restrictive (e.g. INET) or can be restrictive (IGU). It is attempted to use only freely available data, so that such data can be re-distributed. In more restrictive data cases (IGU, GB), it is not allowed to download the data and distribute it to others. However, it is allowed to let other potential users know of the location of such data and supply them with tools that allow them to carry out the same data download and subsequent incorporation of the data into a gas network data set.

**Note:**

In case that other users are aware of other data sources that might be useful to this project, please get in touch and supply us with a brief description of the data and the location of such data, so that additional tools can be developed to incorporate the data in this project. Please use the following email address: [developers.gas\(at\)scigrid.de](mailto:developers.gas@scigrid.de)

## 3.2 Gas consumers data set

**European gas consumers (CONS)** is a further dataset for the SciGRID\_gas project which was generated by extracting information from another non-OSM data source, and combining it with spatial information and distribution models. This work was carried out by Javier Sandoval as part of his Master’s Thesis [San21]. The raw data generated is a time series of daily gas consumptions on a NUTS-3 level ([Wik21]) for all member states of the EU. A very detailed description of the data generation process can be found in [San21]. Here a brief summary of how the data was generated will be given, followed by its limitations and the extend of the data (data density).

### 3.2.1 Input data sets for the CONS data set

Several different data sets and information were used to generate the gas demand data, such as the spatial framework, gas consumed, number of household and more. Each of those input data sets will be listed and briefly described below.

#### Spatial disaggregation

Here for the spatial framework NUTS region were used, which is a geographical system dividing the European Union into regions of similar size in respect of similar number of inhabitants. Information associated with NUTS regions can be used for socio-economic analysing, policy analysis in respect of farming, and collecting, developing and harmonizing the European regional statistics. NUTS-0 is the coarsest division, describing the national level, meaning each country is a single NUTS-0 typology. NUTS-1 is the next coarsest division, with only 103 regions, NUTS-2 consists of 286 regions, whereas NUTS-3 consists of 1364 regions in the EU (including Norway which uses a similar classification and the former EU member UK, but excluding Malta and Cyprus). For the SciGRID\_gas project the spatial information was given through shape files, such as:

- “NUTS\_RG\_01M\_2016\_4326\_LEVL\_1.shp”
- “NUTS\_RG\_01M\_2016\_4326\_LEVL\_2.shp”
- “NUTS\_RG\_01M\_2016\_4326\_LEVL\_3.shp”.

In the SciGRID\_gas project the gas consumption was modelled on the NUTS-3 level. For countries (e.g. Ukraine) which are not part of the EU and are not described by the NUTS-3, no disaggregation smaller than the national level could be carried out.

#### Consumer type disaggregation

Next to the spatial disaggregation, a consumer type disaggregation was carried out. This information was available from the European Statistics department [Eur21b]. This data source supplied annual gas consumption on a national scale, broken up into the three sectors *Residential*, *Commercial* and *Industrial*. In addition, the European Statistical Office (EuroStat) also made available disaggregated final energy consumption of household and several industry consumers [Eur21c], which was a further input to the process. The three different sectors are defined as follow:

- *Residential*: Gas used by households due to space heating, cooking and hot water generation
- *Industrial*: Gas consumed by industrial facilities
- *Commercial, trade and services*: Gas consumed by smaller commercial, trade and service businesses, such as restaurants, workshops and other service providers (CTS).

## Further independent variables

Previously it had been demonstrated by the DemandRegio project [GGB+20] that the gas demand depends on independent variables, such as temperature (household space heating uses gas in many countries), and GDP (industry production is related to gas consumption). Here a list of all further input data sets is given:

- Temperature: Average daily air temperature at 2 m above ground, for the years 2010 to 2019, on a spatial resolution of 0.25 degrees through ERA5 [CCCS21]. Here a method was implemented, that generated a single daily value for each NUTS-3 region. This single daily value was used in determining “heating degree days”, which was the variable used.
- Number of households: Annual number of household by degree of urbanization on a NUTS-2 level [Eur21f], which was disaggregated onto a NUTS-3 level during the simulation process using population information.
- Population: Annual number of people living in NUTS-3 regions [Eur21g].
- Average dwelling area: Average size in units of square meters of dwelling by household type and degree of urbanization on a NUTS-0 level for the year 2012 only [Eur21a].
- GDP: Gross domestic product (GDP) at current market prices by NUTS-3 regions [Eur21e] on an annual basis for the years 2010 to 2019.
- Number of employees: Annual number of employees on a NUTS-2 level for the years 2010 to 2019 [Eur21d], which was disaggregated onto a NUTS-3 level during the simulation process using GDP information.
- Energy consumption: National annual gas consumption per the three sectors [Eur21b].
- Industry sector energy consumption: Information on different type of industry was available [Eur21b], which was linked to the number of employees by type of industry [Eur21h], resulting in larger gas consumption for regions with energy-intensive activities.
- National gas cooking: National values of cooking with gas [Eur21c] next to other energy sources
- National gas water heating: National values of water heating with gas [Eur21c] next to other energy sources
- National gas space heating: National values of space heating with gas [Eur21c] next to other energy sources.

Below for each sector and subsector the independent input variables are listed:

- *Residential:*
  - Cooking:
    - \* National gas cooking
    - \* Number of households.
  - Space Heating:
    - \* National gas space heating
    - \* Average dwelling area
    - \* Number of households
    - \* Temperature.
  - Water heating:
    - \* National gas water heating
    - \* Population.
- *Industrial:*
  - Energy Consumption: for sector industry

- Number of employees in industry
- Industry sector energy consumption
- GDP.
- *CTS*:
  - Energy Consumption: for sector CTS
  - Number of employees in CTS
  - Temperature
  - GDP.

In addition, Sandoval incorporated a weekday dependent gas consumption for the sectors *Industry* and *CTS* (see [San21]).

With the above information Sandoval was able to estimate a gas consumption for all three main sectors on daily time steps for the years 2010 to 2019 (incl.). In addition a similar data set was available for Germany through the DemandRegio project ([GGB+20]) for the year 2015. Sandoval was able to calibrate his results with the DemandRegio gas consumption and derive fitting parameters, which were applied to all European countries, resulting in a daily time series of gas consumption for all NUTS-3 regions. The resulting gas demand time series can be accessed through the SciGRID\_gas project.

### 3.2.2 Incorporating the CONS data into the SciGRID\_gas data frame

This above described gas demand data set will be part of the data repository on Zenodo. In addition, Sandoval generated gas demand time series for NUTS-2 and NUTS-1 level as well, which will also reside in the same repository. If any of these data sets need to be incorporated into the SciGRID\_gas project, then one needs to copy the data, in respect of the NUTS-levels, into one of the following folder: “Eingabe/Consumers/Nuts\_3”, “Eingabe/Consumers/Nuts\_2” and “Eingabe/Consumers/Nuts\_1”.

However, due to the volume of the gas data demand sets size (NUTS-3 is about 140 MB of data), the time series data was reduced to the following attribute values and will be incorporated into the SciGRID\_gas network data set as described below:

- *min\_demand\_M\_m3\_per\_d*: The minimal value of gas consumed of the 10-year time series of the combined sectors (sum) data set.
- *max\_demand\_M\_m3\_per\_d*: The maximal value of gas consumed of the 10-year time series of the combined sectors (sum) data set.
- *mean\_demand\_M\_m3\_per\_d*: The mean value of gas consumed of the 10-year time series of the combined sectors (sum) data set.
- *median\_demand\_M\_m3\_per\_d*: The median value of gas consumed of the 10-year time series the combined sectors (sum) data set.
- *min\_demand\_household\_M\_m3\_per\_d*: The minimal value of gas consumed of the 10-year time series of the sector household only.
- *max\_demand\_household\_M\_m3\_per\_d*: The maximal value of gas consumed of the 10-year time series of the sector household only.
- *mean\_demand\_household\_M\_m3\_per\_d*: The mean value of gas consumed of the 10-year time series of the sector household only.
- *median\_demand\_household\_M\_m3\_per\_d*: The median value of gas consumed of the 10-year time series of the sector household only.

- *min\_demand\_industrial\_M\_m3\_per\_d*: The minimal value of gas consumed of the 10-year time series of the sector industry only.
- *max\_demand\_industrial\_M\_m3\_per\_d*: The maximal value of gas consumed of the 10-year time series of the sector industry only.
- *mean\_demand\_industrial\_M\_m3\_per\_d*: The mean value of gas consumed of the 10-year time series of the sector industry only.
- *median\_demand\_industrial\_M\_m3\_per\_d*: The median value of gas consumed of the 10-year time series of the sector industry only.
- *min\_demand\_commercial\_M\_m3\_per\_d*: The minimal value of gas consumed of the 10-year time series of the sector CTS only.
- *max\_demand\_commercial\_M\_m3\_per\_d*: The maximal value of gas consumed of the 10-year time series of the sector CTS only.
- *mean\_demand\_commercial\_M\_m3\_per\_d*: The mean value of gas consumed of the 10-year time series of the sector CTS only.
- *median\_demand\_commercial\_M\_m3\_per\_d*: The median value of gas consumed of the 10-year time series of the sector CTS only.

### Consumers elements

As described in the above subsection, a number of attribute values have been generated for each NUTS region. In addition, only the following two attributes are part of the “Consumers” component:

- *nuts\_id*: Where the attribute name will be followed by an underscore and a number indicating the nuts level, e.g. “nuts\_id\_3” indicating that this is a NUTS-3 level region.
- *exact*: Accuracy in respect of geo-referencing (see [Chapter 5.3](#)).

This data set contains NUTS-3 regions of the European Union and other non-EU countries. Hence for those 1506 elements of type *Consumers*, consisted of 1348 values for NUTS-3 regions, and another 158 non-NUTS regions, e.g. locations in Albania, Switzerland and Turkey. For countries such as Ukraine, Belarus, Moldavia and Russia no NUTS values are defined.

### Nodes elements

Overall, there are 1506 *Nodes* elements in the CONS data set. Each node contains information in respect of the standard attributes, such as “name” and “node\_id”, next to “elevation\_m” and the attribute “exact” describing the accuracy in respect of the geo-referencing. As this info was available for all nodes, the data density is 100% for these attributes.

## 3.2.3 Copyright and disclaimer for the INET data set

### Copyright



Open Access: This document and the CONS data set are licensed under a Creative Commons Attribution 4.0 International License, which permits the user to share, adapt, distribute and reproduce in any medium or format, as long as the

user gives appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

A list of the sources used for the generation of the CONS data set can be found in [San21].

## Disclaimer

The CONS data set is supplied on a best-effort basis only, using available information as documented gathered through the process described in [San21]. While every effort is made to make sure the information is accurate and up-to-date, we do not accept any liability for any direct, indirect, or consequential loss or damage of any nature—however caused—which may be sustained as a result of reliance upon such information.

### 3.2.4 Summary CONS data

The CONS data set supplies information on gas consumption on a NUTS-3 level for Europe. This gas demand was generated through an interplay of independent variable information such as GDP, temperature and others, in combination with heuristic models, resulting in a daily time series spanning the years 2010 to 2019. This time series information was reduced to single values of minimum, median, mean and maximum gas demand for each NUTS-3 region for the three different consumer sectors: household, industry and commercial. The time series data and the summary data are accessible through the Zenodo web page, where the latter has been incorporated into a SciGRID\_gas data set.

Below a table summarises the number of elements for each component:

Table 3.1: CONS component summary

Component Name	Count
<i>BorderPoints</i>	0
<i>Compressors</i>	0
<i>Consumers</i>	1506
<i>LNGs</i>	0
<i>Nodes</i>	1506
<i>PipeSegments</i>	0
<i>Productions</i>	0
<i>Storages</i>	0

In addition, the map in Figure 3.1 visualizes the data for Europe.

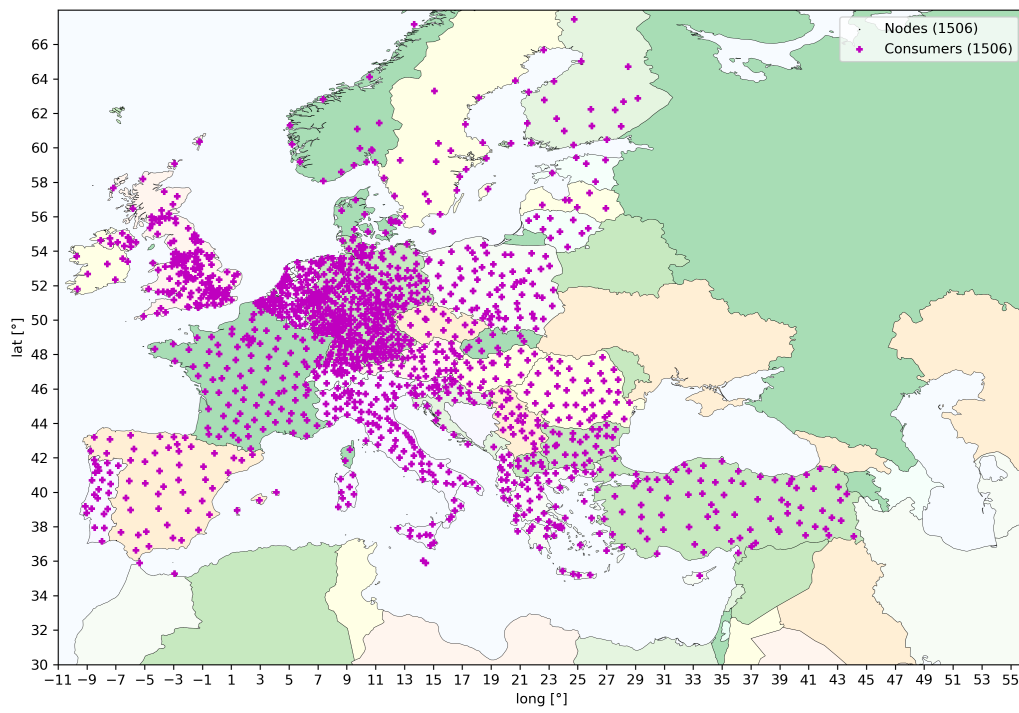


Figure 3.1: Overview map of the CONS data set for Europe.

### 3.3 Data summary

SciGRID\_gas is based on open source data. To generate a gas pipeline network data set, one needs to access different data sets that were found throughout the project and presented here. Emphasis was given to depict the number of elements per component and the data density for each data set.

### 3.4 Summary

Gas component data sets come in different forms, licenses, formats and detail. The SciGRID\_gas project can process such data and combine them to a consistent and reliable network data set.

The underlying gas component data sets were categorized into two different groups:

- OSM data: This is data originating from the OSM data base, containing well geo-referenced locations of gas facilities, such as pipe locations or gas storage facilities. However, it comes with very few meta information.
- Non-OSM data: These are all other data sources, which can “supply” detailed information on some of the gas facilities attributes. However, this information is sparse, as published only for a few facilities. Here, the INET data set was introduced as an example of the non-OSM data set, and the pathway of converting the raw data from the www into SciGRID\_gas project component structure.

Here detailed information on one or several data sources have been given, and should be used as a reference for later data processes.

## CONCLUSION

This document here is the documentation of one of the data sets that is part of the SciGRID\_gas project. This document here started off with the introduction of the SciGRID\_gas project, such as funding, duration and goals. In a subsequent chapter the data structure within the SciGRID\_gas project was described, such as components, elements, attributes and attribute values, so that the transmission data set could be an input to certain gas flow model. The third chapter introduced the CONS data set, which is a data set that was generated as part of the project containing information on the gas demand (estimated consumption) on a regional level (NUTS-3 regions in the EU).

This data set containing 1506 nodes, with 1348 consumer attributes, such as overall minimum or maximum daily gas consumption.



## 5.1 Glossary

Dataset abbreviations can be found in [Table 5.1](#).

Table 5.1: Dataset abbreviations

Name	Abbreviation	Description
Raw InternetDaten data set	INET	Label/name for the raw InternetDaten data set
Raw Gas Infrastructure Europe data set	GIE	Label/name for the raw Gas Infrastructure Europe data set
Raw Gas Storage Europe data set	GSE	Label/name of the raw Gas Storage Europe data set
Raw Norwegian data set	NO	Label/name for the raw Norwegian data set
Raw Long-term planning and short-term optimization data set	LKD	Label/name for the raw Long-term planning and short-term optimization data set
Raw International Gas Union data set	IGU	Label/name for the raw International Gas Union data set
Raw EntsoG-Map data set	EMAP	Label/name for the raw EntsoG-Map data set
Raw consumer data set	CONS	Label/name for the raw natural gas consumer data set
Merged and filled IGG data set	IGG	Filled data sets, for which the <b>INET</b> , <b>GIE</b> and <b>GSE</b> data sets were merged
Merged and filled IGGI data set	IGGI	Filled data sets, for which the <b>INET</b> , <b>GIE</b> , <b>GSE</b> and <b>IGU</b> data sets were merged
Merged and filled IGGIN data set	IGGIN	Filled data sets, for which the <b>INET</b> , <b>GIE</b> , <b>GSE</b> , <b>IGU</b> and the <b>NO</b> data sets were merged
Merged and filled IGGINL data set	IGGINL	Filled data sets, for which the <b>INET</b> , <b>GIE</b> , <b>GSE</b> , <b>IGU</b> , <b>NO</b> and the <b>LKD</b> data sets were merged
Merged and filled IGGIELGN data set	IGGIELGN	Filled data sets, for which the <b>INET</b> , <b>GIE</b> , <b>GSE</b> , <b>IGU</b> , <b>EMAP</b> , <b>LKD</b> , <b>GB</b> , and the <b>NO</b> data sets were merged

The glossary terms can be found in [Table 5.2](#).

Table 5.2: Glossary

Name	Abbreviation	Description
component		A gas network consists of different components, such as: pipelines, compressors, LNG terminals, storages, entry points and production sites
element		Elements are instances of components. Hence, “10 compressor elements” refers to a data set that contains information for 10 compressor stations
attribute		Gas facilities, such as pipelines or compressors, can be described with a large set of parameters, such as pipeline diameter, or compressor capacity. Those parameters are referred to as attributes
facility		General term used for a gas appliance, such as a single compressor station, or a single LNG terminal
PipeLine		A gas pipeline entity, which has one start and one end point, however, can run via many nodes
PipeSegment		A gas pipeline that has only one start and one end point, but no nodes in-between
LNG	LNG	Liquefied natural gas
CNG	CNG	Compressed natural gas
flow duration curve	FDC	It is the cumulative frequency curve that shows the percentage of time specified flow where equal or exceeded during a given period. The temporal information, when certain events occur, is lost
Energiewende		German term for the change in using primary energies, the move away from coal to renewable energies, such as wind or solar
gas component data set		Raw input data, associated with components of the gas transmission grid
gas network data set		Output data, a coherent network of gas transmission components
OSM	OSM	Data that is available from openstreetmap.org
non-OSM	Non-OSM	Data that is not part of the OSM data set
gas type		There are two types of gas: High (H) and Low (L) calorific gas
mean absolute error	MAE	mean difference between input values and estimated values
data density		The ratio of the number of usable (not missing) attribute values over number elements of the component, in units of [%]
Transmission System Operator	TSO	An entity entrusted with the transportation of natural gas/electricity, as defined by the European Union
gas transmission network		This describes the physical gas transmission grid, however, it excludes any facilities/components that would be part of a distribution network and their facilities
gas component data set		The term “gas component data set” is used for raw data sets of gas network facilities. However, not all elements (e.g. compressors) need to be connected to pipelines, where the emphasis is on the term <b>component</b>
gas network data set		A “gas component data set” can be converted into a “gas network data set”, by connecting all non-pipeline elements to nodes and all nodes are connected to pipelines. Hence, the emphasis here is on the term <b>network</b>
Nomenclature des unités territoriales statistiques	NUTS	Geographical system dividing the European Union into regions of similar size in respect of number of inhabitants

## 5.2 Unit conversions

Table 5.3: Unit conversions

From Unit	To Unit	MultiVal
LNG Mt	LNG Mm <sup>3</sup>	2.47
gas tm <sup>3</sup> h <sup>-1</sup>	gas Mm <sup>3</sup> d <sup>-1</sup>	24/1000
LNG Mm <sup>3</sup>	gas Mm <sup>3</sup>	584
LNG t	gas Mm <sup>3</sup>	1442.48
GWh (H)	gas Mm <sup>3</sup>	0.0879757777
GWh (L)	gas Mm <sup>3</sup>	0.1023541453

For some elements of some components, the calorific value was given through the references. Hence during the conversion process from GWh to M m<sup>3</sup>, the elements calorific value was used, however, wherever the element specific calorific value was not known, the default values from Table 5.3 was used in dependence of the gas type of the element. If no gas type was known, then high calorific gas is assumed.

## 5.3 Attribute *exact*

Each element of type *Nodes* has an attribute *exact*. With this, the SciGRID\_gas project is trying to let the user know, how well the actual location of the *Nodes* elements are known. The actual location (latitude-longitude pair) can be spot on (verifiable through satellite imagery) or can be unknown by 10's or 100's of km, where city names or country names are known only. Here the attribute value for *exact* is being given, ranging from "1" to "5" as listed in Table 5.4 below.

Table 5.4: Unit conversions

Exact value	Description	Uncertainty [km]
1	The exact location of this node is known, as one was able to verify the facility through satellite data.	0
2	Here the lat/long is not known exactly. However, one assumes that the location is within a small region (e.g. Krummhörn). Hence, not being much larger than 10 km	10
3	Here so little is known about the exact location, and one only knows that the location is within a large region (e.g. Hamburg). Hence, the actual location could be out by 10 km or more but less than 100 km	100
4	Here so little is known about the exact location, and one only knows that the location is within a state (e.g. Niedersachsen). Hence, the actual location could be out by 100 km or more but less than 1000 km	1000
5	Here so little is known about the exact location, and one only knows that the location is within a country (e.g. Ukraine). Hence, the actual location could be out by 1000 km or more.	> 1000

## 5.4 Location name alterations

Location names should be changed into the 26 letters used in the English language.

For names from the individual countries please follow the suggested approach:

- Germany/Austria: *Umlaute* to be replaced with the letter followed by an ‘e’, e.g.: ü = ue.
- France/Belgium: Omit accent de gues and accent de graphs, e.g.: ó = o.
- Sweden: Please change the last three letters of the Swedish alphabet and replace e.g.: ä = a.
- Poland: Please change any letter that cannot be found in the English alphabet, knowing that for some letters that one can only use a single letter instead of the three different letters used in the Polish alphabet, e.g.: z = z.
- Spain/Portugal: Please change any letter that cannot be found in the English alphabet, e.g.: ñ = n.
- Greece: Please do not use Greek letters. Please try to write the Greek words with Latin letters.
- Denmark: Please change any letter that contains non-English letters, e.g.: “å” with ”aa”.
- Slovakia, Czech Republic, Hungary, Rumania, Latvia, Lithuania, Estonia, Bulgaria, Slovenia, Croatia: PLEASE use your common sense, based on the examples from the other countries above.

## 5.5 Country name abbreviations

For convenience we provide a short list of names and two-digit codes (see Table 5.5) for the probably most important countries associated with the European Transmission Grid.

Table 5.5: Country codes

Country name	Country code	Country name	Country code
Albania	AL	Kosovo	XK
Armenia	AM	Latvia	LV
Austria	AT	Liechtenstein	LI
Azerbaijan	AZ	Lithuania	LT
Belarus	BY	Luxembourg	LU
Belgium	BE	Malta	MT
Bosnia and Herzegovina	BA	Moldova	MD
Bulgaria	BG	Montenegro	ME
Croatia	HR	Netherlands	NL
Cyprus	CY	Norway	NO
Czech	CZ	Poland	PL
Denmark	DK	Portugal	PT
Estonia	EE	Romania	RO
Finland	FI	Serbia	RS
France	FR	Slovakia	SK
Georgia	GE	Slovenia	SI
Germany	DE	Spain	ES
Greece	GR	Sweden	SE
Hungary	HU	Switzerland	CH
Iceland	IS	Turkey	TR
Ireland and Northern Ireland	IE	Belarus	UA
Italy	IT	Great Britain	GB
Russia Federation	RU	Europe	EU
Ukraine	UA		

## 5.6 Acknowledgement

We acknowledge the contribution of Dr. Ontje Luensdorf from the German Aerospace Center (DLR), Institute for Networked Energy Systems to the SciGRID\_gas project.



## BIBLIOGRAPHY

- [AFW14] M. Ahmed, B.T. Fasy, and C. Wenk. *New Techniques in Road Network Comparison*. Penguin Random House, New York, NY, 2014.
- [AG99] H. Alt and L. Guibas. *Discrete geometric shapes: matching, interpolation, and approximation-a survey*. Sack JR, Urrutia J, Handbook of Computational Geometry, Elsevier, New York, NY, 1999.
- [CCCS21] 14 06 2018. [Online] Copernicus Climate Change Service. Era5 hourly data on single levels from 1979 to present. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>, 2021. Accessed: 2021-01-15.
- [DPDi20] J.C. Diettrich, A. Pluta, J. Dasenbrock, and W. i. *SciGRID\_gas: The combined IGGIN gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, 2020. URL: <https://zenodo.org/record/4288459#.YG7aFj9CSUk>, doi:10.5281/zenodo.4288458.
- [DPi20] J.C. Diettrich, A. Pluta, and W. i. *SciGRID\_gas: The INET gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.4008975>, doi:10.5281/zenodo.4008975.
- [DPM20a] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID\_gas: The combined IGG gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.4009129>, doi:10.5281/zenodo.4009129.
- [DPM20b] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID\_gas: The combined IGGI gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://zenodo.org/record/4288468#.YG7aBz9CSUk>, doi:10.5281/zenodo.4288467.
- [DPM20c] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID\_gas: The combined IGGIELGN gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, 2020. URL: <https://zenodo.org/record/4642569#.YG7ZhT9CSUk>, doi:10.5281/zenodo.4642568.
- [DPM20d] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID\_gas: The combined IGGINL gas transmission network data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, 2020. URL: <https://zenodo.org/record/4288440#.YG7aHz9CSUk>, doi:10.5281/zenodo.4288439.
- [DPM20e] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID\_gas: The raw INET data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.3985249>, doi:10.5281/zenodo.3985249.
- [DPM20f] J.C. Diettrich, A. Pluta, and W. Medjroubi. *SciGRID\_gas: The raw LKD data set*. DLR-Institut für Vernetzte Energiesysteme e.V., Oldenburg, Germany, Aug 2020. URL: <https://doi.org/10.5281/zenodo.3985271>, doi:10.5281/zenodo.3985271.
- [Ent17] EntsoG. North West GRIP, Min Report. [https://www.entsog.eu/sites/default/files/files-old-website/publications/GRIPs/2017/entsog\\_GRIP\\_NW\\_2017\\_main\\_xs.pdf](https://www.entsog.eu/sites/default/files/files-old-website/publications/GRIPs/2017/entsog_GRIP_NW_2017_main_xs.pdf), 2017. Accessed: 2021-03-17.

- [Eur21a] EuroStat. Average size of dwelling by household type and degree of urbanisation. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc\\_hcmh02&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_hcmh02&lang=en), 2021. Accessed: 2021-02-08.
- [Eur21b] EuroStat. Complete energy balances. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg\\_bal\\_c&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_bal_c&lang=en), 2021. Accessed: 2021-02-08.
- [Eur21c] EuroStat. Disaggregated final energy consumption in households - quantities. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg\\_d\\_hhq&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nrg_d_hhq&lang=en), 2021. Accessed: 2021-02-07.
- [Eur21d] EuroStat. Employment by age, economic activity and nuts 2 regions (nace rev. 2). [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst\\_r\\_lfe2en2&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_r_lfe2en2&lang=en), 2021. Accessed: 2021-02-08.
- [Eur21e] EuroStat. Gross domestic product (gdp) at current market prices by nuts 3 regions. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama\\_10r\\_3gdp&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_10r_3gdp&lang=en), 2021. Accessed: 2021-03-31.
- [Eur21f] EuroStat. Number of households by degree of urbanisation and nuts 2 regions. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst\\_r\\_lfsd2hh&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_r_lfsd2hh&lang=en), 2021. Accessed: 2021-02-08.
- [Eur21g] EuroStat. Population on 1 january by age group, sex and nuts 3 region. [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo\\_r\\_pjangrp3&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_r_pjangrp3&lang=en), 2021. Accessed: 2021-01-10.
- [Eur21h] EuroStat. Sbs data by nuts 2 regions and nace rev. 2 (from 2008 onwards). [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs\\_r\\_nuts06\\_r2&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs_r_nuts06_r2&lang=en), 2021. Accessed: 2021-03-16.
- [Gas19] GasTerra. Aardgas in Nederland. <https://www.gasterra.nl/uploads/fckconnector/20d5b75a-2ad6-529e-b024-1f5bd43bbd5e>, 2019. Accessed: 2021-03-17.
- [GGB+20] F. Gotzens, B. Gillessen, S. Burges, W. Hennings, J. Müller-Kirchenbauer, S Seim, P. Verwiebe, T. Schmid, F. Jetter, and T. Limme. *Harmonization and development of methods for a spatial and temporal resolution of energy demands (DemandRegio)*. Forschungsstelle für Energiewirtschaft e.V., 2020. Accessed: 2021-05-18.
- [Hel18] D. Helle. OpenStreetMap - Deutschland. <https://www.openstreetmap.de/>, 2018. Accessed: 2019-12-12.
- [Kha13] Y. Khalid. What is Pickle in python? <https://pythontips.com/2013/08/02/what-is-pickle-in-python/>, 2013. Accessed: 2019-10-10.
- [KingSpalding18] King&Spalding. LNG in Europe 2018: An Overview of LNG Import Terminals in Europe. <https://globalnghub.com/wp-content/uploads/2018/09/King.pdf>, 2018. Accessed: 2018-09-01.
- [KKS+17] F. Kunz, M. Kendzioriski, W.-P. Schill, J. Weibezahn, J. Zepter, C. von Hirschhausen, and P. Hauser. *Electricity, Heat, and Gas Sector Data for Modeling the German System*. Deutsches Institut für Wirtschaftsforschung, Daten Dokumentation 92, Berlin, 2017.
- [LSS+19] P. Lustenberger, F. Schumacher, M. Spada, P. Burgherr, and B. Stojadinovic. Assessing the performance of the european natural gas network for selected supply disruption scenarios using open-source information. *Energies*, 12(4685):1–28, 2019. doi:{10.3390/en12244685}.
- [MMK16] C. Matke, W. Medjroubi, and D. Kleinhans. SciGRID - An Open Source Reference Model for the European Transmission Network (v0.2). <https://power.scigrd.de>, 2016. Accessed: 2019-09-09.
- [Mic20] Microsoft. Bing maps api. <https://www.microsoft.com/en-us/maps/licensing>, 2020. Accessed: 2018 to 2021.
- [Nis20] A. Nisbet. Open topo data api. <https://www.opentopodata.org>, 2020. Accessed: 2020 to 2021.

- [San21] Javier Enrique Sandoval. Estimation and simulation of a gas demand time series for the european nuts 3 regions. Master's thesis, Carl von Ossietzky Universität Oldenburg, Germany, Fak. 5, Institute of Physics (PPRE) D-26111 Oldenburg, Germany, 6 2021. Supervised by Prof. Dr. Carsten Agert, Dr. Herena Torio and Dr. Wided Medjroubi.
- [San19] B. Sandvik. World Borders. [http://thematicmapping.org/downloads/world\\_borders.php](http://thematicmapping.org/downloads/world_borders.php), 2019. Accessed: 2019-07-07.
- [SAB+17] M. Schmidt, D. Aßmann, R. Burlacu, J. Humpola, I. Joormann, N. Kanelakis, T. Koch, D. Oucherif, M.E. Pfetsch, L. Schewe, R. Schwarz, and M. Sirvent. *GasLib—A Library of Gas Network Instances*. 2017. doi:{10.3390/data2040040}.
- [sl19] scikit-learn. 1.1. Linear Models (scikit learn). [https://scikit-learn.org/stable/modules/linear\\_model.html](https://scikit-learn.org/stable/modules/linear_model.html), 2019. Accessed: 2019-08-08.
- [UoO14] USA University of Oregon. Comparing distributions: Z Test. <http://homework.uoregon.edu/pub/class/es202/ztest.html>, 2014. Accessed: 2020-07-07.
- [Wik20a] Wikipedia. Bootstrapping (statistics). [https://en.wikipedia.org/wiki/Bootstrapping\\_\(statistics\)](https://en.wikipedia.org/wiki/Bootstrapping_(statistics)), 2020. Accessed: 2019-06-06.
- [Wik20b] Wikipedia. Cross-validation (statistics). [https://en.wikipedia.org/wiki/Cross-validation\\_\(statistics\)#Exhaustive\\_cross-validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)#Exhaustive_cross-validation), 2020. Accessed: 2019-07-07.
- [Wik20c] Wikipedia. Jackknife resampling. [https://en.wikipedia.org/wiki/Jackknife\\_resampling](https://en.wikipedia.org/wiki/Jackknife_resampling), 2020. Accessed: 2019-08-08.
- [Wik20d] Wikipedia. Lasso (statistics). [https://en.wikipedia.org/wiki/Lasso\\_\(statistics\)](https://en.wikipedia.org/wiki/Lasso_(statistics)), 2020. Accessed: 2020-04-04.
- [Wik20e] Wikipedia. Limited-memory BFGS. [https://en.wikipedia.org/wiki/Limited-memory\\_BFGS](https://en.wikipedia.org/wiki/Limited-memory_BFGS), 2020. Accessed: 2020-06-06.
- [Wik20f] Wikipedia. Out-of-bag error. [https://en.wikipedia.org/wiki/Out-of-bag\\_error](https://en.wikipedia.org/wiki/Out-of-bag_error), 2020. Accessed: 2019-07-07.
- [Wik20g] Wikipedia. Transmission system operator. [https://en.wikipedia.org/wiki/Transmission\\_system\\_operator/](https://en.wikipedia.org/wiki/Transmission_system_operator/), 2020. Accessed: 2019-09-09.
- [Wik20h] Wikipedia. JAGAL. <https://en.wikipedia.org/wiki/JAGAL>, 2020. Accessed: 2020-01-01.
- [Wik21] Wikipedia. Nomenclature of territorial units for statistics. [https://en.wikipedia.org/wiki/Nomenclature\\_of\\_Territorial\\_Units\\_for\\_Statistics](https://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics), 2021. Accessed: 2021-04-10.
- [BMW11] BMWi. Forschung für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung. [https://www.bmw.de/Redaktion/DE/Publikationen/Energie/6-energieforschungsprogramm-der-bundesregierung.pdf?\\_\\_blob=publicationFile&v=12](https://www.bmw.de/Redaktion/DE/Publikationen/Energie/6-energieforschungsprogramm-der-bundesregierung.pdf?__blob=publicationFile&v=12), 2011. Accessed: 2019-02-02.
- [BMW20] BMWi. Home page of BMWi. <https://www.bmw.de/Navigation/DE/Home/home.html>, 2020. Accessed: 2020-03-03.
- [BundesregierungDeutschland20] Bundesregierung Deutschland. Home page of Bundesregierung Deutschland. [https://www.bundesregierung.de/Webs/Breg/DE/Themen/Energiewende/\\_node.html](https://www.bundesregierung.de/Webs/Breg/DE/Themen/Energiewende/_node.html), 2020. Accessed: 2020-01-01.
- [EntsoG20] EntsoG. Home page of EntsoG. <https://www.entsog.eu/>, 2020. Accessed: 2020-03-03.
- [GasIEurope20] Gas Infrastructure Europe. Home page of Gas Infrastructure Europe. <https://agsi.gie.eu>, 2020. Accessed: 2020-01-01.
- [GasSEurope20] Gas Storages Europe. Home page of Gas Storages Europe. <https://www.gie.eu/index.php/transparency/gse-transparency-template>, 2020. Accessed: 2020-01-01.

- [Gassco20a] Gassco. Data page of facilities from Gassco. <https://www.npd.no/en/about-us/information-services/available-data/map-services/>, 2020. Accessed: 2020-01-01.
- [Gassco20b] Gassco. Home page of Gassco Norway. <https://www.gassco.no/en/>, 2020. Accessed: 2020-01-01.
- [IGU20] IGU. Home page of International Gas Union. <https://www.igu.org/>, 2020. Accessed: 2018-10-01.
- [nationalGrid20] nationalGrid. Home page of National Grid UK. <https://www.nationalgrid.com/uk/>, 2020. Accessed: 2018-10-01.