SciGRID_gas: The raw GIE data set

Release 1.0

J.C. Diettrich & A. Pluta & W. Medjroubi

Jul 17, 2020

CONTENTS
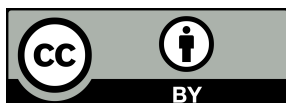
Impressum

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

LIST OF FIGURES

LIST OF TABLES

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, 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 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. In addition heuristic tools are required to fill data gaps, so that a complete gas network data set can be generated.

Here, this document describes one of the non-OSM data set, called the "GIE" data set, which originated from "Gas Infrastructure Europe". 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 GIE 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 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 [W16] 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 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 - December 2020

- **Funding agency**: Federal Ministry for Economic Affairs and Energy (*Bundesministerium für Wirtschaft und Energie*), Germany

- **Funding code**: Funding Code: 03ET4063

- **Project partner**: DLR Institute for Networked Energy Systems

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 difficult, 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 operators, 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 commercially 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 research centre DLR Institute of Networked Energy Systems in Oldenburg aims at building 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 are published under the GPLv3 license.

The main purpose of the SciGRID_gas project is therefore to open the door to new gas transmission network models and innovative ideas in energy system modelling by providing freely available and well-documented data on the European gas transmission network.

The input data itself is based on data available from 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

- remove all gas facilities, that are not connected to pipelines.

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 like pipelines, compressors, LNG terminals, and their attributes, such as diameters, capacities, etc.

Here a single data source will be introduced, the **GIE** data set (GIE = Gas Infrastructure Europe). At later stage, descriptions of other data sources will follow, and will be made available on the project webpage.

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

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 GIE 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.

Prior to the description of the data structure, the overall pathway of the data flow within the SciGRID_gas project will be explained, as it is believed, that such overview will help the reader.

2.1 Data structure description

This section contains information about 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 summarized as a picture in Figure 2.1.
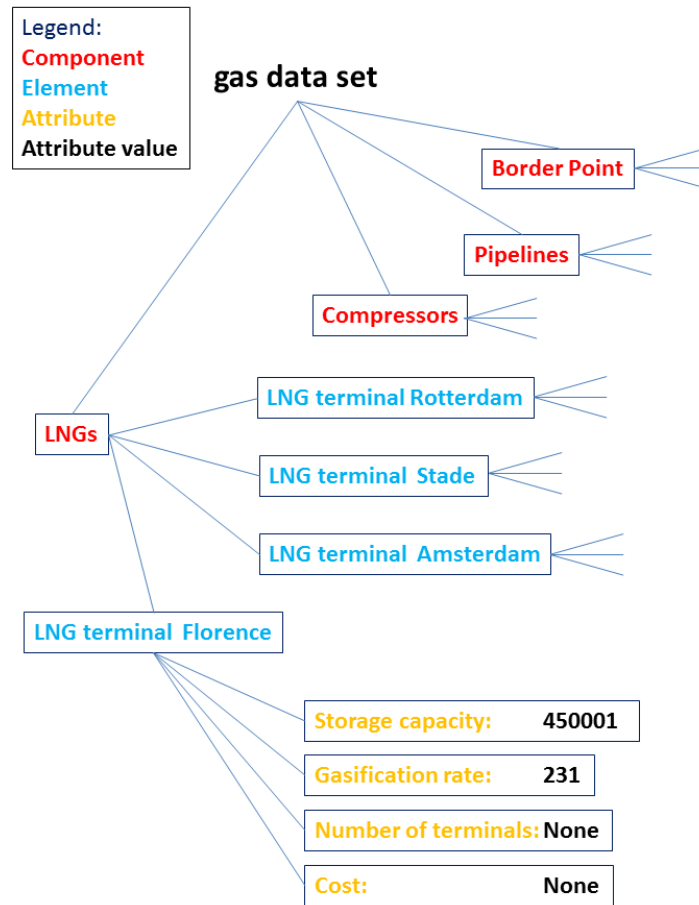
Figure 2.1: Data structure for the SciGRID_gas data set

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].

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 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", and the emphasis is on the term **component**.

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. Hence, the emphasis here is on the term **network**.

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 labels of the values that are associated with the elements. Examples for 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 boolean (*True*/*False*) is also allowed, more likely a "1" will stand for *True* and "0" for *False*. However, some attribute values might not be given in the data source, therefore a no value attribute value does exist. In the Python code it is termed *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, where each component contains several elements, and each element has several attributes, which each come with a value, where "None" stands for unknown value. The heuristic processes described in this document at a later stage will fill those "None" values with generated values.

Gas component types

A gas transmission network consists of different components, such as pipelines, compressors, etc. For the Sci-GRID_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* are one of the main components of the gas pipeline network. *PipeLines* allow for the transmission of the gas from one node to another. However, each pipe is unique. They might have different diameter, capacity or max pressure. In addition, a single *PipeLine* can connect several nodes. Therefore it could go from "Radeland" to "Bottrub" and then follow on to "Frankfurt". However, *PipeLines* do not need to connect more than 2 nodes, but can. The order of those nodes is important, and indicates the flow direction.

- *PipeSegments*: *PipeSegments* are almost identical to *PipeLines*, However, are only allowed to connect two nodes. Thus they have one start node and one end node, and are not passing via other nodes or other component elements in between, such as compressors or LNG terminals. Hence, any pipeline can easily be converted to multiple pipe-segments.

- *Compressors*: *Compressor* represent compressor stations, which are important. Gas travelling through the gas pipeline loses pressure due to friction on the pipeline walls and other factors. This will reduce the throughput of the gas amount. Hence, every so often (~ every 150 km), a compressor station is required, which increases the pressure of the gas, and Hence, allows the gas to flow through the gas pipeline. A gas compressor station contains several gas compressors units (turbines). Knowing the individual gas turbines is of an advantage, as those turbines can be combined in different ways, such as in series, or parallel, or combinations of those two options.

- *LNGs*: (LNG terminals and LNG storages) Some of the gas, which is being used throughout Europe, is supplied via ships to LNG terminals and LNG storage facilities. (From here onwards the acronym "*LNGs*" will be used instead "LNG terminals and LNG storage facilities".) As the transmission of gas would be extremely inefficient due to its volume, the gas state is changed to the liquid form (LNG gas), and then shipped. Ships arriving in Europe need special LNG terminals that can store LNG gas and subsequently re-gasify it. The storage and re-gasification of the gas are combined in the *LNGs* component and need to be part of any gas network for Europe.

- *Storages*: Part of the gas network will be gas storages. Gas storages are being used as gas pipeline capacities or gas production capacities might not be able to cover high demand periods, such as during the winter. Hence, large gas storage units are being filled during the summer periods while the overall demand is low, and if capacities of net supply allow it. This gas is then used during the winter period, and can compensate for shortcomings of the gas network or gas supply. Almost every country has their own gas storage units, ranging from smaller units to compensate for daily fluctuations to larger units, which compensate seasonal fluctuations. For the SciGRID_gas model the larger seasonal storage units are of more importance than the smaller ones,

as we are interested in the transmission gas pipeline network. However, any gas storage can be added and implemented into the gas network data set.

- *Consumers*: Part of the gas pipeline network is the knowledge of gas demand. Gas is added to the network at LNG terminals and European boundary cross border points. One type of users is the gas power plants. These can be added to the SciGRID_gas model, as this will specify local gas demand. In addition other consumers, such as city gas providers and large industries can also be added to the network data set.

- *Production*: 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 through Europe.

- *BorderPoints*: BorderPoints are cross border points (between different countries), which are mostly for the purpose of accounting the gas flow. Most large gas pipelines have cross border stations, e.g. Ellund (lat/long: 54.80181, 9.289079) at the border between Germany and Denmark, with gas facilities on both sides of the border.

- *EntryPoints*: These are special border points, as they are at the borders of the European Union and will be the gas entry points for the SciGRID_gas model data set.

- *InterConnectionPoints*: These are points between gas transmission operators, and will be found mainly within Europe, in particular at country borders. However, they can also be found within a single country, if there is more than one gas transmission operator.

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. The overall structure of those elements is described in the following part:

- *id*: A string, that is the ID of the element, and must be unique.

- *name*: A string that is the name of the element, such as "Compressor Radeland".

- *source_id*: A list of strings that are the sources of the element. As several elements from different sources could have been combined in a single element, one might need to know which are the original ids of the original sources.

- *node_id*: This is the ID of a geo-referenced point 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, that starts at one point and finishes at a different point, this entry would be a list of at least two node_id values.

- *lat*: This is the latitude value of an element. For pipelines, lat is a list of latitude values if known. The geo-referenced projection of the element that is being used in the SciGRID_gas project is: World Geodetic system 1984 (epsg:4326).

- *long*: The longitude, analogue to lat.

- *country_code*: This is a string indicating the official 2-digit country code (Alpha-2 code, see Chapter 4.4 for list of countries and their code). It represents the location of the element. As pipelines can pass through more than one country, the country code for pipes is the list of country codes of the countries the pipeline is passing through.

- *comment*: This is an arbitrary comment that is associated with the element.

- *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. Throughout the SciGRID_gas project, they have been coded as "dictionaries". They are called:

- *param*

- *method*

- *uncertainty*.

The structure within each dictionary is the same, However, their meaning is different. First of all 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 , such as number of turbines, overall turbine power, energy source of turbine and more.

So 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".

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.

2.2 Summary

The SciGRID_gas software is designed to construct a gas transmission network data set form different 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*.

DATA SOURCES

Two thirds of the gas used in Europe is imported from non-EU states, and all gas required for the consumption needs to be distributed through the existing gas transmission pipelines in Europe. In the future gas consumption might rise, leading to additional pressure on the current infrastructure. In addition, gas facilities could play a vital role in reducing CO_2 emission, as excess electricity could be converted to gas, that could be stored and transmitted throughout Europe with the existing gas network. Hence, a reliable network data set for the European transmission network is essential. The data required for such models ranges from pipeline diameter, gas pressure within the pipeline, actual pipeline length, pipeline capacity, and underground storage volume to name just a few.

However, such data is the property of the transmission system operators (TSOs) and is therefore 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.

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 [D18]. 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.

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^3/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."

The data available 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.

One of the main challenges for SciGRID_gas is that, gas transmission data is incomplete and accumulated from different sources. Also such different sources can have different properties for one and the same facility. Hence, it is important to know, which data set supplies which information. Hence, this chapter here will introduce the relevant data sets (e.g. INET), starting off with the components, the elements for each component and then the attributes for each element.

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[1], 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 in the format as described in later chapter releases.

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 [GasIEurop20] 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 [FMWP+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 our 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.

---

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

The following is a list of the data sources that will be used throughout the project and an indication 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)

- **EMap** (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)

- **GasLib** (http://gaslib.zib.de/) (N-CRRD)

- **INET** (see `Refs_InternetData`) (N-CRRD).

Each data set and source comes with a different copyright regulation. 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

3.2 Gas Infrastructure Europe (GIE) data set

Gas Infrastructure Europe (**GIE**) is a data set with vital meta-data within the SciGRID_gas project, where GIE stands for Gas Infrastructure Europe.

'Gas Infrastructure Europe (GIE) is an association representing the sole interest of the infrastructure industry in the natural gas business, such as Transmission System Operators, Storage System Operators and LNG Terminal Operators. GIE has currently 68 members in 25 European countries.' (https://www.gie.eu/).

GIE is the umbrella organisation for the following three gas components:

- **Storage**: GSE - Gas Storage Europe representing the Storage System Operators (SSO)

- **LNG**: GLE - Gas LNG Europe representing the LNG Terminal Operators (TO)

- **Transmission**: GTE - Gas Transmission Europe representing the Transmission System Operators (TSO)

The storage and the LNG information can be retrieved through an API supplied by GIE. However, there is no further information on the gas transmission part.

The APIs for the gas storage and the LNG terminals are:

- Aggregated LNG Storage Inventory (ALSI) (https://alsi.gie.eu/api/data/)

- AGSI+ AGGREGATED GAS STORAGE INVENTORY (https://agsi.gie.eu/api/data/)

Documentation for the APIs can be found on the web under: https://agsi.gie.eu/GIE_API_documentation_v001.pdf.

3.2.1 Pre requirements for accessing the GIE data set

A private key is required for the GIE data set so that one can download data from the GIE API.

As stated in the documentation for the GIE_API:

> The API service is available to the public free of charge. Registration on the AGSI+ or ALSI website is mandatory for non-data providers to be able to use the API. Registration will result in a personal API key that is required within the API URL. The only purpose of this registration is to enable us to assess and improve the performance of our systems where and if required (user count, user activity, most popular data set types). Your account information and settings can be updated (and cancelled) at any time after signing in. Your data will be stored and securely handles as long as your account remains active.

For this you will need to go to the following link: https://agsi.gie.eu/#/login where on the right hand side you will need to fill in the registration details.

Under "Access to:" please select "Both AGSI+ and ALSI"

After registration you will have access to your private key. Copy the key and past it into the following file:

/SciGRID_gas/Eingabe/GIE/GIE_PrivateKey.txt

This is your private key, Hence, do not share it with others.

3.2.2 Data manipulation of the GIE data set

Gas infrastructure providers are requested to publish certain gas flow information. This data is accessible via the GIE URLs, and contains a vast amount of meta-data for storages and LNG terminals throughout Europe. Whenever the data is downloaded from the GIE API, the data needs modification, so that it is conform to the SciGRID_gas data model. Several tools have been written to achieve this. The GIE specific tools are described below for **Storages** and **LNGs**.

Processes for retrieving the data from the GIE API

First of all, one could access some meta-data on the storages and LNG terminals through the following internet links:

- LNG: https://www.gie.eu/index.php/gie-publications/databases/lng-database

- Storages: https://www.gie.eu/index.php/gie-publications/databases/storage-database

These data sets come as Excel books and contain information, such as name of facility, country, type of facility, and eic_code. Other information, such as max hourly capacity or LNG storage capacity, is also given, however, discarded due to copyright reasons.

LNG terminals The following information are used from the LNG table above:

- country

- type

- eic_code

- short_name

- name

- nameShort

This information (column heading and column data) needs to be placed into a CSV document. This file needs to be named "GIE_LNG.csv" and needs to be stored in the folder "/SciGRID_gas/Eingabe/GIE/".

Other additional fields from this table are:

- Region

- Status

- Investment

- Start-up year

- Type

- Operator short name

- Max. Hourly Cap. $[\mathrm{m}^3(\mathrm{N})/\mathrm{h}]$

- Nom. Annual Cap. billion $[\mathrm{m}^3(\mathrm{N})/\mathrm{a}]$

- Possible additional Nom. Annual Cap. billion $[\mathrm{m}^3(\mathrm{N})/\mathrm{a}]$

- LNG storage capacity $[\mathrm{m}^3\mathrm{LNG}]$

- Number of tanks

- Max. ship class size receivable $[\mathrm{m}^3\mathrm{LNG}]$

- Number of jetties

- Min. sea depth alongside [m]

- Max. send out pressure [bar]

- TPA regime

- PCI list

- Operator long name

This list can change during the runtime of the project.

The EIC-code, facility code and country code is subsequently used to request time series information for each location from the GIE API.

The retrieved time series contain two useful values:

- the working LNG volume in the LNG storage tank

- the gas flow amount from the storage to the gas-pipeline, in units of GWh/d (see Table 3.1)

From the so created time series, one can determine the maximum working gas volume in the LNG storage tank in units of million LNG cubic meters. Also the maximum and medium gas flow from the storage to the gas pipeline is determined, in units of GWh per day.

Prior to estimating the maximum and median value from the retrieved time series, the time series was quality assured. This was done by removing any outliers/spikes.

Table 3.1: GIE incorporated attributes

Field identifier	Description	Units	Example
status	E (estimated) C (confirmed) N (no data)	E / C / N	C
gasDayStartedOn	The start of the gas day reported upon	YYYY-MMDD	2015-11-02
lngInventory	The aggregated amount of LNG in the LNG tanks at end of the previous gas day	$1000m^3$LNG (2 digits behind decimal point)	5373.25
sendOut	The aggregated gas flow out of the LNG facility within the gas day	GWh/d (1 digit behind decimal point)	976.5
dtmi	Declared Total Maximum Inventory	$1000m^3$ LNG (2 digits behind decimal point)	8898.99
dtrs	Declared Total Reference SendOut	GWh/d (1 digit behind decimal point)	6650.0
info	Service Announcement (if applicable)	url	https://alsi.gie.eu/#/news/184

The following information is incorporated into the SciGRID_gas data structure: name, max storage volume, max and medium gas flow volumes, facility code, country code, and EIC-code.

Subsequently, the LNG storage volume and flow are converted to their corresponding gas phase values. The final units of the measurements were $[\mathrm{Mm}^3\mathrm{d}^{-1}]$. No geo-coordinates are given for LNG terminals within the GIE data set. This information has been retrieved from the INET data set by a comparison of the name and the country code of the facility.

Storages

A meta-data set for the storages is available as Excel book and can be downloaded from the following URL: https://www.gie.eu/maps_data/downloads/2018/Storage_DB_Dec2018.xlsx.

In this Excel book, the sheet "Storage DB" contains the following columns:

- Country: string indicating the country of the storage

- Concatenate: –missing description–

- Country Code: two letter acronym for country code

- Company code: number of company

- Facility code: code of the facility

- Operator: name of operator

- Facility/Location: name of facility location

- Status: status of storage unit (operational/under construction/planned)

- Investment: string indicating the investment (existing/expansion/new facility)

- Start-up year: year when started operation

- Type: storage type, e.g. depleted field, salt cavern,…

- Notes:

- onshore/offshore: either onshore of offshore location of the gas storage

- Working gas (technical) TWh: maximum working gas volume in units of [TWh]

- Working gas TPA TWh: maximum working gas volume under TPA in units of [TWh]

- Working gas no TPA TWh: maximum working gas volume not under TPA in units of [TWh]

- Withdrawal technical GWh/day: maximum withdrawal rate of gas in units of [GWh/d]

- Withdrawal TPA GWh/day: maximum withdrawal rate of gas under TPA in units of [GWh/d]

- Withdrawal no TPA GWh/day: maximum withdrawal rate of gas not under TPA in units of [GWh/d]

- Injection technical GWh/day: maximum injection rate of gas in units of [GWh/d]

- Injection TPA GWh/day: maximum injection rate of gas under TPA in units of [GWh/d]

- Injection no TPA GWh/day: maximum injection rate of gas not under TPA in units of [GWh/d]

- Access regime: access regime with two options "nTPA", "rTPA", and "No TPA"

- in EU28 number: string ("n" or "y") if part of the 28 EU members

- in EU28 SUM: string ("n" or "y") if part of the 28 EU members

- EU 28 filter: string ("NO" or "YES") if part of the 28 EU members.

A subset of the above data needs to be saved as column data into a CSV file, containing only the following parameters with their data:

- Country

- Type

- EIC Code

- Short Name

- Name

- nameShort

This file needs to be name "GIE_Storages.csv" and needs to be stored in the folder "/SciGRID_gas/Eingabe/GIE/".

These is subsequently used to get access to the time series of the storage data set, by using the facility code, the country and the EIC-code.

All time series consist of the daily "working gas volume", the "daily injection capacity" and the "daily withdraw capacity". Maximum values for each of those parameters are extracted from those time series.

However, as was carried out for the LNG data set, the same testing of the goodness of the data was carried out.

In addition, gas flow values were converted from [GWh/d] to [Mm^3/d].

3.2.3 GIE data density

All GIE components (**Storages** and **LNGs**) have the following mandatory attributes:

- *id*: unique identifier

- *name*: name of the pipe-segment

- *source_id*: a source id

- *node_id*: the id of the start and the end node of the pipe-segment

- *lat*: a list of latitude values

- *longitude*: a list of longitude values

- *country_code*: a string pair indicating the country code of the start and the end point

- *comment*: a user comment.

LNGs

Overall, there are 21 LNG terminals in the GIE data set. In addition to the default attributes, the following non-standard attributes (see Table 3.2) are supplied. The number of attribute values supplied for each attribute is given by the parameter 'data density' (see Chapter 4.1):

Table 3.2: GIE *LNGs* data density

Attribute name	Description	Units	Data density [%]
eic_code	EIC code of LNG terminal		100
facility_code	unique facility code		100
max_cap_store2pipe_M_m3_per_d	maximum gas flow from storage to pipeline	Mm^3/d	100
max_workingGas_M_m3	maximum stored gas in LNG terminals	Mm^3	100
median_cap_store2pipe_M_m3_per_d	medium gas flow from storage to pipeline	Mm^3/d	100
name_short	short name of the facility		100

Storages

Overall, there are 109 storage facilities in the GIE data set. In addition to the default attributes, the following non-standard attributes (see Table 3.3) are supplied and partially populated with data:

Table 3.3: GIE *Storages* data density

Attribute name	Description	Units	Data density [%]
eic_code	EIC code of storage facility		100
facility_code	unique facility code		100
max_cap_pipe2store_M_m3_per_d	maximum gas flow from pipeline to storage	$\mathrm{Mm^3d^{-1}}$	100
max_cap_store2pipe_M_m3_per_d	maximum gas flow from storage to pipeline	$\mathrm{Mm^3d^{-1}}$	100
max_workingGas_M_m3	maximum working gas in storage	$\mathrm{Mm^3}$	100
name_short	short name of the facility		100

Nodes

Overall, there are 115 node points in the GIE data set. In addition to the default attributes, the following non-standard attributes (see Table 3.4) are supplied and partially populated with data:

Table 3.4: GIE *Nodes* data density

Attribute name	Description	Units	Data density [%]
exact	boolean indicating that storage is planed		100
eic_code	EIC code of storage facility		100
facility_code	unique facility code		100
name_short	short name of the facility		100
elevation_m	short name of the facility		100

Data availability and data usage

The API of the GIE (Gas Infrastructure Europe (ASBL)) web portal allows for the user to download time series on the daily gas amount (stored or available) for gas storages and LNG terminals. Here, we do not pass on the downloaded time series information, but only other retrieved constant metadata, such as maximum storage capacity or maximum re-gasification from LNG into the gases state, were derived using the time series data.

3.2.4 Copyright

The generally valid copyright regulations for databases apply.

Data disclaimer

In addition the data disclaimer is given as under: https://agsi.gie.eu/#/disclaimer:

> "All data is provided by the contributors on a voluntary basis and free of charge. The data provided by AGSI is for information purpose only. GSE is using reasonable efforts to invest in ensuring the correctness, completeness, and timeliness of the information provided herein. Data have been carefully checked, are updated at regular intervals and may be subject to changes, removal, or amendments without prior notice. GSE neither assumes any warranty or liability for the correctness and completeness of information/services and entries nor for the mode of presentation."

3.2.5 Summary GIE data

The GIE data set supplies information on gas infrastructure facilities all over Europe, such as gas storages and gas LNG terminals. Data for those facilities are accessible by the SciGRID software through special CSV data files that are downloaded from the GIE web page. The information in those facilities are automatically filtered and reshaped to the data structure of SciGRID_gas. Units are partially converted to align with the other project data. The facilities are further geo-reference by the SciGRID_gas software with the help of the INET data set.

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

Table 3.5: GIE component summary

Component Name	Count
BorderPionts	0
Compressors	0
ConnectionPoints	0
Consumers	0
EntryPoints	0
InterConnectionPoints	0
LNGs	21
Nodes	115
PipeSegments	0
Production	0
Storages	109

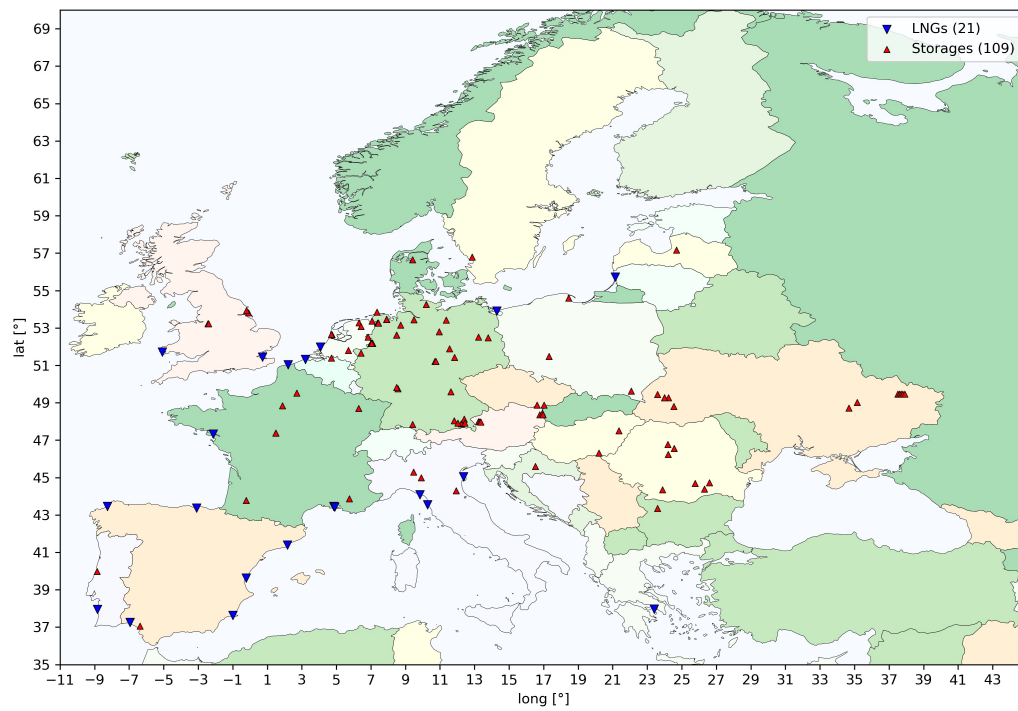In addition, the map in Figure 3.1 visualizes the data for Germany.

Figure 3.1: Overview map of the GIE 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.

FOUR

APPENDIX

4.1 Glossary

The glossary terms can be found in Table 4.1.

Table 4.1: Glossary

Name	Abbreviation	Description	Units
InternetDaten data set	INET	This is the label/name for the InternetDaten data set	
Gas Infrastructure Europe data set	GIE	This is the label/name for the Gas Infrastructure Europe data set	
Gas Storage Europe data set	GSE	This is the label/name for the Gas Storage Europe data set	
Long-term planning and short-term optimization data set	LKD	This is the label/name for the Long-term planning and short-term optimization data set	
International Gas Union data set	IGU	This is the label/name for the International Gas Union data set	
Norwegian data set	NO	This is the label/name for the Norwegian data set	
EntsoG-Map data set	EMAP	This is the label/name for the EntsoG-Map data set	
component		A gas network consists of different components, such as pipelines, compressors LNG terminals and more. However, for a gas transmission network, there is a handful of components only: pipeline, compressor, LNG terminal, storage, entry point, border point, connection point, consumer, node, and production	
element		Elements are instances of component. Hence, we speak of 10 compressor elements, if we have a data set that has 10 compressors. Here then we can refer to the first or the last or any element of such component	
attribute		Gas facilities, such as pipelines or compressors, can be described with a large number of parameters, such as pipeline diameter, or compressor capacity. Those parameters are referred to as attributes. Hence, each component has a list of properties, which are different from one component to another component	
facility		General term used for a gas appliance, such as compressor element, or LNG terminal	

Table 4.2: Glossary

Name	Abbreviation	Description	Units
PipeLine		This is a gas pipeline entity, which has one start and one end point, however can run via many nodes, compressors and other gas network elements	
PipeSegment		This is a gas pipeline, that has only one start and one end point, but no nodes in-between, Hence, only goes from one node to another node	
LNG	LNG	Liquefied natural gas	
CNG	CNG	Compressed natural gas	
flow duration curve	FDC	It is the cumulative frequency curve that shows the percent of time specified flow were equal or exceeded during a given period. The information on occurrence of events 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 the 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		This is the ratio of the number of usable (not missing) attribute values over number elements of the component	[%]
Transmission System Operators	TSO	This is 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 excludes any facilities/components that would be part of a distribution network and their facilities. This projects goal is to create an open source gas network data set that can be used to describe the European gas transmission network	
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. 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", and the emphasis is on the term **component**	
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 removed, resulting in a single network. Therefore the network contains nodes and edges which are connected, and all objects with the exception of pipelines are associated with nodes in this network, whereas pipelines are associated with edges. Hence, the emphasis here is on the term **network**	

4.2 Unit conversions

Table 4.3: Unit conversions

From Unit	To Unit	MultiVal
LNG Mt	LNG Mm^3	2.47
gas tm^3/h	gas Mm^3/d	24/1000
LNG Mm^3	gas Mm^3	584
LNG Mt	gas Mm^3	1442.48

4.3 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.

4.4 Country name abbreviations

For convenience we provide a short list of names and two-digit codes (see Table 4.4) for the probably most important countries associated with the European Transmission Grid.

Table 4.4: 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	EL	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

4.5 Acknowledgement

We acknowledge the contribution of Dr. Ontje Luensdorf from the DLR Institute of Networked Energy System to the SciGRID_gas project.

[B19]   Sandvik B. World Borders. http://thematicmapping.org/downloads/world_borders.php, 2019. Accessed: 2019-07-07.

[D18]   Helle D. OpenStreetMap - Deutschland. https://www.openstreetmap.de/, 2018. Accessed: 2019-12-12.

[FMWP+17] Kunz F., Kendziorski M., Schill W.-P., Weibezahn J., Zepter J., von Hirschhausen C., and Hauser P. *Electricity, Heat, and Gas Sector Data for Modeling the German System*. Deutsches Institut für Wirtschaftsforschung, Daten Documentation 92, Berlin, 2017.

[HL99]   Alt H. and Guibas L. *Discrete geometric shapes: matching, interpolation, and approximation-a survey*. Sack JR, Urrutia J, Handbook of Computational Geometry, Elsevier, New York, NY, 1999.

[JCJAW20a] Diettrich J.C., Dasenbrock J., Pluta A., and Medjroub W. SciGRID_gas: The filled INET data set. 2020. Accessed: 2020-07-08.

[JCJAW20b] Diettrich J.C., Dasenbrock J., Pluta A., and Medjroub W. SciGRID_gas: The raw INET data set. 2020. Accessed: 2020-07-08.

[MBTC14]   Ahmed M, Fasy B.T., and Wenk C. *New Techniques in Road Network Comparison*. Penguin Random House, New York, NY, 2014.

[MBTKSC15] Ahmed M., Fasy B.T., Hickmann K.S., and Wenk C. A path-based distance for street map comparison. *ACM Trans. Spatial Algorithms Syst.*, 1(1):28, 2015. doi:{10.1145/2729977}.

[MDR+17]   Schmidt M., Aßmann D., Burlacu R., Humpola J., Joormann I., Kanelakis N., Koch T., Oucherif D., Pfetsch M.E., Schewe L., Schwarz R., and Sirvent M. *GasLib—A Library of Gas Network Instances*. 2017. doi:{10.3390/data2040040}.

[RLP85]   Graham R.L. and Hell P. On the history of the minimum spanning tree problem. *Annauls of the History of Computer*, 7(1):43–57, 1985. doi:{10.1145/2729977}.

[sl19]   scikit-learn. 1.1. Linear Models (scikit learn). 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.

[W16]   Medjroubi W. SciGRID_power. https://power.scigrid.de, 2016. Accessed: 2019-09-09.

[Wik20a]   Wikipedia. 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, 2020. Accessed: 2019-07-07.

[Wik20c]   Wikipedia. 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), 2020. Accessed: 2020-04-04.

[Wik20e]   Wikipedia. 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, 2020. Accessed: 2019-07-07.

[Wik20g]   Wikipedia. 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.

[Y13]   Khalid Y. What is Pickle in python? https://pythontips.com/2013/08/02/what-is-pickle-in-python/, 2013. Accessed: 2019-10-10.

[BMWi11]   BMWi. Forschung für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung. https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/6-energieforschungsprogramm-der-bundesregierung.pdf?__blob=publicationFile&v=12, 2011. Accessed: 2019-02-02.

[BMWi20]   BMWi. Home page of BMWi. https://www.bmwi.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, 2020. Accessed: 2020-01-01.

[EntsoG20] EntsoG. Home page of EntsoG. https://www.entsog.eu/, 2020. Accessed: 2020-03-03.

[GasIEurop20] Gas Infrastructure Europ. Home page of Gas Infrastructure Europ. https://agsi.gie.eu, 2020. Accessed: 2020-01-01.