Research Centre on
ZERO EMISSION
NEIGHBOURHOODS
IN SMART CITIES

NTNU SINTEF

FME-ZEN
DATA MANAGEMENT PLAN

ZEN MEMO No. 33 – 2021

Sobah Abbas Petersen | Dept. of Computer Science, NTNU

Research Centre on
ZERO EMISSION
NEIGHBOURHOODS
IN SMART CITIES

ZEN MEMO No. 33

Sobah Abbas Petersen, Dept. of Computer Science, NTNU

FME ZEN Data Management Plan

Norwegian University of Science and Technology (NTNU) | www.ntnu.no
SINTEF Community | www.sintef.no

https://fmezen.no

Preface

Acknowledgements

This report has been written within the Research Centre on Zero Emission Neighbourhoods in Smart Cities (FME ZEN). The author gratefully acknowledge the support from the Research Council of Norway, the Norwegian University of Science and Technology (NTNU), SINTEF, the municipalities of Oslo, Bergen, Trondheim, Bodø, Bærum, Elverum and Steinkjer, Trøndelag county, Norwegian Directorate for Public Construction and Property Management, Norwegian Water Resources and Energy Directorate, Norwegian Building Authority, ByBo, Elverum Tomteselskap, TOBB, Snøhetta, Asplan Viak, Multiconsult, Sweco, Civitas, FutureBuilt, Hunton, Moelven, Norcem, Skanska, GK, Nord-Trøndelag Elektrisitetsverk - Energi, Smart Grid Services Cluster, Statkraft Varme, Energy Norway, Norsk Fjernvarme and AFRY.

The Research Centre on Zero Emission Neighbourhoods (ZEN) in Smart Cities

The ZEN Research Centre develops solutions for future buildings and neighbourhoods with no greenhouse gas emissions and thereby contributes to a low carbon society.

Researchers, municipalities, industry and governmental organizations work together in the ZEN Research Centre in order to plan, develop and run neighbourhoods with zero greenhouse gas emissions. The ZEN Centre has nine pilot projects spread over all of Norway that encompass an area of more than 1 million m^2 and more than 30 000 inhabitants in total.

In order to achieve its high ambitions, the Centre will, together with its partners:

- Develop neighbourhood design and planning instruments while integrating science-based knowledge on greenhouse gas emissions;
- Create new business models, roles, and services that address the lack of flexibility towards markets and catalyze the development of innovations for a broader public use; This includes studies of political instruments and market design;
- Create cost effective and resource and energy efficient buildings by developing low carbon technologies and construction systems based on lifecycle design strategies;
- Develop technologies and solutions for the design and operation of energy flexible neighbourhoods;
- Develop a decision-support tool for optimizing local energy systems and their interaction with the larger system;
- Create and manage a series of neighbourhood-scale living labs, which will act as innovation hubs and a testing ground for the solutions developed in the ZEN Research Centre. The pilot projects are Furuset in Oslo, Fornebu in Bærum, Sluppen and Campus NTNU in Trondheim, an NRK-site in Steinkjer, Ydalir in Elverum, Campus Evenstad, NyBy Bodø, and Zero Village Bergen.

The ZEN Research Centre will last eight years (2017-2024), and the budget is approximately NOK 380 million, funded by the Research Council of Norway, the research partners NTNU and SINTEF, and the user partners from the private and public sector. The Norwegian University of Science and Technology (NTNU) is the host and leads the Centre together with SINTEF.

https://fmezen.no
@ZENcentre
FME ZEN (page)

Abstract

This document describes the initial Data Management Plan (DMP) for the FME Zero Emission Neighbourhood (ZEN) Research Centre. A DMP is a document that describes how data in a research project will be managed, from project start up, throughout the research process and in the time after completion of the project. It describes what data will be collected or generated, how the data will be stored, described with metadata, analysed and, if possible, shared. It is aimed as a tool for planning and raising awareness among the researchers and should be a "living document" which is updated during the course of the research project.

This document will describe the guidelines and requirements for every step of the data lifecycle, to ensure that research that is conducted complies with national and international regulations, such as EU's General Data Protection Regulation (GDPR). Furthermore, it will provide recommendations to ensure that the data meets the aims of the FAIR (Findable, Accessible, Interoperable and Re-usable) data principles.

A DMP is required by the Norwegian Research Council (NRC) and EU Horizon 2020 program for all projects that are funded through these programs.

Contents

Tables

Figures

Acronyms

Acronym	Description
DMP	Data Management Plan
FAIR	Findability, Accessibility, Interoperability, Reusability
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
KPI	Key Performance Indicator
NRC	Norwegian Research Council
NSD	Norwegian Centre for Research Data
W3C DCAT	World Wide Web Consortium Data Catalogue Vocabulary
ZEN	Zero Emission Neighbourhood

1. Introduction

This document describes the initial Data Management Plan (DMP) for the FME Zero Emission Neighbourhood (ZEN) Research Centre. A DMP is required by the Norwegian Research Council (NRC) and EU Horizon 2020 program for all projects that are funded through these programs.

NTNU describes a DMP as a document that describes how data in a research project will be managed, from project start up, throughout the research process and in the time after completion of the project. A DMP describes what data will be collected or generated. The plan states how the data will be stored, described with metadata, analysed and, if possible, shared. The plan also addresses issues related to rights, privacy and costs. It is a tool for planning and raising awareness and should be a "living document" which is updated during the course of the research project.

The goal of the DMP is to describe how the data that is produced and collected within the project shall be collected, stored, used (real time or non-real time) and shared within the project partners and other stakeholders. This includes data that are produced through the research activities by the researchers in the project as well as the data that may be gathered from the pilots and case studies conducted within the project, by the researchers and other project partners.

This document will describe the guidelines and requirements for every step of the data lifecycle, to ensure that research that is conducted complies with national and international regulations, such as EU's General Data Protection Regulation (GDPR). Furthermore, it will provide recommendations to ensure that the data meets the aims of the FAIR (Findable, Accessible, Interoperable and Re-usable) data principles.

Data management is important to ensure good and uniform data management practices among all the participants of the project. Good data management practices can support increased transparency, easy access to data within the project and for other stakeholders, reduce the risk of losing data, save time when documenting research and facilitates future reuse and sharing of data.

The following documents have been consulted as guidelines for structuring and the contents of this document:

- Science Europe, Practical Guide to the International Alignment of Research Management, available from https://www.forskningsradet.no/contentassets/e4cd6d2c23cf49d4989bb10c5eea087a/se_rdm_practical_guide_final.pdf
- NTNU's guidance on DMP, available from https://innsida.ntnu.no/wiki/-/wiki/English/DMP+guidance
- SINTEF's guidance on DMP, available from SINTEF's internal website: https://sintef.sharepoint.com/sites/stottetjenester/forskningsdata/datahaandteringsplan---dmp/Sider/default.aspx.

2. Concepts and Definitions

A number of different kinds of data are relevant to the FME-ZEN project. This section provides an overview of the relevant concepts and definitions related to data management and their descriptions from the literature.

Personal data

According to the Norwegian Data Protection Agency, personal data is defined as follows (translated from Norwegian):

> Personal information is all information and assessments that can be linked to an individual. Typical personal information are the name of a person, address, telephone number, e-mail and ID number. An image is considered personal information if people can be recognised, and audio recordings can be personal information even if a name is not mentioned in the recording. Biometrics such as fingerprints, iris patterns, head shape (for facial recognition) are also personal information. Furthermore, a dynamic IP address is also defined as personal information. The registration number of a car can be personal information if it can be linked to a specific person, while not if it is on a company car used by several people. Information about behavior patterns is also considered personal information. Information about what you shop for, what shops you go to, what TV series you watch, where you physically move during a day and what you search for online are all examples of this[1].

In GDPR, personal data is defined, in Article 4, as follows:

> "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".[2]

GDPR requires that personal data is processed with the aim to irreversibly prevent the identification of the individual to whom the data relates to. The GDPR is focused on the protection of personal data, not merely the privacy of personal data.

Of particular interest to the FME-ZEN project is energy related data from buildings. According to the Norwegian Data Protection Agency, data that is collected for learning or making decisions about an individual are also personal data, which includes records about electricity and water usage.[3]

Sensitive Data

GDPR distinguishes between sensitive and non-sensitive personal data. Sensitive personal data is defined as data that reveals a person's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person and data concerning health or a natural person's sex life and/or sexual

[1] https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/
[2] https://gdpr-info.eu/art-4-gdpr/
[3] https://www.datatilsynet.no/personvern-pa-ulike-omrader/overvaking-og-sporing/strommaling/

orientation. Non-personal sensitive information is often accessible through publicly available data sources. Such data on its own does not identify a person; however, when combined with other data, may lead to the identification of data. An example of such data could be the Postcode of a person.
In a project as diverse as FME-ZEN, there may be non-personal sensitive data and information such as politically sensitive information, strategically sensitive information for organisations and the research activities and trade secrets. Examples of such data may include results of research or data pertaining to specific buildings.

Metadata

Metadata is data that provides structured information about other data[4]. Metadata summarises basic information about data, making data discovery and working with particular instances of data easier. The completeness of Metadata depends on the relevant characteristics of the data that is required to describe it, explain it, easier to find it and reuse it. NISO[5] identifies three distinct types of metadata[6]: descriptive metadata, structural metadata, administrative metadata, and they are described below:

- **Descriptive** metadata is descriptive information about a resource. It is used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
- **Structural** metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.
- **Administrative** metadata is information to help manage a resource, such as resource type, permissions, and when and how it was created, when and how it was created, file type and other technical information, and who can access it. There are several subsets of administrative data; two that are sometimes listed as separate metadata types:
    - Rights management metadata, which deals with intellectual property rights, and
    - Preservation metadata, which contains information needed to archive and preserve a resource.

Other types of metadata that have been identified are Reference metadata, which is information about the contents and quality of statistical data; and Statistical metadata, also called process data, which describes processes that collect, process or produce statistical data.
Some common standards for metadata are Dublin Core[7] and W3C DCAT[8]. Metadata is discussed in Section 6.2.

[4] https://www.opendatasoft.com/blog/2016/08/25/what-is-metadata-and-why-is-it-important-data
[5] https://www.niso.org/publications/understanding-metadata
[6] https://marciazeng.slis.kent.edu/metadatabasics/types.htm
[7] https://www.dublincore.org/specifications/dublin-core/dces/
[8] https://www.w3.org/TR/vocab-dcat/

Research Data

Research data is defined by NRC as "registrations / records / reports in the form of numbers, texts, images and sound that are generated or occur during research projects" (translated from Norwegian). NTNU uses this definition and identifies that research data can be both digital and analogue. They may, for example, be documentation of procedures, collected interview data, large response datasets, continuous data streams from sensors, archived material and sensitive personal data and physical objects or digital representations of these[9]. Research data also includes scientific publications and reports related to the research results.

NRC distinguishes between source data and result data. Source data is information that already exists in various formats and that forms the entrance to or basis for new research. Results data are data that are generated through research and that in the next round could be source data for new research activities.

Secondary Data

Secondary data refers to "data that already exist, regardless of the research to be conducted. These may comprise information collected for a different purpose (e.g. public administrative data, clinical data or weather data) or they may be physical or digital collections of objects and texts (e.g. libraries or data reused from previous projects such text corpuses or other scientific collections). Information on the Internet may also be defined as secondary data in this context, and such information is highly heterogeneous"[10].

Data Source

A data source may be the initial location of the data or where physical information is first digitised. Concretely, a data source may be a database, a flat file, live measurements from physical devices, scraped web data. An example of a data source in ZEN could be measurements from sensors on buildings.

Data Series

A data series is a related set of data points, which is a collection of observations obtained through repeated measurements over time. An example of a data series relevant for ZEN is time series data, such as sensor data measured over time.

Datasets

A dataset is a collection of data such as one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables. For example, the values from one or more sensor at any point in time.

[9] https://innsida.ntnu.no/documents/portlet_file_entry/10157/%C3%85pne+data_politikk.pdf/e9cfb26d-c314-4f2a-9c24-32b2cf5916fb?status=0
[10] https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

Open Data

Open data can be defined as "data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike"[11]. A precise meaning of "open" with respect to knowledge is available: "Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness"[12].

NTNU's policy is that research data shall, in principle, be openly available, except when security, privacy, legal or commercial matters so require. NTNU adheres to the FAIR[13] principles for the management of research data and shall use licenses for research data as these require.

FAIR Data Principles

The FAIR Principles for sound data management have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for findable, accessible, interoperable and reusable. Research data must be of quality that makes them accessible, findable and reusable. The concept interoperable entails that both data and metadata must be machine-readable and that a consistent terminology is used. Guidelines for achieving FAIR data in the FME-ZEN project are described in Section 6.

Data Model

A Data Model is an abstract model that organises data description, data semantics, and consistency constraints of data. The data model emphasises on what data is needed and how it should be organised instead of what operations will be performed on the data[14]. The primary goals of using a data model are to ensure that all data objects required by a database are accurately represented and it can help to define the relational tables, primary and foreign keys and stored procedures in databases. It can help to identify missing and redundant data. Open Data Models can facilitate data exchange through integration or bulk import of data.

Open Science

Open Science refers to scientific practice in which processes and results are made openly accessible under conditions that promote quality and knowledge development, including the sharing and utilisation of the research-based knowledge in a socially responsible manner.

Privacy by Design

GDPR draws attention to the concept "Privacy by Design", which states that organisations that develop software need to consider privacy at the initial design stages and throughout the complete development process of new products, processes or services that involve processing personal data. NCR provides guidelines for supporting privacy by design[15].

An overview of the types of data described in this section and examples of them are provided in Table 1.

---

[11] https://opendatahandbook.org/guide/en/what-is-open-data/
[12] https://opendefinition.org/od/2.1/en/
[13] https://www.force11.org/group/fairgroup/fairprinciples
[14] https://www.guru99.com/data-modelling-conceptual-logical.html
[15] https://www.datatilsynet.no/en/about-privacy/virksomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/

Table 1 - Overview of Data Types

Concept/Data Type	Examples
Personal data	Person's name, identification number
Sensitive Personal Data	Person's opinion
Non-sensitive personal data	Electricity consumption from households
Non-personal sensitive data	Specific research results, Specific information related to pilots
Metadata	Type of data, when data was created
Research Data	Data from sensors, simulated data, interview data
Secondary Data	Weather data, data collected during the ZEN project
Open Data	Weather data

3. Categories of Data in FME-ZEN

The main aim for the data management in FME-ZEN is the collection, aggregation, storage, analysis and management of Key Performance Indicator (KPI) data to track the project's and pilots' progress. Several categories of data have been identified in the document ZEN *Data Management and Monitoring: Requirements and Architecture D1.1.2, V2*, February 2019. They are:

- KPI data
- Research data
- Context data

In contrast to the general data types described in Section 2, these three categories have been identified to focus on the data types in the FME-ZEN project. The data that belong to these three categories can also be described as one or more of the data types described in Section 2. For example, KPI data may be research data or secondary data; research data may be sensitive personal or non-personal sensitive data and context data may be secondary data or research data. Examples are included in the following subsections.

3.1 KPI Data

In the context of ZEN, KPIs are a set of quantifiable performance measurements that define sets of values based on measured data from a project, making it easier to measure and track the neighbourhood's performance over time and against other similar projects[16]. A number of KPIs in a set of themes have been identified; some examples are listed below:

- Energy, for example energy efficiency in buildings, in kWh/m2/yr
- Emissions, for example GHG (Greenhouse Gas) Emissions in tCO_2eq
- Power or load (KW)
- Economy, for example Life Cycle cost of projects
- Mobility, for example modal split between different transportation modes

[16] Zero Emission Neighbourhoods in Smart Cities: Definition, Key Performance Indicators and Assessment Criteria: Version 1.0

- Spatial qualities, for example the delivery and spatial distribution of services and facilities in a neighbourhood
- Innovation, to estimate novel ideas towards ZEN and their impact

From the definition of KPI data, it is not likely to include sensitive personal data, such as the identification of a person. However, data such as energy consumption in households can be considered personal data. Similarly, depending on the pilot, the buildings and the organisations that are involved, some of the data may fall into the category of non-personal sensitive data.
Most of the KPI data can also be considered as research data; e.g. produced or derived during the FME-ZEN project.

3.2 Research Data

Research data, as defined in Section 2, are data that have been generated and/or used in the FME-ZEN project. This data may be generated and gathered by third-party applications or applications developed by the ZEN researchers and partners; e.g. simulation data or sensor data. This data which has been collected and used can come from pilots or entities/installations/buildings within pilot areas, prototypes from building and energy related data from buildings. Some examples of these data types are occupant behavior data, energy production and consumption. Research data, in this context, may include real-time data or near real-time data; e.g. sensor readings. Some of the research data gathered may be personal data; e.g. data that relates to occupant behaviour in identifiable spaces. Research data may also be historical data, which can serve as context data or secondary data for other researchers.

3.3 Context Data

In addition to the KPI and research data, the research conducted in FME-ZEN rely on data collected or generated by other sources, such as weather data, environmental conditions, energy systems, etc. This is context data, which is not always of interest solely on its own, but can add value when used in combination with other data; e.g. research data in ZEN. Context data supports the interpretation of results obtained by researchers. They may be secondary data, which may be historical data gathered from sensors or other means, research data such as simulated data or Open Data such as weather and climate related data.

4. Data Management Process

The process of using and generating data in research processes is described by the top part of Figure 1[17]. The first step is the collection or the production of data, which will be stored somewhere. The second step is the processing or analysis of the data. In this step, the data will be quality assured. This is also a process to determine whether the raw data will be preserved. The third step is preparing the data for long-term storage (archiving) and publication. This involves discipline-specific documentation and coding of data (metadata) describing i.e. who is responsible for the data, what the data contains, what has been done with the data, who can use them for which purpose. These metadata

[17] https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

will facilitate other researchers to find and reuse the data. In the fourth stage, the data is preserved over a longer period of time. It is important to ensure that the data is formatted appropriately and adheres to the FAIR data principles such that they are still findable, available, interoperable and reusable to users, despite changes in technology.

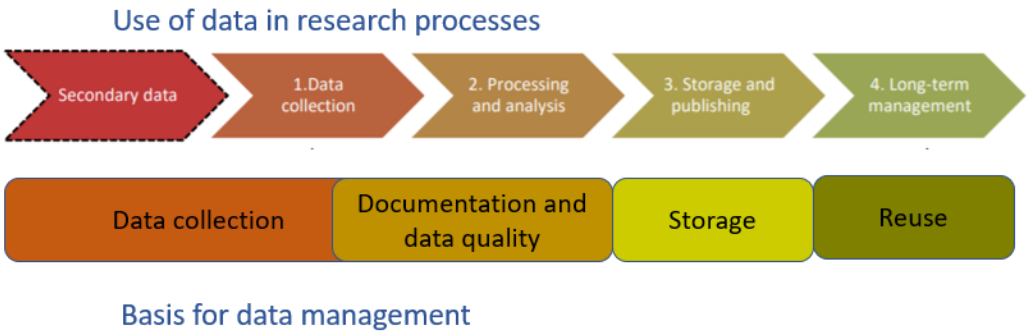

Figure 1 - Data in research processes

The guidelines for developing a Data Management Plan recommended by NTNU and NCR are thus based on the illustrated process and the main categories of guidelines are shown on the bottom part of Figure 1. As such, this document is structured according to this process and the recommended guidelines.

4.1 Data Collection

Data collection may include obtaining new data or existing data (secondary data) for conducting new research or analysis activities. Data may be captured from any source, by a variety of ways. Data may be captured without the source being aware of it or with the consent of the source, automatically by technology (e.g. sensors or cameras) or manually.
In the FME-ZEN project, a number of research activities are based on the development of models (simulation models, process models or analytical models), where the input data may consist of data collected from a building, a pilot neighbourhood or elsewhere and context data (e.g. weather data). The application of the models may result in new data, models, or other. With this in mind, data collection will include the following main activities:

- Preparation for collecting data.
- Collecting the data.
- Storing the data (that is collected as well as generated from the research activities).

Preparation for data collection will include identifying the need for collecting the data, clarifying the methodologies and where the data will be collected from. In particular, if the data collection is related to one or more pilots or ZEN cases, the preparation for data collection is an important activity, to clarify various points with the pilots (the person responsible for the pilot or other relevant persons), and to reach a common agreement about data collection, storage, access and reuse.

When **preparing to collect data**, it is important to determine if any personal data needs to be collected. If any personal data will be collected, there are 2 important activities that must be conducted:

1. Norwegian Data Protection Services' (NSD) notification form must be completed. This is a digital form that all researchers or anyone collecting personal data must fill in prior to

collecting and processing any personal data in a research; see Figure 2. This form is available from:
https://www.nsd.no/en/data-protection-services/notification-form-for-personal-data/

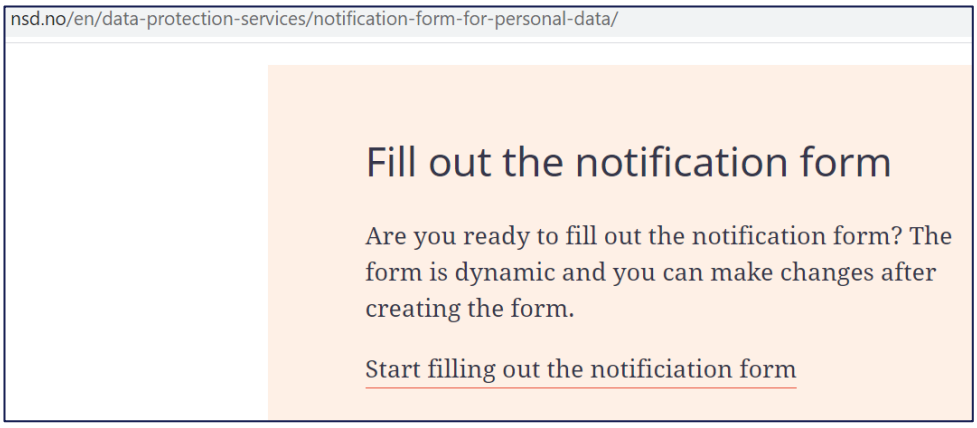

Figure 2 - NSD notification form, prior to collecting personal data

2. Obtain consent from the data providers. The Information and Consent form from NSD[18] must be used to inform the data subjects and to obtain their consent. In addition, most of the questions included in the table should be communicated to the owners of the data.

An example of a consent form is available from Appendix 1 – Consent Form for Data Collection. Guidelines and questions that should be considered when collecting data are provided in Table 2.

Table 2 – Data Collection Guidelines

Data Collection	Questions to consider
Description of the data	What data will be collected and for what purpose? How much data will be collected? What kinds of data will be collected; e.g. numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media? Can we use any existing data (rather than collect new data)?
Usage	What will the data be used for? Who will use the data? Will the data be used by other researchers or partners in the project? Will be the data be useful for stakeholders outside of the project?
Agreement with "pilot"	Is the data collected from a pilot neighbourhood or a ZEN case? Who needs to be contacted, informed about the data collection? Ensure compliance with relevant regulations and norms practised by the pilot.

[18] https://nsd.no/personvernombud/en/help/information_consent/index.html

Data Collection	**Questions to consider**
Target data providers or data sources	Who will provide the data – specific user groups, organisations, etc.? If the data already exists, where will be the data be obtained from?
Data aggregation	Will new data be aggregated with existing data? How can we preserve protection of data subjects through data aggregation?
Informed consent	If the data is personal data (e.g. energy data from households), how will compliance with GDPR be ensured? If the data is personal data, how will you ensure informed consent?
Methodology	How and when will the data be collected; e.g. using technology (e.g. sensors) or other (e.g. interviews)?
Metadata	What metadata will be stored and used to structure the data? What metadata that may be shared and/or reused?
Standards	What standards will be used for the data? Give preference to open and standard formats as they facilitate sharing and long-term re-use of data.
License (for ICT applications)	Which technology, if any, will be used? How will the technology be obtained? Who will pay/take responsibility for obtaining the software license?
Volume of data	How much data will be collected; Mbytes, GBytes, etc.? This will help to prepare for appropriate storage of the data.
License (for the data)	Who will have access to the data? What are the restrictions on accessing and reusing the data, if any?
Storage	How will the data be stored? Where will the data be stored (locally, or elsewhere)?
Anonymity	How will the data be anonymised so that the data subject is not identified?
Access	Who will have access to the data? Will the data be made available to researchers or partners in the project?
Quality of the process	How will you ensure the quality of the data collection process and methods? Who will ensure this?
Risk Assessment	Do you need to conduct a Data Protection Impact Assessment (DPIA)? If so, who will conduct it? According to GDPR, data collection that is likely to involve "a high risk" to people's personal information is required to conduct a DPIA[19].

[19] https://www.privacy-regulation.eu/en/article-35-data-protection-impact-assessment-GDPR.htm

Working with personal data involves some extra precautions. GDPR sets out seven key principles, which should be considered when collecting personal data:

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation (capture only data that is needed).
- Accuracy
- Storage limitation
- Integrity and confidentiality (security)
- Accountability

In addition to collecting new data, existing data may also be reused. Furthermore, there may be situations where existing data is reused by aggregating several existing data sources or by aggregating existing data with new data, which is collected or generated.

4.2 Documentation and Data Quality

FME ZEN is an ongoing project and the datasets and results from the project are likely to evolve over time. When new data is added, the datasets will need a description and metadata. Metadata is important to ensure that both humans and machines can understand and use the information the data set provides. Similarly, researchers from within the project and other partners are likely to share and reuse existing data. Thus, it is important to adhere to open and standard formats as they facilitate sharing and long-term reuse of data.

Dublin Core[7] and W3C DCAT[8] are common standards for metadata which are described in Section 6.2.

It is important to document the data to help other researchers in the project as well as secondary users to understand and reuse the data. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation may also include details on the methodology used, lab protocols, codebook, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data.

It is also important to indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and reuse.

A set of questions to consider to ensure appropriate documentation and quality of the data are provided in Table 3.

Table 3 - Documentation and data quality Guidelines

Documentation and Data Quality	Questions to consider
Documentation	What information is needed for the data to be read and interpreted? How will you capture / create this documentation and metadata?
Metadata	Are there any metadata standards used in your field? If not, how will you ensure that the data will be findable and possible to interpret in the future?
Standards	Which standard data format will be used? How will you ensure that the data will be findable, available, interoperable and reusable in the future?
Organisation of data	How will the data be organised? e.g. How will version control be managed? How will you ensure that the data is well-organised such that it will be easy to find, understand and re-use?
Data Quality	How will you ensure the quality of the data?
Quality of the process	How will you ensure the quality of the documentation of the data?

4.3 Data storage and Backup

Following the Research Councils policy on Open Science, every instance of research data must be stored in a safe and secure manner for a minimum of 10 years[20]. Research data should be "as open as possible, as closed as necessary". When deciding how to store research data, the data should first be classified in order to determine the value and identify the need for security. NTNU guidelines divides data into four categories: public, internal, confidential, and highly confidential.

- **Public**: Information that may be accessible to anyone, without special access rights
- **Internal**: Information which requires some protection and where accessibility should be restricted to selected internal and / or external users.
- **Confidential**: Information that requires strict access control. Examples of such information are some strategy documents, sensitive personal information, health information, exam papers before they are given, some types of research data and work.
- **Strictly confidentia**l: Information with this classification should not be used. Documents with information in this class must be processed in systems that are approved for this.

NTNU's guidelines for storing data are available from: https://innsida.ntnu.no/wiki/-/wiki/Norsk/Lagringsguide.
SINTEF's guidelines for storing data are available from: https://sintef.sharepoint.com/sites/stottetjenester/forskningsdata/forskningsdata---lagring-og-aktivt-arbeid/Sider/default.aspx

[20] https://www.forskningsradet.no/siteassets/tall-og-statistikk-seksjonen/apen-forskning/nfr-policy-open-science-eng.pdf

Of specific relevance to FME-ZEN are situations where data from the pilots and ZEN cases are involved. In situations where the data is stored or backed up by two different partners, it is important to clarify who is responsible and coordinate among the people that are responsible for the data.
A set of questions to consider to ensure appropriate storage and backup of the data are provided in Table 4.

Table 4 – Data Storage and Backup Guidelines

Data Storage and Backup	**Questions to consider**
Data category	What are the categories of data - public, internal, confidential or strictly confidential?
Data protection	How will you control access to the data? Who will be allowed access to the data? How will you ensure control of the data? How will you ensure compliance with data protection rules and regulations?
Storage	Where will the data be stored (locally, or elsewhere)? It is recommended to store data in at least two separate locations.
Protection of sensitive data	Do you have any sensitive data? How will you take care of sensitive data or personal data? How will you anonymise, pseudonymise and/or encrypt personal data?
Data Security	How will you ensure that your data is secure?
Backup	How will the data be recovered in the event of an incident? How will you ensure secure backup of your data?
Risk assessment	Are there known threats or vulnerabilities to the data? If so, how can they be removed or minimised? How will you assess the risks for data security? Who will do this? How will you mitigate risks to your data? How often will you do a risk assessment?

4.4 Legal and Ethical Requirements, Codes of Conduct

Research activities that collect or use personal data must comply with regulations (e.g. EU's GDPR). For FME-ZEN, when working with data related to the pilots or ZEN cases, extra care must be taken to ensure compliance with regulations and codes of conduct agreed with the relevant partners and stakeholders. Since each data collection activity is likely to be different, e.g. within different pilots or cases, attention must be given to the steps described in the earlier sub-sections every time data is collected.
A set of questions to consider to ensure that the legal and ethical requirements are fulfilled in the processing of data are provided in Table 5.

Table 5 – Legal and Ethical Requirements Guidelines

Legal & Ethical Requirements	Questions to consider
Personal Data	Is the data personal data or sensitive information? How will you manage personal data (e.g. access, transfer, removal? Have you obtained consent? How will you comply with regulations?
Intellectual Property Rights (IPR)	Who will be the owner of the data? Who will control the access to the data? (In general, if the research project is conducted by NTNU employees, NTNU will be the owner of the data sets.)
Access	Will the data be openly accessible or partially open after the research project? Who will manage this?
Ownership	Who owns the data? Will the data be shared by other researchers and partners? If so, by whom?
Third party data	Will there be use of third party data? How will you acquire third party data?
Ethics	Are there ethical issues with the data you collect or use? Who will be involved in an ethical review, if that is required?

4.5 Data Sharing and Long-term Preservation

Long-term data preservation concerns the storage of data in a trustworthy data depository, ensuring data security. To be able to share and reuse existing data, the data must be discoverable.
A set of questions to consider to ensure appropriate sharing and long-term preservation of the data are provided in Table 6.

Table 6 – Storage and Long-term Preservation Guidelines

Data Sharing & Long-term Preservation	Questions to consider
Preserving data	Do you need to preserve your data? What are the criteria for preservable data?
Archiving	What data or parts of the data will be archived? How will data be retained/archived? How long will the data be archived for? When will the data be made available?
Discovery	How will the data be discovered?
Persistent identifiers	How will you define persistent identifiers for the data? Who will do this?
Data sharing	Are there restrictions on sharing data? Which parts of the data can be shared or needs restricted access?

Data Sharing & Long-term Preservation	**Questions to consider**
	Who can the data be shared with? Who needs the data for conducting their research; e.g. within the project? How will the data be shared; e.g. transferred to the users?
Data aggregation	Will the data be aggregated with other data for new analyses? How can we ensure that aggregated data will not reveal the identity of the data subject (e.g. a person, or a specific building)?
Data embargo	Is there a need for embargoed data? Which data?
Open Data	Can the data or parts of the data be made available as Open Data?
Responsibility	Who will be responsible for the long-term preservation of the data?

Following the Research Council's policy on Open Access, all research data will in general be made openly accessible, but exceptions are made when the research can or should not be openly accessible. Reasons for restricting access include security concerns, sensitive personal data, legal factors and commercial factors. There are several data-archives that makes it possible to archive and share research data:

- NTNU Open Research Data: NTNU's institutional archive for open research data.
- Zenodo: A general archive for research data, that is run by CERN/EU.
- NSD: NSD's archive is a certified archive by Core Trust Seal and supports both publishing of open data and data with access restrictions.

The definition of Research data also includes scientific publications and reports related to the research results. Such results from the project will be made available through the project's webpages.

5. Data Management Responsibilities and Resources

The roles and responsibilities for the data management activities (data collection, quality assurance, storage, data security and sharing) must be identified. Some of the roles within data management activities in the FME-ZEN project are:

- Data collectors, who may be ZEN researchers – Post Doctoral researchers, PhD students and Master's students. It is important to identify who is responsible for the data collection activity and communicate this to the data providers, e.g. in the Consent form (Appendix 1). For example, if a Master's student is collecting data, the supervisor is usually responsible for the overall activity.
- Responsibility for the quality assurance of the data – the person(s) responsible for the quality assurance of the data must be identified and included in the process as relevant. For example, this may be the supervisor of a researcher or other.

- Responsibility for data storage – in some cases, a specific role may be responsible for storing the data, where the person(s) with this role may be responsible for one or more data collection activities.
- Data security – an important role is ensuring the security of all data (personal, research, etc.). It is important to have the required competences and procedures for data security and assessing the security risks.

The appointment of a Data Protection Officer (DPO) is mandatory in some instances[21]; e.g. if the data processing is carried out by a public authority or body, except for courts acting in their judicial capacity. The DPO for NTNU is Thomas Helgesen (https://www.ntnu.edu/employees/thomas.helgesen, thomas.helgesen@ntnu.no).

6. Guidelines towards FAIR Data

To make your research data more visible and reusable, it should follow the FAIR data principles – Findable, Accessible, Interoperable and Reusable. The ultimate goal of FAIR is to optimise the reuse of data. FAIR is not a standard; however, to achieve FAIR data, standards must be followed and APIs (Application Programming Interfaces) and protocols should be followed. APIs allow programmatic access to many databases and tools. They can directly access or query existing data, without the need to download entire datasets, which can be very large.
This section provides some guidelines to make your data FAIR, which are summarised in Table 7.

FAIR Data and Open Data are not the same; data can be FAIR and Open or FAIR and restricted. But it could be Findable, and the Metadata could be Accessible.

[21] https://gdpr-info.eu/art-37-gdpr/

Table 7 - FAIR Data guidelines

Findable	Accessible	Interoperable	Reusable
• Create a persistent and globally unique identifier, e.g. assign a DOI, which can make citations easy • Metadata and data are indexed • Describe using Metadata, which enhances documentation. • Examples of where data can be archived are: - NSD: https://www.nsd.no/arkiver-data/ - Zenodo: a general repository for research data, maintained by CERN/EU	• Access to data can be open, under embargo or restricted • Metadata is accessible, even if the data is restricted • Metadata is retrievable by the persistent identifier (DOI) • Use a standardised communication protocol to facilitate access and reuse (e.g. http) • Consider format of the data; Open, non-proprietary or common formats will increase accessibility	• Data is documented by using well known schemas to allow data to be combined and exchanged • Data is presented in a machine-actionable format • For software, provide a clear and concise description of the operations along with the corresponding input and output data types. Describe the data formats for input and output. • For data, create an API and share. Open Source API editors can be used to design, describe and document APIs, e.g. https://swagger.io/tools/swagger-https://github.com/OAI/OpenAPI-Specification editor/,	• Assign a license to the data, which documents the author's requirements for reuse. E.g. Creative Commons (https://creativecommons.org/licenses/) • Ensure appropriate and adequate Metadata fields that facilitates reuse. • Metadata and data are registered or indexed in a searchable resource (infrastructure). • For software, document and describe the software; how to install, run and use. Describe the functionalities, provide examples. • For datasets, document and describe how to use them.

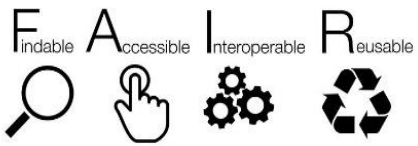

Image CC-BY-SA by SangyaPundir

6.1 Checklist for FAIR Data

A **checklist** to determine how FAIR your data is has been adopted from Sarah Jones & Marjan Grootveld, EUDAT.

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed, so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

6.2 The role of Metadata

The role of Metadata and rich Metadata is important to make research data FAIR.
Metadata summarises basic information about data, making data discovery and working with particular instances of data easier. Examples of Metadata are illustrated in Figure 3.

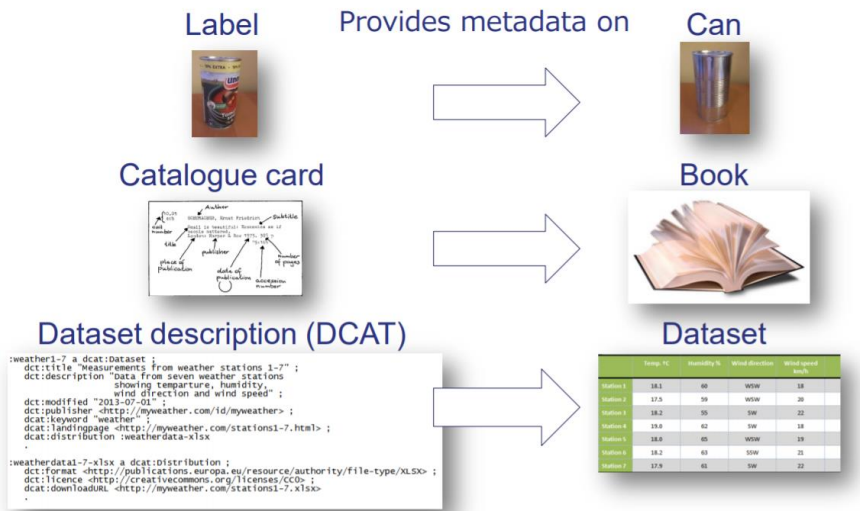

Figure 3 - Examples of Metadata[22]

The relevance of Metadata is further illustrated in Figure 4, where Metadata about research data, e.g. a data series, is made public through a data catalogue, which could be found and reused by others, e.g. a project or city dashboard, or other.

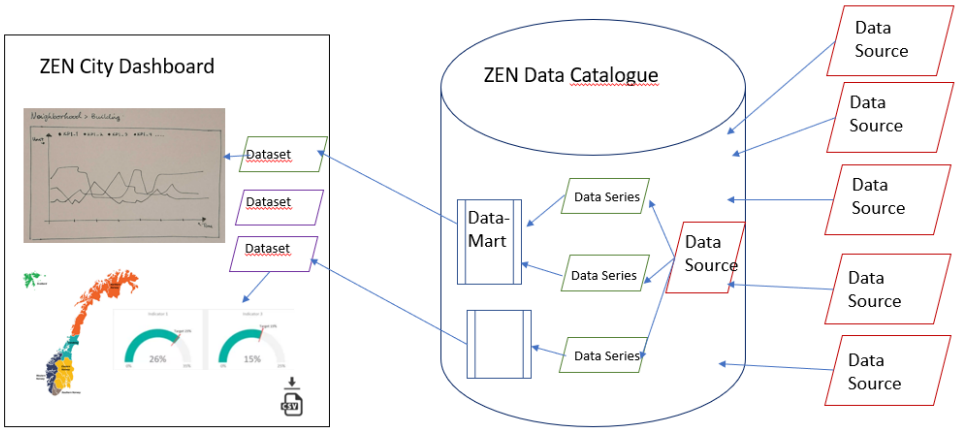

Figure 4 - Relevance of MetaData for FAIR Data

Dublin Core[7] and W3C DCAT[8] are common standards for Metadata.

22 https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_1.4_introduction_to_metadata_management_en_edp.pdf

Dublin Core is a set of 15 generic, widely used elements that provide information about the data. The 15 elements are: Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights. A user guide for Dublin Core is available from: https://www.dublincore.org/specifications/dublin-core/usageguide/.
An example describing Metadata using Dublin Core is shown in Table 8.

Table 8 - Example of Metadata using Dublin Core

Dublin Core Element	**Use**	**Example Value**
Creator	An entity primarily responsible for making the resource.	Researcher Zen
Contributor	An entity responsible for contributing to the resource.	Student Zen
Publisher	An entity responsible for making the resource available.	Researcher Zen
Title	A name given to a resource	Energy profile building X
Date	A period or time associated with an event in the lifecycle of the resource.	01.01.20-01.01.23
Language	Language of the resource.	Norwegian Bokmål
Format	The file format, physical medium or dimension.	CSV
Subject	The topic of the resource.	Energy related data
Description	An account of the resource.	Energy profile of Building X during the period t1-t2.
Identifier	An unambiguous reference to the resource within a given context.	www.xxx.no
Relation	A related resource.	
Source	A related resource from which the resource is derived.	
Type	The nature or genre of the resource.	
Coverage	The spatial or temporal topic; the jurisdiction under which the resource is relevant.	Norway
Rights	Information about rights held in and over the resource.	

W3C DCAT (World Wide Web Data Catalog Vocabulary) supports interoperability of data portals and catalogues across Europe. This format is particularly relevant for creating metadata for some of the research data produced in ZEN, such as time series data. The type of information that should be captured in the metadata are include Title, Date created, Date modified, Description, Publisher, License, Keywords, Update frequency, Distribution. An example describing Metadata using W3C DCAT is shown in Table 9.

Table 9 - Example describing Metadata using W3C DCAT

W3C DCAT Element	Use	Example Value
Title	A name given to the data series or data set.	Energy profile building X
Date Created	Date when the data was created	01.01.20
Date modified	Date when the data was last modified	01.02.20
Description	A description of the data.	Energy profile of Building
Publisher	An entity responsible for making the data available.	Researcher Zen
License	The conditions or restrictions that apply to the use of a Distribution	Public domain
Keywords	The file format, physical medium or dimension.	Energy, Buildings, Zero Emission
Theme	A theme that the data set belongs to	Energy related data
Update frequency	How often is the dataset updated	Once a day

Note that additional elements can be added to the W3C DCAT metadata descriptions.

7. Overview of Useful Links

Description	Link
NTNU DMP guidance	https://innsida.ntnu.no/wiki/-/wiki/English/DMP+guidance
SINTEF DMP guidance	https://sintef.sharepoint.com/sites/stottetjenester/forskningsdata/datahaandteringsplan---dmp/Sider/default.aspx
NSD Support for creating a DMP	https://nsd.no/arkivering/en/data_management_plan.html
NTNU's guidelines for storing data	https://innsida.ntnu.no/wiki/-/wiki/Norsk/Lagringsguide
Template for Information & Consent form from NSD	https://nsd.no/personvernombud/en/help/information_consent/information_requirements.html

Appendix 1 – Consent Form for Data Collection

Request for participation to …..

Background and Purpose
Explain the aim of data collection, which data will be collected and how it will be collected.
The aim of the data collection is to …..
The data that will be collected will include…….
This work is conducted as a part of a the FME-ZEN project, at the Department of XXX, at NTNU.

What does participation in the project imply?
What is the expectation from the data provider, e.g. a participant in a survey or an interview*? *The example text is for collecting data through a questionnaire. Adapt the text as relevant.
The participation includes your response to this online questionnaire. The questionnaire includes a set of statements and the participants are requested to indicate their level of agreement to the statements, using a Likert scale.

What will happen to the information about you?
Explain how the data will be treated, e.g. where it will be stored, who will have access to it, what will happen to the data once the project is completed, etc. Adapt the example text as relevant.
All personal data will be treated confidentially. The data that is gathered will be stored on NTNU servers, ensuring anonymity of the respondents. The participants will not recognisable in any of the publications that report the results from this study.

Voluntary participation
It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason. If you decide to withdraw, all your personal data will be made anonymous.

If you would like to participate or if you have any questions concerning the project, please contact:
XXX, NTNU, email: xxx@ntnu.no.

Consent for participation in the study

State explicitly what you are asking consent for. You can also ask for signatures of participants. Note: you need consent from parents/guardians of minors.
I have received and understood information about the research project. I give consent:

- ☐ to participate in the data collection.

VISION:

«Sustainable neighbourhoods with zero greenhouse gas emissions»

Research Centre on
ZERO EMISSION
NEIGHBOURHOODS
IN SMART CITIES

CENTRE FOR ENVIRONMENT-FRIENDLY ENERGY RESEARCH
The Research Council of Norway

NTNU

SINTEF

https://fmezen.no