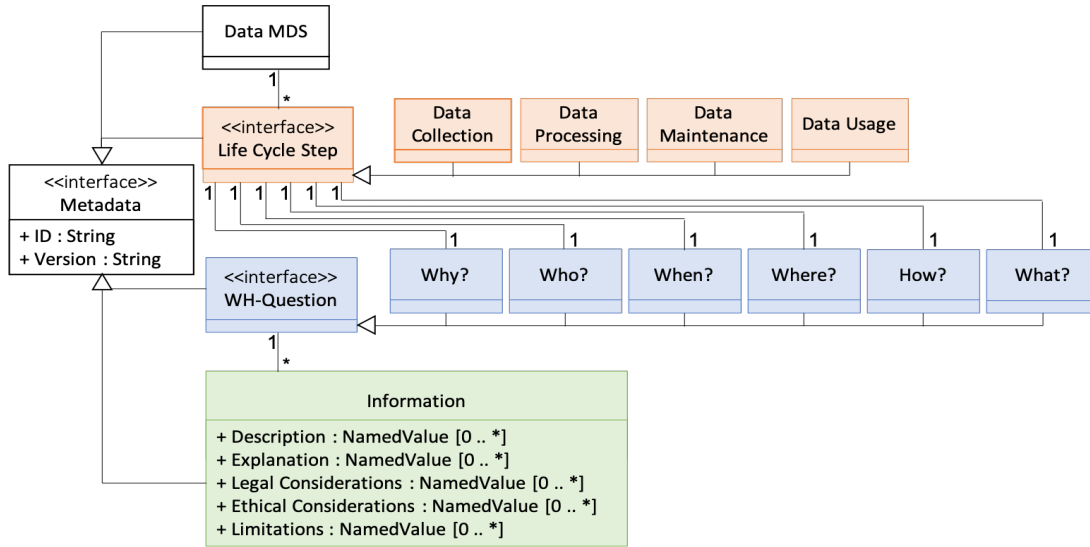


# LiQuID - A Metadata Model for Accountable Data

LiQuID is a novel metadata model for accountable data. It is a hierarchical model comprising three levels: the top level models the different data life cycle steps, we therefore call it **lifecycle level**. For each such step, the second level, named **question level**, structures metadata according to questions relevant for accountable data. The **information level** at the leaf level models the actual information per life cycle step and question. Note that by extending a general metadata interface, each element of the metadata model is uniquely identified and may have multiple versions.



This document provides exhaustive descriptions for the information fields of all combinations of life cycle steps and questions yielding 120 information field descriptions.

## Data Collection

### Why?

Information	Why was the data set created?
Explanation	Why is a dataset for this purpose needed?
Legal Considerations	Why is it legal to create a data set for this cause?
Ethical Considerations	Why is it ethical to create a data set for this cause?
Limitations	What are problems or limitations of this creation goal? What efforts have been made to mitigate the identified problems and limitations?

### Who?

Information	Who (people, organizations) was involved in the data collection process? Provide all information relevant to their identification, their role in the data collection process, all information necessary to assess their qualifications to fulfill this role, and all characteristics which could have an influence on the data set.
Explanation	Why were these particular people involved in the data collection process?
Legal Considerations	Why was it lawful that these people were involved in the data collection process?
Ethical Considerations	Why was it ethically justifiable that these people were involved in the data collection process?

Limitations	<p>What general problems or limitations in the data set could result from the selection of persons involved in the data set collection process and their characteristics or qualifications?</p> <p>What problems or limitations for the overall objective (Why?) could result from the choice of people?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>
-------------	---

## When?

Information	<p>On what date(s) or time frame(s) has the data been collected/ created?</p> <p>It must also be possible to place the data in a temporal context.</p>
Explanation	<p>Why was the data collected at this particular time(s)?</p>
Legal Considerations	<p>Why was it lawful that the data was collected at this time(s)?</p>
Ethical Considerations	<p>Why was it ethically justifiable to collect the data at this time(s)?</p>
Limitations	<p>What general problems or limitation in the data set could result from the choice of time(s) or time frame(s) of the data collection process?</p> <p>What problems or limitations for the overall objective (Why?) could result from the choice of time(s) or time frame(s)?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>

## Where?

Information	<p>Where was the data set collected (country, place, website, ...)?</p> <p>It must also be possible to place the data in a spatial context.</p>
Explanation	<p>Why was the data set collected at this particular place(s)?</p>
Legal Considerations	<p>Why was it lawful to collect the data set at this place(s)?</p>
Ethical Considerations	<p>Why was it ethically justifiable to collect the data set at this place(s)?</p>
Limitations	<p>What general problems or limitations in the data set could result from the choice of place(s) of the data collection process?</p> <p>What problems or limitations for the overall objective (Why?) could result from the choice of place(s)?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>

## How?

Information	<p>What was the methodology/ procedure for data collection?</p> <p>Which methods and tools were exactly used in each step and what was the (technical) environment?</p>
Explanation	<p>Why was the data collected using this particular methodology/ procedure, methods, tools and (technical) environment?</p>
Legal Considerations	<p>Why was it lawful to collect the data using this methodology/ procedure, methods, tools, and (technical) environment?</p>
Ethical Considerations	<p>Why was it ethically justifiable to collect the data using this methodology/ procedure, methods, tools, and (technical) environment?</p>
Limitations	<p>What general problems or limitations in the data set could result from this choice of methodology/ procedure, methods, tools, and (technical) environment of the data collection process?</p> <p>What problems or limitations for the overall objective (Why?) could result from this choice?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>

## What?

Information	What data was collected? What concepts does it cover? What is a general description of the data set? What are the characteristics/ profile of the data set (dependent on data type)? What is the quality of the data set (quality metrics depend on data type)?
Explanation	Why is this data with these characteristics and these quality values the result of the data collection?
Legal Considerations	Why is it lawful to collect this kind of data?
Ethical Considerations	Why is it ethically justifiable to collect this kind of data?
Limitations	What are general problems or limitations of this data set? What are problems or limitations of the dataset with respect to the overall objective (Why?)? What efforts have been made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

## Data Processing

### Why?

Information	What is the goal of this data processing step?
Explanation	Why is this data processing step needed?
Legal Considerations	Why is it lawful to perform this data processing step?
Ethical Considerations	Why is it ethically justifiable to perform this data processing step?
Limitations	What are problems or limitations of the data set resulting from the choice of this data processing step? What efforts have been made to mitigate the identified problems and limitations?

### Who?

Information	Who (people, organizations) was involved in the data processing? Provide all information relevant to their identification, their role in the data processing, all information necessary to assess their qualifications to fulfill this role, and all characteristics which could have an influence on the data set.
Explanation	Why were these particular people involved in the data processing?
Legal Considerations	Why was it lawful that these people were involved in the data processing?
Ethical Considerations	Why was it ethically justifiable that these people were involved in the data processing?
Limitations	What general problems or limitations in the data set could result from the selection of persons involved in the data set processing and their characteristics or qualifications? What problems or limitations for the overall objective (Why?) could result from the choice of people? What efforts have been made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

### When?

Information	On what date(s) or time frame(s) has the data been processed?
Explanation	Why was the data processed at this particular time(s)?
Legal Considerations	Why was it lawful that the data was processed at this time(s)?
Ethical Considerations	Why was it ethically justifiable to process the data at this time(s)?

Limitations	<p>What general problems or limitation in the data set could result from the choice of time(s) or time frame(s) of the data processing?</p> <p>What problems or limitations for the overall objective (Why?) could result from the choice of time(s) or time frame(s)?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>
-------------	---

## Where?

Information	Where was the data set processed (country, place, website, ...)?
Explanation	Why was the data set processed at this particular place(s)?
Legal Considerations	Why was it lawful to processed the data set at this place(s)?
Ethical Considerations	Why was it ethically justifiable to process the data set at this place(s)?
Limitations	<p>What general problems or limitations in the data set could result from the choice of place(s) of the data processing?</p> <p>What problems or limitations for the overall objective (Why?) could result from the choice of place(s)?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>

## How?

Information	<p>What was the methodology/ procedure for data processing?</p> <p>Which methods and tools were exactly used in each step and what was the (technical) environment?</p>
Explanation	Why was the data processed using this particular methodology/ procedure, methods, tools and (technical) environment?
Legal Considerations	Why was it lawful to process the data using this methodology/ procedure, methods, tools, and(technical) environment?
Ethical Considerations	Why was it ethically justifiable to process the data using this methodology/ procedure, methods, tools, and(technical) environment?
Limitations	<p>What general problems or limitations in the data set could result from this choice of methodology/ procedure, methods, tools, and (technical) environment of the data processing?</p> <p>What problems or limitations for the overall objective (Why?) could result from this choice?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>

## What?

Information	<p>What data was the result of the processing?</p> <p>What concepts does it cover?</p> <p>What is a general description of the data set?</p> <p>What are the characteristics/ profile of the data set (dependent on data type)?</p> <p>What is the quality of the data set (quality metrics depend on data type)?</p>
Explanation	Why is this data with these characteristics and these quality values the result of the data processing?
Legal Considerations	Why is it lawful to process this kind of data?
Ethical Considerations	Why is it ethically justifiable to process this kind of data?
Limitations	<p>What are general problems or limitations of this processed data set? What are problems or limitations of the dataset with respect to the overall objective (Why?)?</p> <p>What efforts have been made to mitigate the identified problems and limitations?</p> <p>If there are no problems or limitations, state why.</p>

# Data Maintenance

## Why?

Information	Why will the dataset be further maintained?
Explanation	Why does it make sense to maintain the dataset?
Legal Considerations	Why is it lawful to maintain the dataset?
Ethical Considerations	Why is it ethically justifiable to maintain the dataset?
Limitations	What are problems or limitations of the data set resulting from this maintenance goal? What efforts have been made to mitigate the identified problems and limitations?

## Who?

Information	Who (people, organizations) will be involved in the data maintenance? Provide all information relevant to their identification, their role in the data maintenance, all information necessary to assess their qualifications to fulfill this role, and all characteristics which could have an influence on the data set.
Explanation	Why will these particular people be involved in the data maintenance?
Legal Considerations	Why will it be lawful that these people are involved in the data maintenance?
Ethical Considerations	Why will it be ethically justifiable that these people are involved in the data maintenance?
Limitations	What general problems or limitations in the data set could result from the selection of persons involved in the data set maintenance and their characteristics or qualifications? What problems or limitations for the overall objective (Why?) could result from the choice of people? What efforts have been made or will be made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

## When?

Information	On what date(s) or time frame(s) will the data be maintained? With which frequency?
Explanation	Why was the data processed at this particular time(s)?
Legal Considerations	Why will it be lawful that the data will be maintained at this time(s)?
Ethical Considerations	Why will it be ethically justifiable to maintain the data at this time(s)?
Limitations	What general problems or limitation in the data set could result from the choice of time(s) or time frame(s) of the data maintenance? What problems or limitations for the overall objective (Why?) could result from the choice of time(s) or time frame(s)? What efforts have been or will be made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

## Where?

Information	Where will the data set be maintained (country, place, website, ...)?
Explanation	Why will the data set be maintained at this particular place(s)?
Legal Considerations	Why will it be lawful to maintain the data set at this place(s)?
Ethical Considerations	Why will it be ethically justifiable to maintain the data set at this place(s)?
Limitations	What general problems or limitations in the data set could result from the choice of place(s) of the data maintenance? What problems or limitations for the overall objective (Why?) could result from the choice of place(s)? What efforts have been or will be made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

## How?

Information	What will be the methodology/ procedure for data maintenance? Which methods and tools will exactly be used in each step and what will be the (technical) environment?
Explanation	Why will the data be maintained using this particular methodology/ procedure, methods, tools and (technical) environment?
Legal Considerations	Why will it be lawful to process the data using this methodology/ procedure, methods, tools, and (technical) environment?
Ethical Considerations	Why will it be ethically justifiable to maintain the data using this methodology/ procedure, methods, tools, and (technical) environment?
Limitations	What general problems or limitations in the data set could result from this choice of methodology/ procedure, methods, tools, and (technical) environment of the data maintenance? What problems or limitations for the overall objective (Why?) could result from this choice? What efforts have been or will be made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

## What?

Information	What data will be the result of the data maintenance? What concepts does it cover? What is a general description of the data set? What are the characteristics/ profile of the data set (dependent on data type)? What is the quality of the data set (quality metrics depend on data type)?
Explanation	Why is this data with these characteristics and these quality values the result of the data maintenance?
Legal Considerations	Why will it be lawful to maintain this data?
Ethical Considerations	Why will it be ethically justifiable to maintain this data?
Limitations	What are general problems or limitations of this maintained data set? What are problems or limitations of the dataset with respect to the overall objective (Why?)? What efforts have been or will be made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.

# Data Usage

## Why?

Information	What has the data set been used for? For which other purposes can the published data set be used for?
Explanation	Why can or should the published data set be used for these purposes?
Legal Considerations	Why was it/ is it lawful to use the published data set for these purposes?
Ethical Considerations	Why was it/ is it ethically justifiable to use the published data set for these purposes?
Limitations	What should the published data set not be used for?

## Who?

Information	Who publishes this data set? Who has used/ can use the published data set?
Explanation	Why do these people publish this data set? Why have they previously used the data set? Why can or should they use the published data set in the future?
Legal Considerations	Why was it/ is it lawful for them to publish this data set? Why was it/ is it lawful for them to use the published data set?

Ethical Considerations	Why was it/ is it ethically justifiable for them to publish this data set? Why was it/ is it ethically justifiable for them to use the published data set?
Limitations	What general problems or limitations for data set usage could result from the selection of persons publishing the data set? Who was not able to use or should not be using the published data set?

## When?

Information	When can/ was the published data set be used? When is it available? Until what point in time is it valid?
Explanation	Why can the published data set be used at this time (frame)?
Legal Considerations	Why is it lawful to use the published data set at this time (frame)?
Ethical Considerations	Why is it ethically justifiable to use the published data set at this time (frame)?
Limitations	When should the published data set not be used?

## Where?

Information	Where is the data set published/ available? Where (place, geographically) can the published data set be used?
Explanation	Why is the published data set made available at this place? Why can the published data set be used at this place?
Legal Considerations	Why is it lawful to publish the data set at this place? Why is it lawful to use the published data set at this place?
Ethical Considerations	Why is it ethically justifiable to publish the data set at this place? Why is it ethically justifiable to use the published data set at this place?
Limitations	What general problems or limitations for data set usage could result from the selection of place where the data set is made available? Where should the published data set not be used?

## How?

Information	What is a recommended process for using the published data set? What are recommended methods, tools, and technical environments where the published data set can be used?
Explanation	Why should the published data set be used following this process and using these methods, tools, and technical environments?
Legal Considerations	Why is it lawful to use the published data set following this process and using these methods, tools, and technical environments?
Ethical Considerations	Why is it ethically justifiable to use the published data set following this process, and using these methods, tools, and technical environment?
Limitations	What process, tools, methods, and technical environment should not be used for using the published data set? What can users do if they run into problems/ limitations?

## What?

Information	What data is published for use? What concepts does it cover? What is a general description of the data set? What are the characteristics/ profile of the data set (dependent on data type)? What is the quality of the data set (quality metrics depend on data type)?
Explanation	Why is this data with these characteristics and these quality values published?
Legal Considerations	Why is it lawful to publish this data?
Ethical Considerations	Why is it ethically justifiable to publish this data?
Limitations	What are general problems or limitations of this data set? What are problems or limitations of the dataset with respect to the overall objective (Why?)? What efforts have been made to mitigate the identified problems and limitations? If there are no problems or limitations, state why.