DEUTSCHE BUNDESBANK EUROSYSTEM

INEXDA
Working group on data access
Final report

Technical Report 2020-05
May 2020
Version 1.0

Deutsche Bundesbank, Research Data and Service Centre

Written by all members of the INEXDA working group on data access

Research Data and Service Centre

Contributors to this report

Members of the INEXDA Working Group on Data Access

Institution	Name
Banca d'Italia	Laura Bartiloro
Banco de España	Ana Esteban
	Sonia Serra
Banco de México	Claudia Katheri Velázquez Villegas
Banco de Portugal	Paulo Guimarães
Banque de France	Vincent Guegan
Deutsche Bundesbank	Stefan Bender (Chair)
	Jannick Blaschke
	Christian Hirsch
European Central Bank	Juan Alberto Sánchez Hernández
Eurostat	Aleksandra Bujnowska

Disclaimer: The views expressed in this document reflect those of the contributors and are not necessarily those of the institutions they represent.

Keywords: INEXDA, data access, annodata, data sharing

Citation: Members of the INEXDA working group on data access (2020). Final report of the INEXDA working group on data access, Technical Report 2020-05, Deutsche Bundesbank, Research Data and Service Centre.

This publication is available on the INEXDA website (www.inexda.org).

Contents

Executive summary 4

1 Background 5
1.1 Working group tasks 5
1.2 Organisation and members 5
1.3 Calendar of events 5
1.4 Structure of this report 6

2 Outcome of the working group: The INEXDA annodata schema 7

3 Conclusion and the way forward 10

A Glossary 11

B Detailed description of the INEXDA annodata schema 12

C Entity relationship model of the INEXDA annodata schema 32

Executive summary

The INEXDA Working Group on Data Access focused on procedures for granting access to micro data for research and analytical purposes. Micro data access procedures must ensure ongoing compliance with the legal requirements and satisfy technical, organisational, and administrative conditions set forth by the data provider.

This report presents the main outcome of the working group, which is the development of a common taxonomy for describing all legal, technical, and organisational rules that have to be considered when making micro data available for research and analysis. We call this set of metadata items annodata (derived from "annotation to data") as they are designed to complement existing metadata schemas for micro data.

More specifically, the INEXDA annodata schema comprises the following three dimensions:

1. **Rules for individual micro dataset access**. Access protocols comprise criteria that have to be fulfilled to be granted data access. This typically requires information on the legal basis, affiliation of the researcher, and degree of anonymisation of the requested data, among others. A prerequisite for a clear assignment of access protocols to datasets is that datasets are unambiguously identified (e.g. by Digital Object Identifiers).
2. **Rules for combining micro datasets**. To fully leverage on micro data for analytical and research purposes, we need to be able and allowed to link them to other data. First, this requires information on applied methods and the underlying assumptions used when creating this link. The latter is especially important for researchers when deciding whether to use this link in their own research or analysis. Second, this requires information on the legal feasibility of combining datasets.
3. **Information on project and requesting parties**. The characteristics of researchers or analysts affect the decision on the set of access rules (laid out in 1) and rules for combining datasets (laid out in 2) that are valid for a specific project. Finally, in a project, all information on micro data and requesting parties is combined.

This working group laid the foundation for assessing a potential harmonisation of access procedures in INEXDA member institutions. Furthermore, the INEXDA annodata schema is a notable step towards the automation of workflows providing access to micro data.

1 Background

1.1 Working group tasks

Official statistics are increasingly making micro data available to help facilitate robust analytical evidence aimed at addressing important societal challenges of our time. Providing micro data for scientific research and analytical purposes requires developing a set of rules and procedures to grant access for the use of such data. For instance, (micro data) access procedures could comprise an evaluation of the requesting researcher's current CV or the obligation to sign a confidentiality declaration. Spanning a project's life cycle from request to publication, they must ensure ongoing compliance with the legal framework and satisfy technical, organisational, and administrative requirements set forth by the data provider.

During its fifth meeting in April 2019, the INEXDA Network,[1)] an international cooperative project of national central banks, national statistical institutes, and other international organisations, decided to set up a dedicated working group focusing on access to micro data. The INEXDA Working Group on Data Access focused on procedures for granting micro data access for research and analytical purposes, such as scientific external or internal research projects, and internal analytical studies for policy evaluation. The tasks set out in the working group's mandate were to:

1. take stock of existing access modes (e.g. on-site access, remote execution) and procedures currently used by micro data-providing institutions;
2. identify common features and differences, and discuss the potential harmonisation of procedures to agree on a standardised access procedure which may later serve as a best-practice guideline;
3. *if possible*: decide on a software solution that could fulfil the requirements laid down in (2).

1.2 Organisation and members

The following institutions participated in the working group: Banca d'Italia, Banco de España, Banco de México, Banco de Portugal, Banque de France, the Deutsche Bundesbank, the European Central Bank, and Eurostat. Stefan Bender from the Deutsche Bundesbank served as the working group's chairperson. The Deutsche Bundesbank also provided secretarial support.

1.3 Calendar of events

The INEXDA network launched the working group with a first workshop in Frankfurt am Main, Germany, in April 2019. The workshop's main aim was to bring together senior experts from central banks, national statistical institutes, international organisations, and research bodies to

1 The International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) provides a platform for exchanging experiences on statistical handling of granular data for central banks, national statistical institutes, and international organisations. As such, it supports the G20 process, notably the Data Gaps Initiative 2 recommendation aiming to promote the exchange of (granular) data as well as metadata. The network was founded in January 2017. For more information, see Bender, Stefan and the Members of the INEXDA network (2019), INEXDA – The granular data network, IFC Bulletins chapters 49.

share experiences of real-world approaches to facilitate access to micro data for research and analytical purposes.

The workshop revolved around the following topics:

- Data discovery centres: providing information on where to find the data
- Data access centres: how to grant access to the data: nationally and internationally
- Data hubs: access to multi-source remote data
- International initiatives on micro data access and sharing

Subsequently, the working group held two further meetings: in October 2019 in Frankfurt am Main, Germany, and in December 2019 in Porto, Portugal.

1.4 Structure of this report

This concluding report provides an overview of the working group's outcome. It is structured as follows: Section 2 explains the outcome of the working group revolving around the development of a common taxonomy to appropriately describe data access modes to micro data. Furthermore, Section 2 presents the main lessons learned while working on this common taxonomy. Section 3 concludes and states the way forward. Finally, the Annex contains a glossary, a detailed description of the INEXDA annodata schema and an entity relationship model of the same.

2 Outcome of the working group: The INEXDA annodata schema

The main outcome of the INEXDA Working Group on Data Access is the INEXDA annodata schema. Annodata (derived from "annotation to data") are structured information on the set of legal, technical, and organisational rules that make up micro data access procedures. In contrast to traditional metadata schemas,[2] which typically focus on the question "How are the data produced?", annodata help answer the question "How is access to the data managed?."[3]

The working group started developing a set of metadata items (which were later named *annodata* catalogue) designed to describe (micro data) access procedures in data-providing INEXDA institutions in a structured way. All members of the working group agreed that defining such a common schema is a necessary condition for the stocktake. Moreover, it will ease the development of best practices and support harmonisation and standardisation activities by e.g. ensuring an easy adaptation of an annodata-based software in different institutions.

When discussing all institution and dataset-specific requirements, it became clear that the schema had to be rather detailed. The working group further confirmed this initial assessment after collecting information necessary for an automation of micro data access-related tasks and decisions. However, manually filling out such a schema for a stocktake would be an extremely cumbersome and difficult task without the support of a dedicated software solution. For this reason, all members of the working group decided to focus on the design and fine-tuning of the INEXDA annodata schema and postpone the stocktake and associated harmonisation and standardisation activities to a later date when a first technical proof of concept is in place.

Members of the working group decided to structure the INEXDA annodata schema in eight parts representing eight types of annodata. All parts are interlinked with each other and can occur 1-n times. The purpose of parts 1 to 3 is to provide access rules for individual dataset families.[4] Parts 4 to 6 comprise rules for the combination of dataset families. This is important because different dataset families may be governed by different access regimes. Finally, parts 7 to 8 comprise information on the requesting party, e.g. a researcher or analyst, and the project, which need to be collected. Figure 1 illustrates the annodata schema's eight parts.

The remainder of this chapter provides a brief overview of the different annodata types. The complete INEXDA annodata schema can be found in Annex 2.[5] The internal structure of the single types of the annodata schema can be seen in the entity relationship model of the INEXDA annodata schema in Annex 3.

The first three parts of the schema are at a dataset family level:

1. **Access regime**: An access regime relates access protocols to access modes. Access modes

2 Such as the INEXDA metadata schema, which can be found here: Bender, Stefan, Hausstein, Brigitte and Christian Hirsch (2019), An introduction to INEXDA's metadata schema, IFC Bulletins chapters 49.
3 For more information on the general idea behind the annodata concept, see Bender, S., J. Blaschke, H. Doll, A. Gordon, C. Hirsch, D. Hochfellner, J. Lane (2019). The Annodata Framework: Putting FAIR data into practice. Technical Report 2019-03, Deutsche Bundesbank, Research Data and Service Centre.
4 For a definition of the term used in this section, see the glossary in Annex 1.
5 For simplicity, the schema refers to "data" instead of "micro data".

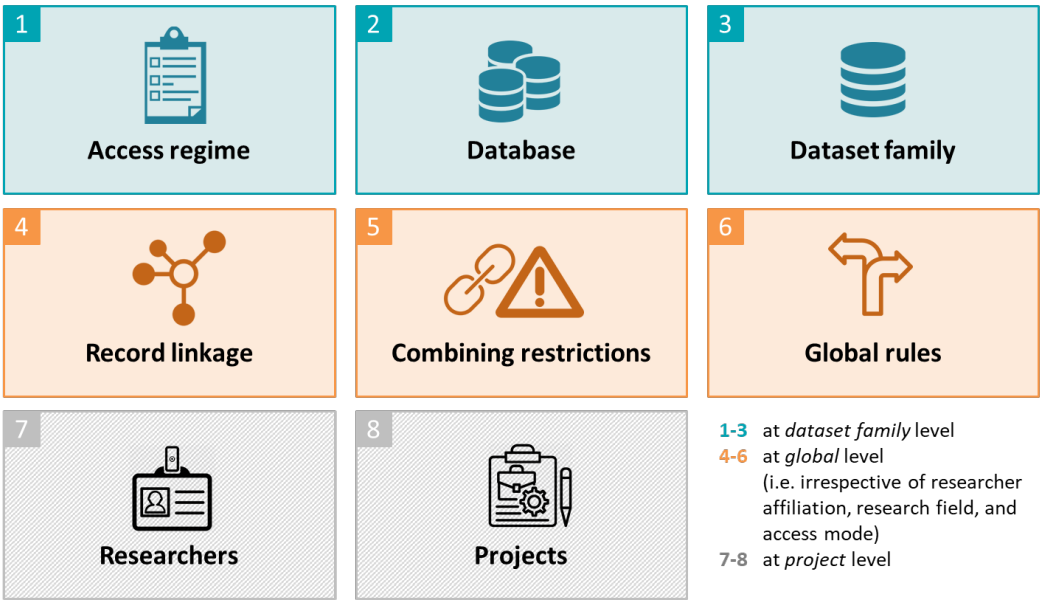


Figure 1: The eight parts of the INEXDA metadata schema

describe how researchers get access to a dataset family (e.g. on-site access, remote execution). Access protocols comprise criteria that have to be fulfilled to be granted data access. Besides the access modes, the shape of access protocols is determined by a combination of the legal basis, affiliation of the researcher, and degree of anonymisation of the requested data. As multiple dataset families can have the same access rules and procedures, an access regime can be assigned to 1-n dataset families. For example, two dataset families could be governed by the same set of rules laid out in a single legal framework. In this case, they could be assigned to a single access regime. On the other hand, one dataset family can only belong to one access regime ensuring an unambiguous assignment of applicable access rules.

2. **Database**: A database is a collection of 1-n dataset families. For instance, the "Money Market Statistical Reporting (MMSR)" would be a database.
3. **Dataset family**: A dataset family is a group of datasets which only differ in their temporal coverage. For example, "MMSR_formally-anonymised" would be a dataset family that consists of 1-n datasets, such as "MMSR_formally-anonymised-2020-03" or "MMSR_formally-anonymised-2020-04", and that belongs to the database MMSR. Applying the same logic "MMSR_non-anonymised" would be a different dataset family. The next three parts of the INEXDA annodata schema provide information at a global level, i.e. rules applying for multiple dataset families. These parts are necessary when a project combines multiple dataset families:
4. **Record linkage**: This part focuses on the technical feasibility of combining two or more dataset families. This includes information on applied methods as well as the underlying assumptions used when creating the link. Transparency over all techniques that have been applied when producing the mapping table is essential to inform researchers about the scope, precision, underlying assumptions, and time dependency of the mapping. This helps researchers assess whether the mapping is sufficient for their research purposes.
5. **Combining restrictions**: Combining restrictions refers to any legal or organisational rules prohibiting the combination of information from two or more dataset families. For example, combining a formally anonymised dataset with its equivalent non-anonymised version might lead to a decryption of the hash identifier and should not be allowed.

6. **Global rules**: This part provides information on the handling of conflicting rules or procedures that might occur when combining information from two or more dataset families with different access regimes. For example, two access regimes might require different contracts. In this case, the global rules would define whether (i) both contracts, (i) only contract A, or (iii) only contract B should be signed. Finally, the last two parts have information on the requesting party that needs to be connected to the rules outlined above. Therefore, the last two parts of the INEXDA annodata schema are at a project level:
7. **Researcher**: This part contains information on a researcher who requests access to the micro data described in parts 1 to 6.
8. **Project**: Finally, the last type describes those research projects for which any micro data described in parts 1 to 6 have been requested.

3 Conclusion and the way forward

The INEXDA Working Group on Data Access developed a common taxonomy for managing and exchanging information related to micro data access. The INEXDA annodata schema consists of eight types of annodata with information at dataset family, global, and project level. In a next step, INEXDA will work on the development of a proof of concept for a software solution to fill out the schema so that it can be used as a stocktake for micro data access processes used in INEXDA institutions.

Based on this stocktake, INEXDA will identify common features and differences, and assess a potential harmonisation of access procedures. In addition, INEXDA will continue working on tools to automate tasks and decisions in the context of micro data access management. In the context of this work, we expect to gain new insights that will be reflected in future revised versions of the INEXDA annodata schema.

Overall, the working group laid the foundation for subsequent INEXDA working groups on more granular topics, such as detailed requirements for the accreditation of institutions, contracts for research projects/bodies, modes for data provision, and statistical disclosure control.

A Glossary

(Micro data) access mode: An access mode describes an access protocol by which a specific type of researcher conducting a project in a specific type of research field is granted access to a specific type of dataset family.

(Micro data) access procedure: All steps, tasks, and decisions required in the context of data access management.

(Micro data) access protocol: Criteria that have to be fulfilled to be granted access under a specific access mode. Access protocol criteria are imposed by a combination of the legal basis, affiliation of the researcher, and degree of anonymisation of the requested data. Examples include application form, data provision, or signing a contract.

(Micro data) access regime: A collection of access protocols clustered in access modes.

(Micro data) access rule: Any requirement defined in the legislation or technical, organisational, and administrative requirements set forth by the data provider.

(Micro) database: A collection of data on a specific subject covering multiple (micro) dataset families.

(Micro) dataset family: Group of (micro) datasets of the same (micro) database which only differ in their temporal coverage.

(Micro) dataset: A set of data of a specific (micro) database covering a specific reference period.

Formally anonymised data: De-identified data which do not allow reporting agents or any other legal or natural person, entity, or branch to be directly identified (i.e. secure use files).

B Detailed description of the INEXDA annodata schema

(1) Access regime *(Occ. 1-n)*[1]		
Item	**Controlled field**	**Drop-down elements / Free text**
1.1 Identification of access regime *(Occ. 1)* Identification of the access regime.	**1.1.1 Access regime ID** *(Occ. 1)* Unique ID for the access regime.	[Free text]
	1.1.2 Full name of access regime *(Occ. 0-1)* Name of the access regime.	[Free text]
1.2 Access mode *(Occ. 1-n)* An access mode describes an access protocol by which a specific type of researcher conducting a project in a specific type of research field is granted access to a specific type of dataset family.	**1.2.1 Type of access mode** *(Occ. 1)* Mode of access that is available under the conditions stated above.	▪ **Secure on-site access**: Provision in the premises of the institution in a dedicated secure environment. ▪ **On-site access**: Provision via the institution's IT infrastructure. ▪ **Remote execution**: Researcher sends code and RDC sends back results. ▪ **On-site in partner institution**: On-site access in partner institution. ▪ **Remote on-site access**: Remote access for internal researchers (e.g. teleworking). ▪ **Remote access**: Researcher can access data remotely from own institution. ▪ **Download**: Scientific use files or public use files.
	1.2.2 Degree of anonymisation *(Occ. 1)* Information on the degree of anonymisation of the dataset families for which this access mode is available.	▪ **Non-anonymised**: No anonymisation (i.e. raw data). ▪ **Formal anonymisation**: Deletion of direct identifiers such as names, addresses, and other identifier (e.g. LEI). No direct identification possible (i.e. secure use files). ▪ **Factual anonymisation**: Identification possible only with significant effort (i.e. scientific use files).

[1] Occ. = Occurrence, i.e. the number of instances this item shows up in the database. For example, 1-n means that the item is a mandatory field (i.e. 1) and can appear up to n times.

		▪ **Perturbed data**: Data modified using statistical methods. ▪ **Full anonymisation**: No identification possible (i.e. public use files).
	1.2.3 Researcher type *(Occ. 1)* Type of researcher for which this access mode is available.	▪ **Internal (limited rights)**: Internal researcher with limited rights (e.g. temporary, not-accredited). ▪ **Internal (extended rights)**: Internal researcher with extended rights (e.g. permanent, accredited). ▪ **External (limited rights)**: External researcher with limited rights (e.g. externals without internal co-author). ▪ **External (extended rights)**: External researcher with extended rights (e.g. externals with internal co-author).
	1.2.4 Research field *(Occ. 1)* Research fields for which the access mode is available.	▪ **Any field**: Any field. ▪ **Scientific research**: Scientific research. ▪ **Monetary policy**: Monetary policy. ▪ **Tasks of the ESCB**: Exercising ESCB tasks.
	1.2.5 Access protocol *(Occ. 1)* Protocol which the researcher has to go through before being granted access via this mode (e.g. certificates that the researcher has to sign before being granted access via this mode).	[Free text]

2. Database *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
2.1 Identification of database *(Occ. 1)* Identification of the database.	**2.1.1 Database ID** *(Occ. 1)* Unique ID for the database.	[Free text]
	2.1.2 Full name *(Occ. 1)* Full name of the database.	[Free text]
	2.1.3 Acronym *(Occ. 0-n)* Acronym for the database.	[Free text]
2.2 Collection basis *(Occ. 0-n)* Basis for collection of the database (e.g. legal basis, commercial data, …).	**2.2.1 Name of collection basis** *(Occ. 1)* Name of the collection basis.	[Free text]
	2.2.2 Type of collection basis *(Occ. 1)* Type of the collection basis. Legal mandate (e.g. data from commercial providers).	▪ **Legal mandate**: Data collected under legal mandate. ▪ **Commercial providers**: Data from commercial data providers. ▪ **Other**: Data from other collection basis.
	2.2.3 Link to relevant documents *(Occ. 0-n)* Link to the collection basis.	[Free text]
2.3 Stakeholder *(Occ. 1-n)* The person, business or functional group responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.	**2.3.1 Stakeholder group ID** *(Occ. 1)* Unique ID for the stakeholder group.	[Free text]
	2.3.2 Stakeholder type *(Occ. 1)* Person, business or functional group responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.	[Free text]

	2.3.3 Name of group *(Occ. 0-1)* Name of the group.	[Free text]
	2.3.4 Group members *(Occ. 1-n)* List of group members.	[Free text]
	2.3.5 Email of group *(Occ. 1)* Email address of the group.	[Free text]
	2.3.6 Direct contact person *(Occ. 0-1)* Name of at least one direct contact person.	[Free text]

3. Dataset family *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
3.1 Identification of dataset family *(Occ. 1)* A dataset family is a partition of a database such that is governed by only one access regime.	**3.1.1 Dataset family ID** *(Occ.1)* Unique ID for the dataset family.	[Free text]
	3.1.2 Full name *(Occ. 1)* Full name of the dataset family.	[Free text]
	3.1.3 Acronym *(Occ. 0-n)* Acronym for the dataset family. If applicable.	[Free text]
3.2 Related database *(Occ. 1)* The name of the database from which the dataset family originates.	**3.2.1 Database ID** *(Occ. 1)* Unique ID of the related database.	[Free text]
3.3 Applicable access regime *(Occ. 1)*	**3.3.1 Access regime ID** *(Occ. 1)* Unique ID of the applicable access regime under which researchers can access the data.	[Free text]

A collection of access criteria clustered in access modes.		
3.4 Unit description *(Occ. 1-n)* The item "Unit Description" describes information about available identifier(s) in the dataset family (see INEXDA metadata schema[2]).	**3.4.1 Name of unit in dataset family** *(Occ. 1)* Name of the unit in the dataset family (e.g. banks, non-financial corporations, securities).	[Free text]
	3.4.2 Variable name of unit in dataset family *(Occ. 1-n)* Name of the variable which holds the unit.	[Free text]
3.5 Temporal coverage *(Occ. 1)* The sample period of the data object, i.e. the time period during which the data were collected or observations were made (see INEXDA metadata schema).	**3.5.1 Start of sample period - Year** *(Occ. 0-n)* Year (start of sample period).	[Drop-down menu with years]
	3.5.2 Start of sample period - Month *(Occ. 0-n)* Month (start of sample period).	[Drop-down menu with months]
	3.5.3 Start of sample period - Day *(Occ. 0-n)* Day (start of sample period).	[Drop-down menu with days]
	3.5.4 End of sample period - Year *(Occ. 0-n)* Year (end of sample period).	[Drop-down menu with years]
	3.5.5 End of sample period - Month *(Occ. 0-n)* Month (end of sample period).	[Drop-down menu with months]
	3.5.6 End of sample period - Day *(Occ. 0-n)*	[Drop-down menu with days]

[2] Note that some items may already be included in the INEXDA metadata schema but are also listed here as they are also relevant for the INEXDA annodata schema. For an introduction to the INEXDA metadata schema, see https://www.bis.org/ifc/publ/ifcb49_32.pdf

	Day (end of sample period).	
	3.5.7 Waves *(Occ. 0-n)* Dates of all waves, separated by a semicolon.	[Free text]
3.6 Anonymisation *(Occ. 1)* Information on the degree of anonymisation of the datasets in the dataset family.	**3.6.1 Degree of anonymisation** *(Occ. 1)* Degree of anonymisation of the datasets belonging to the dataset family.	▪ **Non-anonymised:** No anonymisation (i.e. raw data). ▪ **Formal anonymisation:** Deletion of identifiers, names, and addresses. No direct identification possible (i.e. scientific use files). ▪ **Factual anonymisation:** Data perturbation. Identification possible only with significant effort (i.e. secure use files). ▪ **Full anonymisation:** No identification possible (i.e. public use files).
	3.6.2 Comment *(Occ. 0-1)* Any other comments.	[Free text]
3.7 Identification of dataset *(Occ. 1-n)* A dataset is a partition of a database and organised in dataset families.	**3.7.1 Dataset ID** *(Occ. 1)* Unique ID of the dataset (e.g. DOI).	[Free text]
	3.7.2 Full name *(Occ. 1)* Full name of the dataset.	[Free text]
	3.7.3 Acronym *(Occ. 0-n)* Acronym for the dataset. If applicable.	[Free text]
	3.7.4 Applied technique for creation *(Occ. 0-1)* Name of the technique that has been applied to the 'clean copy' version of the dataset.	[Free text]

	3.7.5 Applied code for creation *(Occ. 0-1)* Link to the applied code.	[Link]
	3.7.6 Physical location *(Occ. 1-n)* Link to the location where data files are stored.	[Link]
	3.7.7 Format *(Occ. 1-n)* Formats in which file can be retrieved from the specified location.	▪ **csv:** Character-separated values. ▪ **text:** Text format. ▪ **dta:** Stata data format.

4. Record linkage - mapping table *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
4.1 Name of mapping table *(Occ. 1)* Mapping table (i.e. a table that contains a mapping of different identifiers, such as ISIN, MFI code) which contains at least one identifier from the dataset family.	**4.1.1 Mapping table ID** *(Occ. 1)* Unique ID for the mapping table.	[Free text]
	4.1.2 Full name *(Occ. 1)* Full name of the mapping table.	[Free text]
	4.1.3 Acronym *(Occ. 0-n)* Acronym for the mapping table. If applicable.	[Free text]
4.2 Time period *(Occ. 1)* Time period for which the mapping table is valid.	**4.2.1 Valid from** *(Occ. 1)* Date from which the mapping table is valid (format: YYYY-MM-DD).	[Free text]
	4.2.2 Valid until *(Occ. 1)* Date until which the mapping table is valid (format: YYYY-MM-DD).	[Free text]
	4.3.1 Applied technique *(Occ. 1)*	[Free text]

4.3 Creation of this version *(Occ. 1-n)* Time period for which the mapping table is valid.	Name of the technique that has been applied to produce the mapping table (e.g. record linkage).	
	4.3.2 Applied code *(Occ. 1)* Name of the technique that has been applied to produce the mapping table (e.g. record linkage).	[Link]
	4.3.3 Documentation *(Occ. 0-n)* Link to a document describing the process of mapping table production. Needs to be provided to the researcher.	[Link]
4.4 File location *(Occ. 1-n)* File location of mapping table.	**4.4.1 File location** *(Occ. 1)* Link to the location where data files are stored.	[Link]
	4.4.2 File format *(Occ. 1)* Format in which the file can be retrieved from the specified location.	▪ **csv:** Character-separated values. ▪ **text:** Text format. ▪ **dta:** Stata data format.
4.5 Unit description *(Occ. 1-n)* Description of units in mapping table.	**4.5.1 Name of unit in mapping table** *(Occ. 1)* Name of the unit in the mapping table (e.g. banks, non-financial corporations, securities).	[Free text]
	4.5.2 Variable name of unit in mapping table *(Occ. 1-n)* Name of the variable which holds the unit.	[Free text]
4.6 Relation to other mapping table *(Occ. 0-n)*	**4.6.1 Relation type** *(Occ. 1)* Type of the relation between this mapping table and the other mapping table named below.	▪ **Is New Version Of:** Indicates that this mapping table here is a new version of the mapping table named below.

Describes relations to other mapping tables.		▪ **Is Part Of:** Indicates that this mapping table here is part of the mapping table named below. ▪ **Is Variant Of:** Indicates that this mapping table here is a variation of the mapping table named below.
	4.6.2 Related mapping table ID *(Occ. 1)* Unique ID of the other mapping table.	[Free text]

5. Global (i.e. irrespective of researcher affiliation, research field, and access mode) combining restrictions *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
5.1 Global restrictions on combining dataset families *(Occ. 0-n)* Global (i.e. irrespective of researcher affiliation, research field, and access mode) restrictions on the combining of two dataset families.	**5.1.1 Dataset family combination ID** *(Occ. 1)* Unique ID for the combination of dataset families A and B.	[Free text]
	5.1.2 ID of dataset family A *(Occ. 1)* Unique ID of dataset family A.	[Free text]
	5.1.3 ID of dataset family B *(Occ. 1)* Unique ID of dataset family B.	[Free text]
	5.1.4 Restriction type *(Occ. 1)* Type of combining restriction between dataset families A and B.	▪ **May not be combined with:** Indicates that dataset families A and B may not be used in the same research project. ▪ **May be combined with:** Indicates that dataset families A and B may be combined with each in the same research project.
5.2 Global restrictions on combining units *(Occ. 1-n)*	**5.2.1 Unit combination ID** *(Occ. 1)* ID for the combination of the specified units A and B.	[Free text]

Global (i.e. irrespective of researcher affiliation, research field, and access mode) restrictions on the combining of two variables.	**5.2.2 Name of unit A** *(Occ. 1)* Name of unit A.	[Free text]
	5.2.3 Variable name of unit A *(Occ. 1)* Name of the variable which holds unit A.	[Free text]
	5.2.4 Name of unit B *(Occ. 1)* Name of unit B.	[Free text]
	5.2.5 Variable name of unit B *(Occ. 1)* Name of the variable which holds unit B.	[Free text]
	5.2.6 Restriction type *(Occ. 1)* Type of combining restriction between the specified units A and B.	▪ **May not be combined with:** Indicates that the specified units A and B may not be used in the same research project. ▪ **May be combined with:** Indicates that the specified units A and B may be combined with each other in the same research project.

6. Global (i.e. applicable to all dataset families) rules *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
6.1 Access protocol family *(Occ. 1-n)* Group of access protocols.	**6.1.1 Access protocol family ID** *(Occ. 1)* Unique ID of the access protocol family.	[Free text]
	6.1.2 Name of access protocol family *(Occ. 1)* Name of the access protocol.	[Free text]
	6.1.3 Project status *(Occ. 1)* Status of the project to which the access protocol belongs. The respective access	▪ **0. Pre-application:** Researcher has contacted RDC but not (yet) submitted a valid project application.

	protocols assigned to a project status are necessary requirements that need to be satisfied in order to proceed to the next project status.	▪ **1. Assessment of application:** RDC has received a valid application and is in the process of assessing it, preparing data, and performing project-related administrative tasks. ▪ **2. Project approved and ready:** Researcher can start working with the data. ▪ **3. Project running:** Researcher is working on the project. ▪ **4. Project finished or rejected:** Researcher has confirmed that the project is finished and can be archived. In addition, this element indicates rejected projects. ▪ **5. Project archived:** Project has been archived and will be deleted in a pre-defined number of years.
	6.1.4 Ranking within project status *(Occ. 0-1)* Ranking of the access protocol family within the project status (where 1 is the highest rank).	[Free text]
	6.1.5 Access protocols *(Occ. 1)* Name of access protocols belonging to the access protocol family.	[Free text]
6.2 Decision rule *(Occ. 0-n)* Global decision rule that applies when two exclusive annodata (controlled field) items have to be evaluated (e.g. when requesting access to more than	**6.2.1 Annodata (controlled field) items** *(Occ. 1)* Annodata (controlled field) items for which the decision rule applies.	[Free text]
	6.2.2 Annodata element *(Occ. 1)*	[Free text]

one dataset family, a decision will be taken which annodata element of the two annodata items prevail). Example: **Annodata item** = Access mode **Annodata (controlled field) item** = Type of access mode **Annodata element** = On-site access	Annodata element for which the decision rule applies.	
	6.2.3 Decision rule *(Occ. 1)* Decision rule in relation to a different specification of the same annodata (controlled field) item.	▪ **Additive:** Annodata (controlled field) item will always be added to the set of applicable annodata (controlled field) items. ▪ **Mandatory:** Annodata (controlled field) item must have the same specification (the same element) for all requested dataset families. ▪ **Hierarchical:** Annodata element with the highest ranking will overwrite lower ranked elements included in the ranking group.
	6.2.4 Ranking between peers *(Occ. 0-1)* If hierarchical: Ranking of annodata elements (where 1 is the highest rank).	[Free text]
	6.2.5 Peer group *(Occ. 0-1)* If hierarchical: Name of peer group in which ranking is applied (e.g. contract type).	[Free text]

7. Researcher *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
7.1 Personal information *(Occ. 1)* Personal information on the researcher.	**7.1.1 Researcher ID** *(Occ. 1-n)* Researcher ID.	[Free text]
	7.1.2 Researcher ID type *(Occ. 1-n)* Type of the researcher ID.	[Free text]
	7.1.3 Surname *(Occ. 1)*	[Free text]

	Surname of the researcher.	
	7.1.4 First name(s) *(Occ. 1)* First name / given names of the researcher.	[Free text]
	7.1.5 Maiden name *(Occ. 1)* Maiden name of the researcher.	[Free text]
	7.1.6 Gender *(Occ. 1)* Gender of the researcher.	▪ **Female:** Female. ▪ **Male:** Male.
	7.1.7 Nationality *(Occ. 1)* Nationality of the researcher.	[Free text]
	7.1.8 Street, number *(Occ. 1)* Address (no P. O. box allowed).	[Free text]
	7.1.9 City *(Occ. 1)* City.	[Free text]
	7.1.10 Postal code *(Occ. 1)* Postal code, ZIP code.	[Free text]
	7.1.11 Country *(Occ. 1)* Country.	[Free text]
	7.1.12 Telephone / mobile phone *(Occ. 0-1)* Personal telephone / mobile phone.	[Free text]
	7.1.13 Date of birth *(Occ. 0-1)* Date of birth (DD-MM-YYYY).	[Free text]

	7.1.14 Place of birth *(Occ. 0-1)* Place of birth.	[Free text]
7.2 Professional affiliation *(Occ. 1)* Professional affiliation of the researcher.	**7.2.1 Scientific institution** *(Occ. 1)* Is researcher a staff member of, or affiliated with, a university, another higher education organisation, or an organisation or institution conducting scientific research?	▪ **Yes:** Yes. ▪ **No:** No.
	7.2.2 Researcher type *(Occ. 1)* Type of researcher for which this access mode is available.	▪ **Internal (limited rights)**: Internal researcher with limited rights (e.g. temporary, not-accredited). ▪ **Internal (extended rights)**: Internal researcher with extended rights (e.g. permanent, accredited). ▪ **External (limited rights)**: External researcher with limited rights (e.g. externals without internal co-author). ▪ **External (extended rights)**: External researcher with extended rights (e.g. externals with internal co-author).
	7.2.3 Job title *(Occ. 1)* Job title.	[Free text]
	7.2.4 Name of institution *(Occ. 1)* Name of institution.	[Free text]
	7.2.5 Department *(Occ. 0-1)* Department at institution.	[Free text]
	7.2.6 Email *(Occ. 1)* Work email address of the researcher.	[Free text]
	7.2.7 Telephone *(Occ. 0-1)*	[Free text]

	Work phone number of the researcher.	
	7.2.8 Fax *(Occ. 0-1)* Work fax number of the researcher.	[Free text]
	7.2.9 Street, number *(Occ. 1)* Address of the institution (no P. O. box allowed).	[Free text]
	7.2.10 City *(Occ. 1)* City.	[Free text]
	7.2.11 Postal code *(Occ. 1)* Postal code, ZIP code.	[Free text]
	7.2.12 Country *(Occ. 1)* Country.	[Free text]
	7.2.13 Website of institution *(Occ. 0-1)* Website of the institution.	[Link]
	7.2.14 Validation date *(Occ. 1)* Validation date of the above (DD-MM-YYYY).	[Free text]
7.3 Research projects *(Occ. 1)* Current and past research projects of researcher.	**7.3.1 Project ID** *(Occ. 0-n)* Project ID of current or past project.	[Free text]
8. Project *(Occ. 1-n)*		
Item	**Controlled field**	**Drop-down elements / Free text**
8.1 Identification *(Occ. 1)* Identification of the project.	**8.1.1 Project ID** *(Occ. 1)* Unique ID of the (planned) research project.	[Free text]

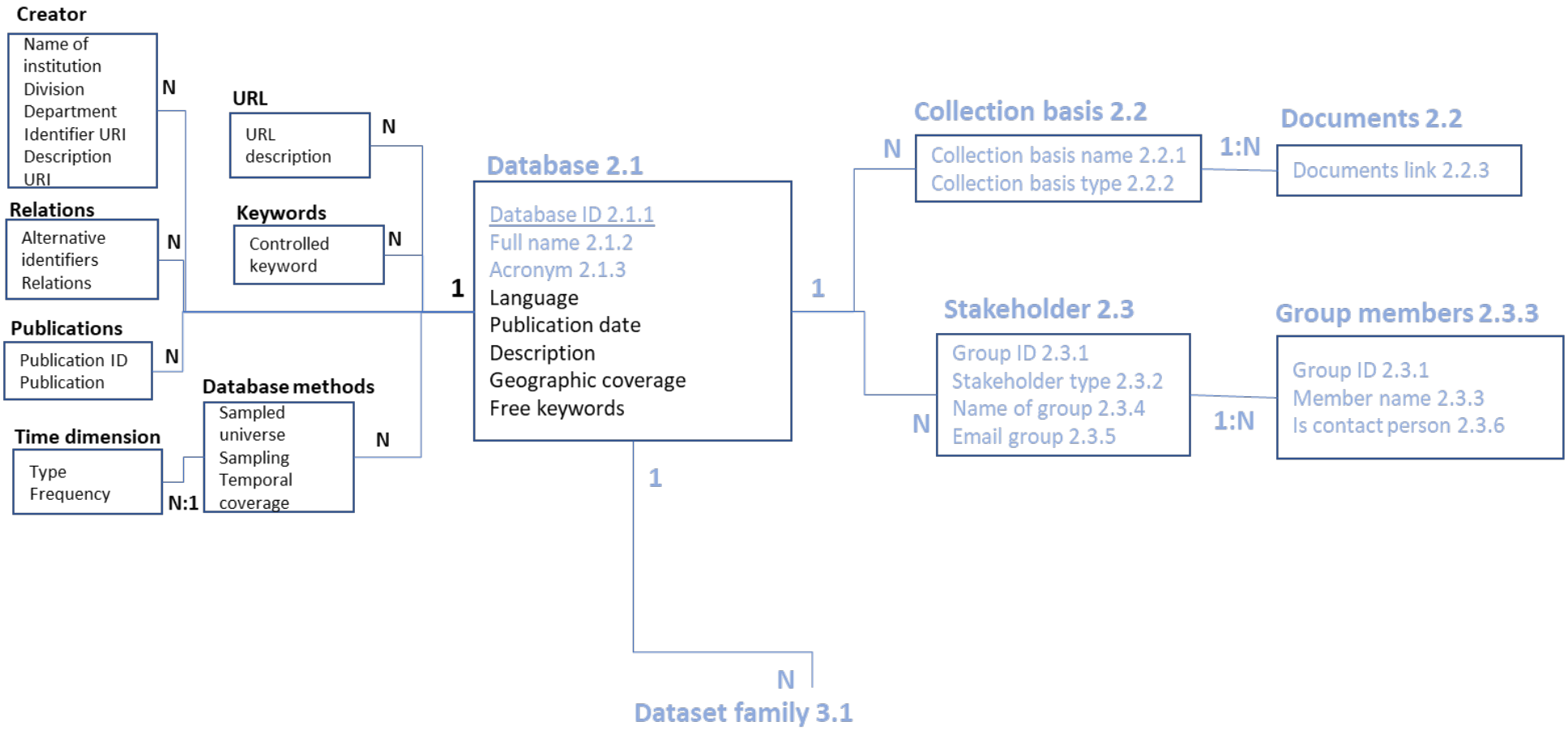
	8.1.2 Project name *(Occ. 1)* Name of the (planned) research project.	[Free text]
8.2 Participating researchers *(Occ. 1-n)* Information on participating researchers.	**8.2.1 Researcher ID** *(Occ. 1-n)* Researcher ID.	[Free text]
	8.2.2 Researcher ID type *(Occ. 1-n)* Type of the researcher ID.	[Free text]
	8.2.3 Lead researcher *(Occ. 1)* Is the researcher the project's lead researcher?	▪ **Yes:** Yes, the project's lead researcher. ▪ **No:** No, not the project's lead researcher.
	8.2.4 Data access *(Occ. 1)* Does the researcher have access to data?	▪ **Yes:** Yes, researcher has data access. ▪ **No:** No, researcher does not have data access.
	8.2.5 Date of induction to the project *(Occ. 1)* Date as of which the researcher is part of the project (DD-MM-YYYY).	[Free text]
	8.2.6 Date of leaving the project *(Occ. 1)* Date until which the researcher was or will be part of the project (DD-MM-YYYY).	[Free text]
8.3 Project description *(Occ. 1)* Description of the (planned) research project.	**8.3.1 Research question** *(Occ. 1)* Motivation of research question, ending by clearly stating the theoretical hypotheses. Explanation of applied methodology, choice of empirical method(s), including the strategy for identifying causal relationships.	[Free text]
	8.3.2 Research field *(Occ. 1)*	▪ **Any field:** Any field.

	Research fields for which the access mode is available.	▪ **Scientific research:** Scientific research. ▪ **Monetary policy:** Monetary policy. ▪ **Tasks of the ESCB:** Exercising ESCB tasks.
	8.3.3 JEL codes *(Occ. 1-n)* JEL codes.	[Free text]
	8.3.4 Keywords *(Occ. 1-n)* Keywords.	[Free text]
	8.3.5 Availability of results *(Occ. 1)* Researcher's intention to make the research results available to the scientific community (e.g. publications, conferences, working papers …).	[Free text]
8.4 Dataset families to be used *(Occ. 1-n)* Internal data used in the (planned) research project.	**8.4.1 Dataset family ID** *(Occ. 1)* Unique ID of dataset family used in the project.	[Free text]
	8.4.2 Type of dataset family *(Occ. 1)* Type of the dataset family.	▪ **Internal dataset family:** Internal dataset family. ▪ **External dataset family:** External dataset family.
8.5 Project organisation *(Occ. 1)* Organisational information on the (planned) research project.	**8.5.1 Responsible business area** *(Occ. 1)* Responsible business area.	[Free text]
	8.5.2 Link to researcher project folder *(Occ. 1)* Link to the project folder in which the researcher can work.	[Link]
	8.5.3 Link to data steward project folder *(Occ. 1)*	[Link]

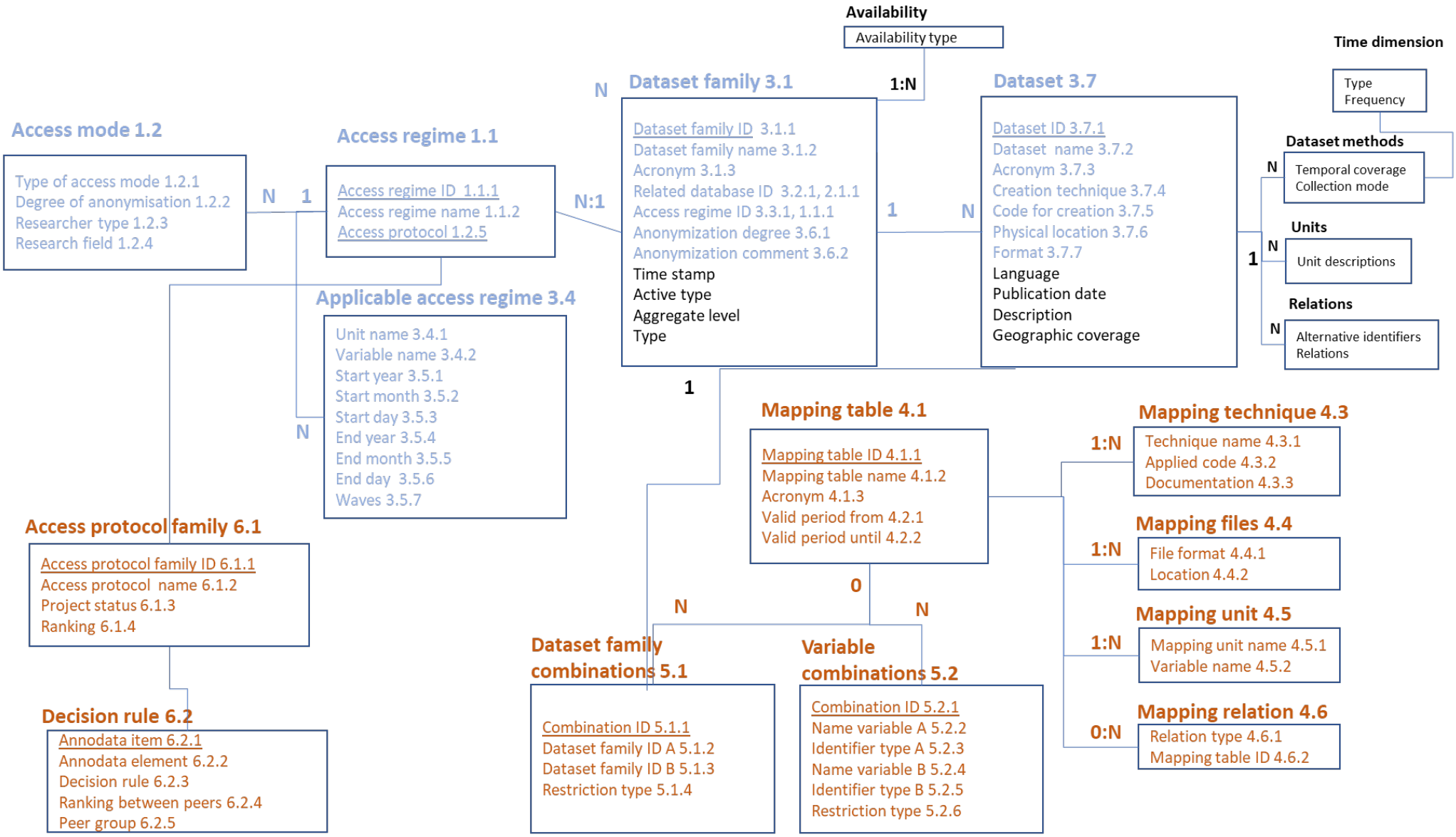
	Link to the folder where information on the project is stored (e.g. contracts, application, etc.) that should only be accessible by the data stewards.	
	8.5.4 Date of application *(Occ. 1)* Date on which the application was received (YYYY-MM-DD).	[Free text]
	8.5.5 Date of approval *(Occ. 0-1)* Date of project's approval (YYYY-MM-DD).	[Free text]
	8.5.6 Date of rejection *(Occ. 0-1)* Date of project's rejection (YYYY-MM-DD).	[Free text]
	8.5.7 Date of project start *(Occ. 0-1)* Start of project (YYYY-MM-DD).	[Free text]
	8.5.8 Date of project end *(Occ. 1)* Planned end of project (YYYY-MM-DD).	[Free text]
	8.5.9 Cost reimbursement *(Occ. 0-1)* Cost reimbursement for researchers.	▪ **No reimbursement:** No reimbursement. ▪ **Only reimbursement:** Only reimbursement. ▪ **Reimbursement and remuneration:** Reimbursement and remuneration.
8.6 Publications *(Occ. 0-n)* Publications from research project.	**8.6.1 Author** *(Occ. 1-n)* Name of the author (surname, first name(s)).	[Free text]
	8.6.2 Publication title *(Occ. 1)* Title of the publication.	[Free text]

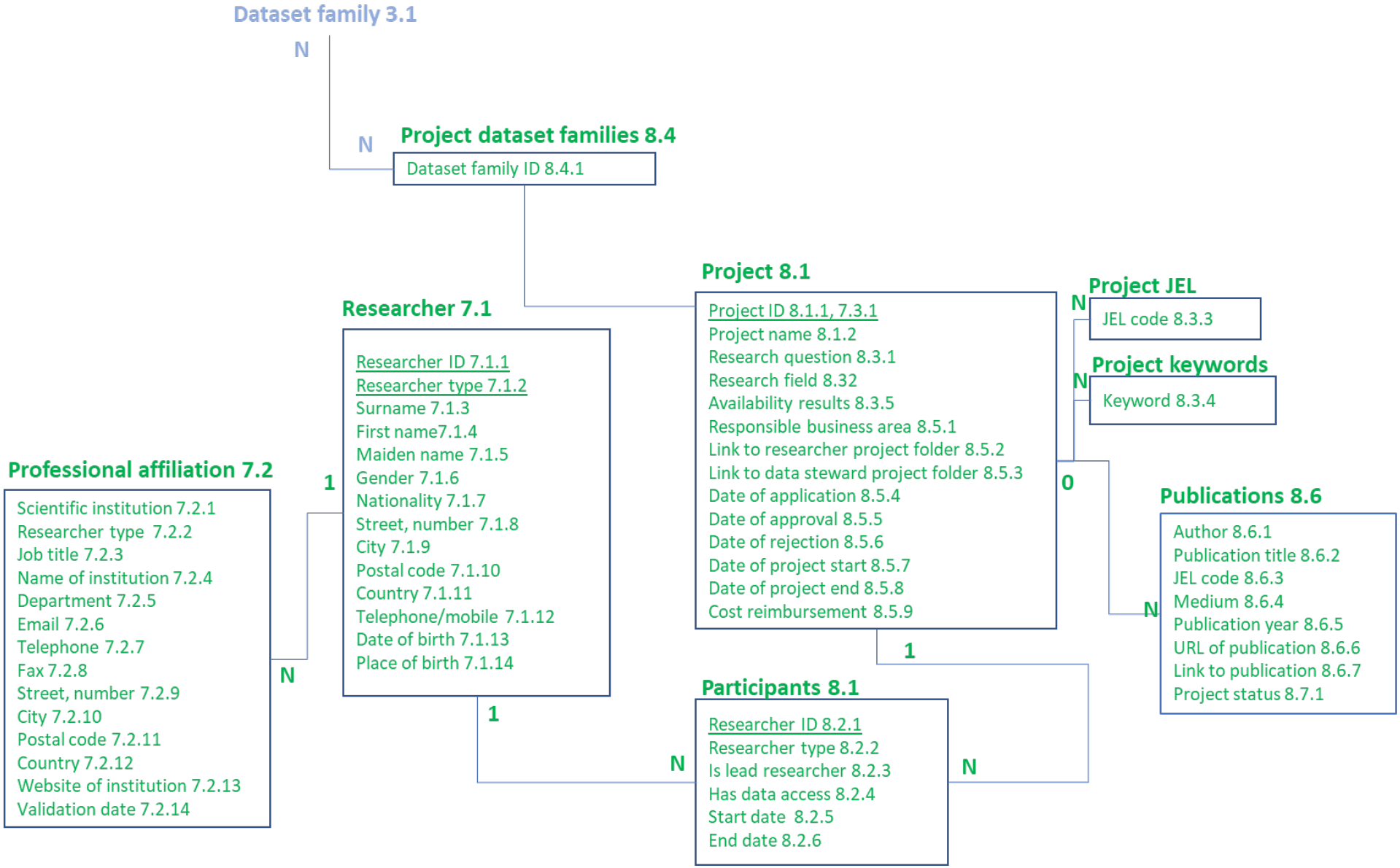
	8.6.3 JEL code *(Occ. 1-n)* JEL code(s) of the publication.	[Free text]
	8.6.4 Medium *(Occ. 1)* Medium in which the project results were published.	▪ **Bachelor thesis:** Bachelor thesis. ▪ **Book:** Book. ▪ **Code:** Code. ▪ **Economics blog:** Economics blog. ▪ **Journal:** Journal. ▪ **Master thesis:** Master thesis. ▪ **PhD thesis:** PhD thesis. ▪ **Presentation:** Presentation. ▪ **White paper:** White paper. ▪ **Working paper:** Working paper.
	8.6.5 Publication year *(Occ. 1)* Year of publication (YYYY-MM-DD).	[Free text]
	8.6.6 URL of publication *(Occ. 0-1)* DOI or link to publication on website.	[Free text]
	8.6.7 Link to publication *(Occ. 0-1)* Link to the physical location of publication (file).	[Free text]
8.7 Project status *(Occ. 0-1)* Current status of the project.	**8.7.1 Project status** *(Occ. 1)* Basic list of project status information that needs to be maintained manually.	▪ **Not yet active:** Not yet active. ▪ **Rejected:** Project has been rejected. ▪ **Active:** Project is active. ▪ **Finished:** Project is finished. ▪ **Archived:** Project has been archived.

C Entity relationship model of the INEXDA annodata schema[3]



[3] In black: Items referring to the INEXDA metadata schema.


Availability
Availability type
1:N
Time dimension
Type
Frequency
Dataset methods
Temporal coverage
Collection mode
Units
Unit descriptions
Relations
Alternative identifiers
Relations
N
1
Access mode 1.2
Type of access mode 1.2.1
Degree of anonymisation 1.2.2
Researcher type 1.2.3
Research field 1.2.4
N
1
Access regime 1.1
Access regime ID 1.1.1
Access regime name 1.1.2
Access protocol 1.2.5
N:1
N
Dataset family 3.1
Dataset family ID 3.1.1
Dataset family name 3.1.2
Acronym 3.1.3
Related database ID 3.2.1, 2.1.1
Access regime ID 3.3.1, 1.1.1
Anonymization degree 3.6.1
Anonymization comment 3.6.2
Time stamp
Active type
Aggregate level
Type
1
N
Dataset 3.7
Dataset ID 3.7.1
Dataset name 3.7.2
Acronym 3.7.3
Creation technique 3.7.4
Code for creation 3.7.5
Physical location 3.7.6
Format 3.7.7
Language
Publication date
Description
Geographic coverage
1
Applicable access regime 3.4
Unit name 3.4.1
Variable name 3.4.2
Start year 3.5.1
Start month 3.5.2
Start day 3.5.3
End year 3.5.4
End month 3.5.5
End day 3.5.6
Waves 3.5.7
N
Access protocol family 6.1
Access protocol family ID 6.1.1
Access protocol name 6.1.2
Project status 6.1.3
Ranking 6.1.4
Decision rule 6.2
Annodata item 6.2.1
Annodata element 6.2.2
Decision rule 6.2.3
Ranking between peers 6.2.4
Peer group 6.2.5
Mapping table 4.1
Mapping table ID 4.1.1
Mapping table name 4.1.2
Acronym 4.1.3
Valid period from 4.2.1
Valid period until 4.2.2
0
N
N
Dataset family combinations 5.1
Combination ID 5.1.1
Dataset family ID A 5.1.2
Dataset family ID B 5.1.3
Restriction type 5.1.4
Variable combinations 5.2
Combination ID 5.2.1
Name variable A 5.2.2
Identifier type A 5.2.3
Name variable B 5.2.4
Identifier type B 5.2.5
Restriction type 5.2.6
1:N
Mapping technique 4.3
Technique name 4.3.1
Applied code 4.3.2
Documentation 4.3.3
1:N
Mapping files 4.4
File format 4.4.1
Location 4.4.2
1:N
Mapping unit 4.5
Mapping unit name 4.5.1
Variable name 4.5.2
0:N
Mapping relation 4.6
Relation type 4.6.1
Mapping table ID 4.6.2


Dataset family 3.1
N
N
Project dataset families 8.4
Dataset family ID 8.4.1
Project 8.1
Project ID 8.1.1, 7.3.1
Project name 8.1.2
Research question 8.3.1
Research field 8.32
Availability results 8.3.5
Responsible business area 8.5.1
Link to researcher project folder 8.5.2
Link to data steward project folder 8.5.3
Date of application 8.5.4
Date of approval 8.5.5
Date of rejection 8.5.6
Date of project start 8.5.7
Date of project end 8.5.8
Cost reimbursement 8.5.9
N
Project JEL
JEL code 8.3.3
N
Project keywords
Keyword 8.3.4
0
N
Publications 8.6
Author 8.6.1
Publication title 8.6.2
JEL code 8.6.3
Medium 8.6.4
Publication year 8.6.5
URL of publication 8.6.6
Link to publication 8.6.7
Project status 8.7.1
Researcher 7.1
Researcher ID 7.1.1
Researcher type 7.1.2
Surname 7.1.3
First name7.1.4
Maiden name 7.1.5
Gender 7.1.6
Nationality 7.1.7
Street, number 7.1.8
City 7.1.9
Postal code 7.1.10
Country 7.1.11
Telephone/mobile 7.1.12
Date of birth 7.1.13
Place of birth 7.1.14
1
Professional affiliation 7.2
Scientific institution 7.2.1
Researcher type 7.2.2
Job title 7.2.3
Name of institution 7.2.4
Department 7.2.5
Email 7.2.6
Telephone 7.2.7
Fax 7.2.8
Street, number 7.2.9
City 7.2.10
Postal code 7.2.11
Country 7.2.12
Website of institution 7.2.13
Validation date 7.2.14
N
1
N
Participants 8.1
Researcher ID 8.2.1
Researcher type 8.2.2
Is lead researcher 8.2.3
Has data access 8.2.4
Start date 8.2.5
End date 8.2.6
1
N