Food and Agriculture Organization
of the United Nations

FAO Statistics Division (ESS)

New CountrySTAT
Guidelines

Revision: September 2016

Contents

How to use this document

This document represents the first operative guidelines to familiarize with the FENIX approach and technologies. It is primary addressed to the FENIX users, in particular to those users who not only access the FENIX platform to search, download and analyse a resource, but also to upload, edit and delete a resource in the FENIX platform.

This document is organized as follows:

- Section 1 provides an overview of CountrySTAT with a focus on its potentiality as an useful tool to browse, visualize and analyse data.

- Section 3 provides an operative and step by step guideline to manage data focusing on creating, editing and deleting data.

- Appendix A is a digression which describes in details the metadata editor providing users with the definition of each metadata field to compile.

Chapter 1

Introductory Considerations

1.1   Overview

CountrySTAT is a collection of software tools, methods and standards developed to facilitate acquisition, management and analysis of large and diversified data sets. CountrySTAT key features may be listed as follows:

- Based on a Open Source software, meaning no licensing constraints and full redistribution rights

- Service-Oriented Architecture (SOA)

- Data sharing through web services (APIs)

- Diversified data collection and upload methods

- Capability to handle different type of data

- Data discovery system based on metadata

- Advanced analytical capacities with the embedded **R** statistical package

- Expandable: possibility to plug-in new functionalities and applications

1.2   CountrySTAT approach

CountrySTAT wants to be a bridge between Information Technology and policy making. It offers a collection of IT tools to manipulate data in order to better understand the complexity of different social phenomena.

Since global food and agriculture challenges require fast, informed and targeted decisions, we need to create the conditions for evidence-based policies, in order to answer questions, such as:

- Where are hungry people?

- How much food is produced?

- How can agro-business become more profitable?

The fundamental requirements that allow us to answer those questions are:

- Mobilization of the national and international community

- High quality, comparable and timely data

However, unfortunately there is still a great way to go, because more often than not, provided data are affected by issues, such as: lack of data and metadata harmonization, poor data dissemination systems, restrictive or unclear data policy, gap between produced data and its availability.
This may lead to unpleasant consequences, such as: potentiality of data not fully exploited, data is not widely accessible or shared because the systems often work in isolation.

Therefore what we need to do is:

- Avail data as **public good**

- Ensure **interoperability** between IT systems

- Include expert-knowledge and users requirements in the IT system design

- Strengthen the international, regional and national institutional networks

This **data sharing network** is composed of three components: data policy, statistical governance and procedures and tools. These components are based on: institutional framework, data management and data availability. The data sharing network strength includes:

- Enabling of data exchange at all levels

- Creating a common technical language across data owners, statisticians, analysts, etc.

- Methods and tools to interpret and understand data, its reliability and its exact meaning (metadata)

- Enabling data harmonization and integration within and across countries and regions

Chapter 2

Data Analysis

Another tool implemented in the Fenix architecture, alongside the Data Management, is the **Search and Visualize** tool.
In the following paragraphs, the focus is therefore on the analysis of previously stored datasets.

2.1 Retrieving a previously stored dataset

In order to start the analysis process on a stored dataset, this dataset has to be retrieved.
This can be done by clicking on the "pulsating folder" button within the "Search and Visualize" section (as shown in figure 2.1)
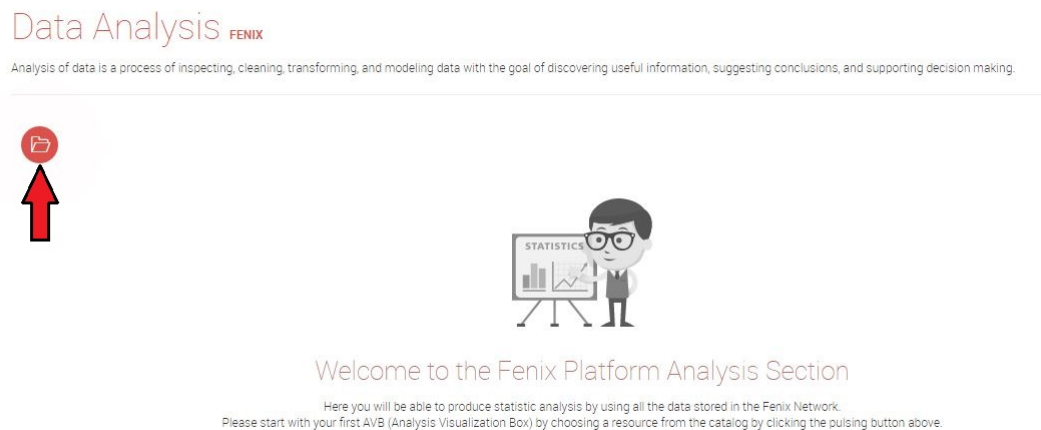

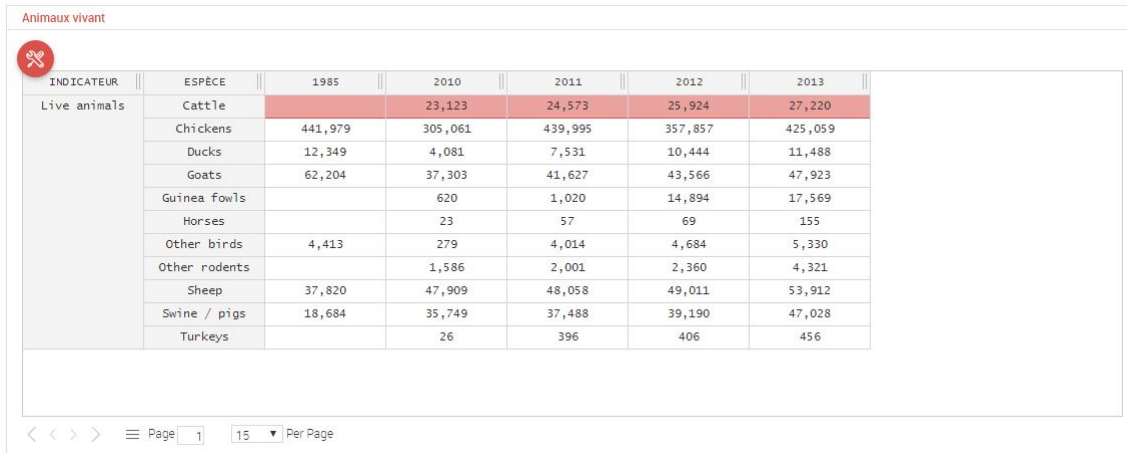Figure 2.1: How to select a resource in data analysis

By clicking on it, the catalogue will pop up (the same shown in figure 3.2).
This catalogue works exactly in the same way as the one used in the Data Management section.
Once all the filters have been imposed, a list of compliant resources, from which to choose, will be presented to the user.

2.2 Analysing datasets

Once the desired dataset has been chosen, the actual data, in table form, will be shown (as in figure 2.2).

Figure 2.2: Interface of the analysis tool

Let's analyse figure 2.2.

Moving the mouse on the table area, will make four buttons appear on the top right. In order, from left to right, they will:

- filter the resource (cogwheel);

- download the resource (arrow pointing down);

- duplicate the resource to facilitate a comparative analysis (entwined circles);

- close the current resource (circled "x").

On the top left area there are three more buttons which, in order, will:

- enter the resource settings (three horizontal lines), in which the user can:
    - download the entire resource, or its metadata;
    - visualize the resource through charts and maps;
    - reduce or enlarge the resource visualization box

- show the metadata linked to the displayed resource (meta);

- enter the visualization options (screwdriver & wrench tool)

Two or more resources can be analysed at once, and thus compared if needed, by clicking on the "folder" button right on top of the open resource (the very same shown in figure 2.1)

Chapter 3

Working with CountrySTAT

The CountrySTAT platform is built on the FENIX architecture to receive different types of resource. In the FENIX framework a resource is considered as an identifiable collection of information organized in a well-defined structure.

Depending on the kind of resource managed, different FENIX components have been built, in order to configure the FENIX environment to properly receive the resource in the DataBase. In the following paragraphs, the focus is on the "dataset" type of resource, which identifies a well-structured set of data organized in a data matrix.

The Data Management is the tool that allows the user to manage all the data-resources available in the FENIX web platform. It allows not only to maintain and update the resource already uploaded into the FENIX system, but also to upload new data-resources, helping the users with the procedures concerning the upload of the three different components that a FENIX "dataset" resource requires: data, metadata and Data Structure Definition (DSD).

3.1 The catalogue

The catalogue allows the user to retrieve a resource already uploaded into the FENIX system. It is one of the Data management's main tools and it is accessible just by clicking on the "Search" box (Figure 3.1).

The process to retrieve a resource in the DB is driven by the metadata information. That is why the menu on the left of the screen of the catalogue contains exactly the same entities of the metadata Editor (described more in details in the following, Appendix A).

By adding and filling the fields on the left of the catalogue, it is possible to select the criteria by which to filter the resources stored in the FENIX data base. Only the resources compliant with the imposed filter(s) will be retrieved (as shown in Figure 3.2).
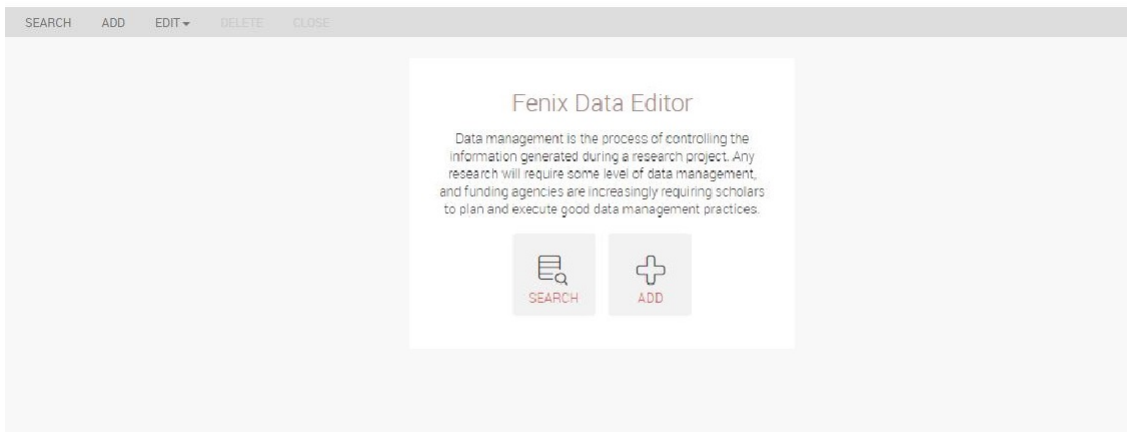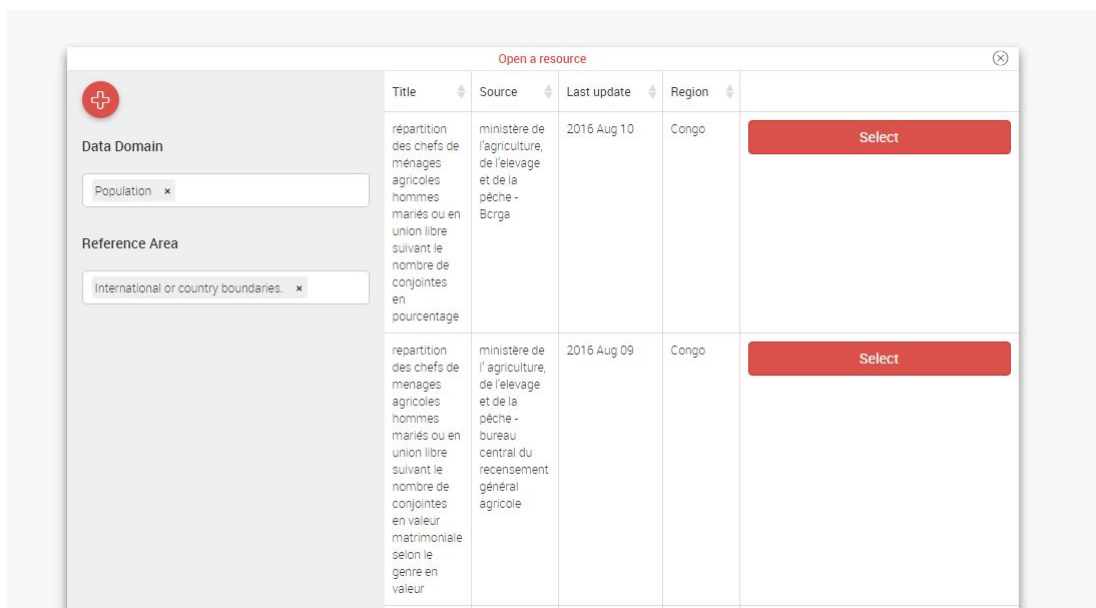
Figure 3.1: Data Management

Figure 3.2: Filtering the catalogue

9

3.2   Preliminary Operations

Three things need to be prepared and understood beforehand, in order to successfully upload a dataset onto the new CountrySTAT:

1. Have the metadata ready, as shown in the excel file accompanying this document;

2. Understand which codelist needs to be used in which situation. (Please refer to 3.2.1);

3. Understand the importance of attributes (Please refer to 3.2.2)

4. Properly prepare the dataset to be uploaded. (Please refer to 3.2.3)

3.2.1   Codelists

The codelists currently in use in the new CountrySTAT are the following:

- **CPC 2.1** for commodities and services;

- **HS** for trade data;

- **Forestry Products** for forest-related products;

- **Fishery products** for fishery-related products;

- **Scientific fishery products** a more in-depth fishery classification;

- **GAUL** for geo data;

- **CountrySTAT Indicators** for all the indicators;

- **UM** for all units of measurement

- **Flag** to indicate missing and estimated values

Central Product Classification

The Central Product Classification (CPC) constitutes a complete product classification covering all goods and services based on a set of internationally agreed concepts, definitions, principles and classification rules.
It provides a comprehensive framework within which data on products can be collected and presented in a format that allows for economic analysis supporting decision-taking and policy-making.
The current version, CPC version 2.1, is the result of a scheduled review of the CPC structure and detail, in order to ensure the classification?s relevance for describing current products in the economy.

Harmonized System

The Harmonized Commodity Description and Coding System generally referred to as "Harmonized System" or simply "HS" is a multipurpose international product nomenclature developed by the World Customs Organization (WCO).
The Harmonized Commodity Description and Coding System (HS) is an exhaustive nomenclature of internationally traded commodities (goods) classified according to the following criteria:

- raw or basic material

- degree of processing

- use or function

- economic activities

GAUL

The Global Administrative Unit Layers (GAUL) is an initiative implemented by FAO within the Bill & Melinda Gates Foundation, Agricultural Market Information System (AMIS) and AfricaFertilizer.org projects.

The GAUL compiles and disseminates the best available information on administrative units for all the countries in the world, providing a contribution to the standardization of the spatial dataset representing administrative units. The GAUL always maintains global layers with a unified coding system at country, first (e.g. departments) and second administrative levels (e.g. districts). Where data is available, it provides layers on a country by country basis down to third, fourth and lowers levels.

The overall methodology consists in:

- collecting the best available data from most reliable sources;

- establishing validation periods of the geographic features (when possible);

- adding selected data to the global layer based on the last country boundaries map provided by the UN Cartographic Unit (UNCS);

- generating codes using GAUL Coding System;

- distribute data to the users

Because GAUL works at global level, unsettled territories are reported. The approach of GAUL is to deal with these areas in such a way to preserve national integrity for all disputing countries .

GAUL is released once a year and the target beneficiary of GAUL data is the UN community and other authorized international and national partners. Data might not be officially validated by authoritative national sources and cannot be distributed to the general public. A disclaimer should always accompany any use of GAUL data.

GAUL keeps track of administrative units that has been changed, added or dismissed in the past for political causes. Changes implemented in different years are recorded in GAUL on different layers. For this reason the GAUL product is not a single layer but a group of layers, named "GAUL Set".

3.2.2 Attributes

An attribute is a characteristic, most of the times dichotomous, embedded in the indicator it accompanies. It serves the purpose of better explaining what the indicator is about.

Examples of attributes include:

- Male, Female

- Rural, Urban

- Agricultural, Non-agricultural

- and so on

3.2.3 The correct way to prepare a dataset

The dataset has to be properly formatted in order to be uploaded onto the FENIX web platform:

- The Excel spreadsheet must be **formatted as text**;

- The name of the Excel file **must not** contain any accent and/or symbol and/or numbers and/or empty spaces

- Each column containing labels **must** be deleted. The appropriate codelist will link the coded columns to their rightful label

- Always convert the spreadsheet to a **CSV file** at the very end

- **Do not** reopen the CSV file afterwards

3.2.4 Important rules for creating tables

Here below is a list of important rules to follow in order to create a properly formatted table:

- **Never** use a comma or a dot for separating the thousands.
  e.g. one thousand must be written as 1000;

- **Always** use a dot for separating decimals.
  e.g. three dot four must be written as 3.4;

- **Always** start the table from the cell A1 (upper-left corner);

- **Never** use the "merge" function;

- **Never** write any text out of the table;

- **Never** have more than one sheet within the same Excel file;

- **Never** have columns with the same name in the same table;

- **Never** have more than one table in the same Excel file

Here below are examples of correctly formatted tables will be provided:

Figure 3.3: This is how a table should look like prior to column deleting

Figure 3.4: This is the same table, properly formatted

3.3 Add a new resource

By clicking on "Add" on the home page of the "Data management" section it is possible to upload a new dataset in the FENIX database. The procedure requires to input three different components of the resource in order to ensure the compliance with all the FENIX tools and to facilitate exchange and sharing of data:

1. **Metadata** extensively documented in (Section 3.4). In addition a wide and detailed explanation of all metadata fields is provided by Appendix A.

2. **Data Structure Definition** extensively documented in (Section 3.5).

3. **Data**, principal component of the resource, it has to be uploaded through a properly formatted CSV file.

3.4 Metadata Editor

By clicking on "Add" on the home page of the "Data management" section, you will be prompted to the metadata editor tool. The user will be requested to fill several metadata fields divided according to their competence area.
In addition, the **Copy Metadata** function, in the upper right part of the metadata editor, allows the user to copy all the metadata information already uploaded in another resource in the FENIX platform via ID number.
For further details, refer to Appendix A.

Figure 3.5: Metadata Editor

3.5 DSD Editor

This section describes the Data Structure Definition (DSD), required to upload a new data-resource in the FENIX platform.

A data matrix is composed of rows and columns. The Data Structure Definition Editor allows users to build a new data matrix in order to make the FENIX environment ready to receive the data: it basically consists in creating the same structure as the one of the table to be uploaded (analysed more in detail in the following paragraphs). In particular, the Figure 3.6 shows how the DSD editor looks like while providing a summary of the requested information.
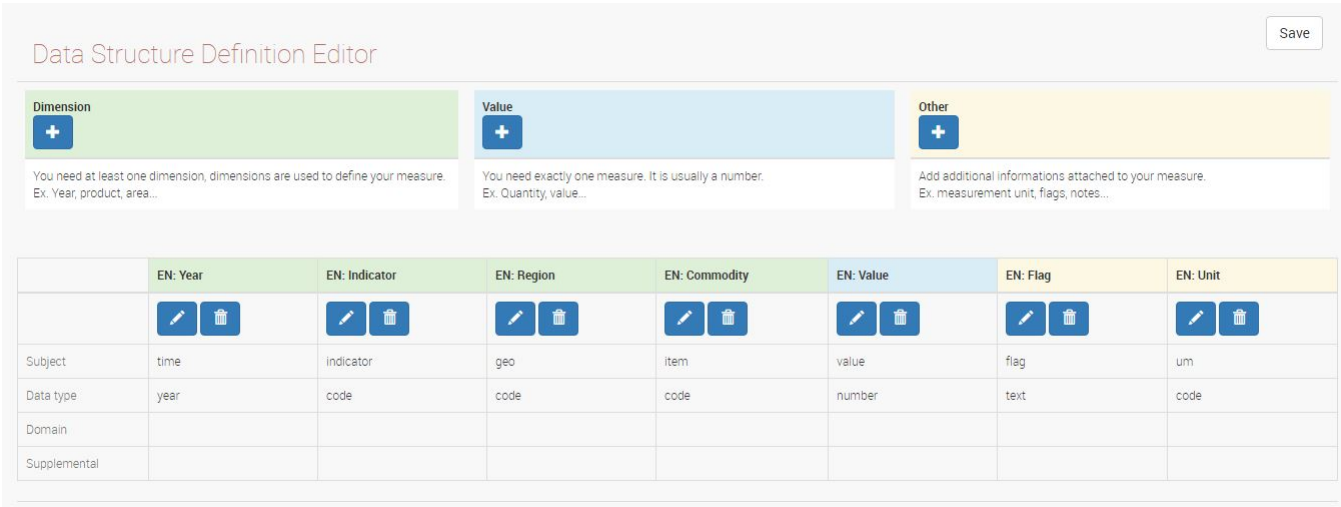
Figure 3.6: DSD Editor

The interface accomplishes this process through the definition of a new data structure: the user adds the columns one by one, specifying their content and their role.

The result of this process consists in building an empty grid that will be manually filled by the user with the actual data, through the upload of the table.

3.5.1 DSD mandatory elements

This section provides details on the requested element to describe each column of the data matrix. Following the template provided by the DSD Editor Interface, the following subsections explain, one by one, the attributes to be specified, in order to characterize each dimension. In this context, columns are often defined as dimensions. A dimension is a collection of reference information about a measurable event. Events are known as "facts" and dimensions categorize and describe data-related facts and measures. An example of dimension is the list of countries the data refers to.

Title

The title of the column represents the label displayed in the data grid. The Title manually typed by the user does not need to correspond to the one in the CSV file.

Subject

FENIX dimensions have been built in order to be sufficiently general and receive a wide variety of data. FENIX application offers a predefined list of values used to classify different type of dimensions according to their semantic value. This choice has several implications because different "Subjects" are associated with different Data types. It also has additional constrains in the following steps.
The current subjects are:

1. Item

2. Indicator

3. Gender

4. Residence

5. Food

6. Sector

7. Field management

8. Agricultural

9. Time

10. Geo

Data types

After selecting the appropriate Subject for the column, the option to select the **Data type** might be presented.
If data in the column is coded, "code" must be selected. The same logic applies for "year", "month", "date" and "number"

3.6 Data uploading

Once the dataset has been properly formatted following the procedure described just above, it's ready to be effectively uploaded. This can be done by clicking on the "Choose File" button (as shown in figure 3.7).
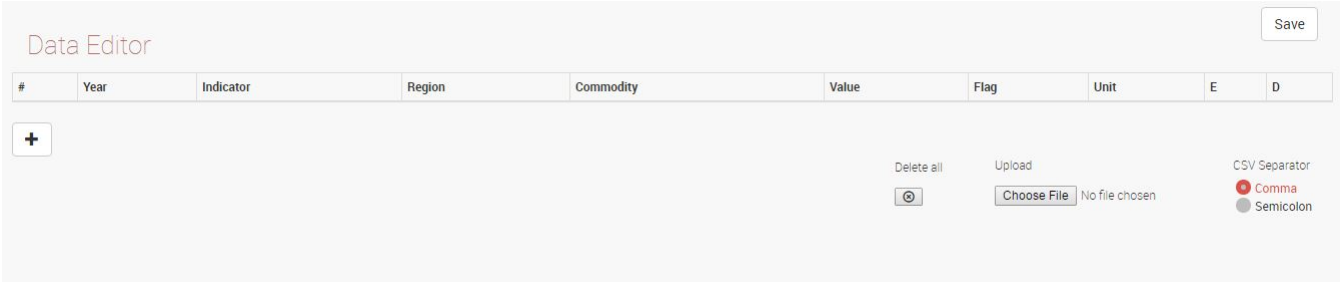
Figure 3.7: Data editor

After the file has been chosen, the system will prompt you to align the columns of the CSV with the ones created via DSD editor. Once done simply click OK.
If everything works properly the uploaded data will be visible on screen, otherwise a message error will appear. Once the dataset is uploaded it must be stored onto the system by clicking on the "save" icon on the upper right part of the page (as shown in figure 3.8).

Figure 3.8: Saving the uploaded dataset

3.7 Operation on already existing resources

Once a new resource has been identified and stored as "current resource", several operations can be undertaken: first of all it is possible to update its components (data and metadata, Figure 3.5), it is then possible to delete the whole data set if the resource is no longer meaningful.

3.7.1 How to modify or delete a single row

By clicking on the "pencil" tool on the right-hand side of any row a pop up will appear allowing the selected row to be modified (as shown in figure 3.9). Each of the specified row's entries can be freely modified. Once all changes have been completed, the new entries must be saved by clicking the "tick" symbol. If the modification of the row is no longer intended, the "x" button must be clicked instead.
By clicking on the "waste bin" symbol on the right-hand side of any row, the selected row can be deleted.
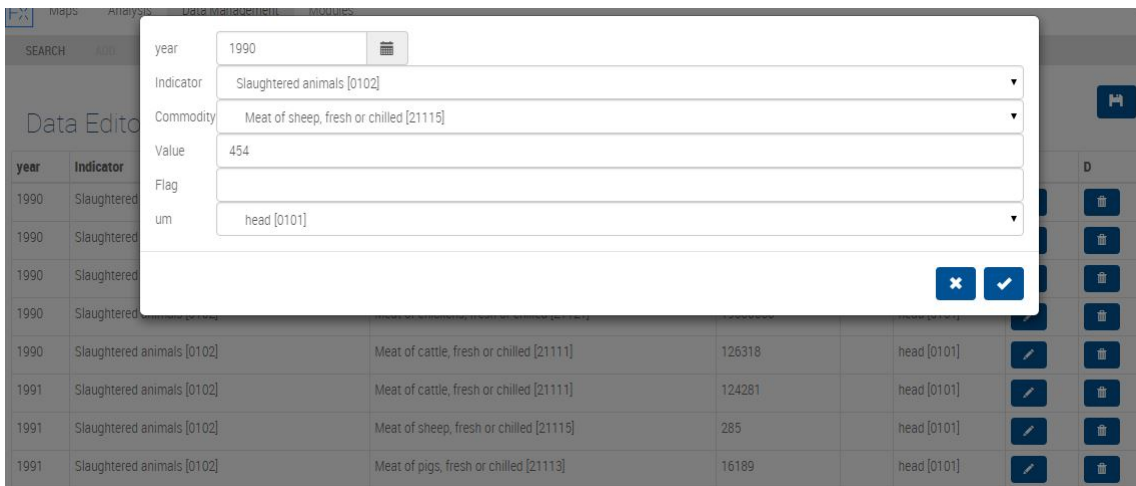
Figure 3.9: Modifying an already existing row

3.7.2 How to delete the entire dataset

By clicking on the "Delete all" symbol at the bottom of the page the whole dataset will be deleted.
However this operation will leave all the medatada and the DSD structure untouched.
By deleting the whole dataset an updated one can easily be re-uploaded without having to create a new resource.

3.7.3 How to add a new row

By clicking of the "+" symbol on the bottom left of the page, a pop up will appear allowing the user to enter a new row (as shown in figure 3.10). The structure of this new row is the same as all the other rows already present in the dataset.

Figure 3.10: Adding a new row

Appendix A

Detailed description of the metadata entities and elements

This section provides details on the definitions, structures and properties of each metadata entity and element. It also describes the relationships between metadata entities. Each element description includes the following information:

- The name of the element, delimited by the symbol $<< \cdots >>$

- The description of the element

- Supplementary information within brackets:

  - The element's obligation/condition status
  - The data type.

- The extended name ("friendly name") of the metadata element (in bold).

A.0.4 Identification Information (ME_Identification)

<<title>>

Textual label used as title of the resource.

(mandatory - data type: free text)
Title

<<creationDate>>

Creation date of the resource.

(mandatory - data type: date)
Creation date

<<language>>

Language used by the resource for textual information. This field is addressed to highlight some particular inconsistencies in the language (or languages) used in the resource, if any. For example to alert that the resource is not completely homogeneous in the language used for textual information. Otherwise it can be leaved empty.

(mandatory - data type: OJ_CodeList)
Language(s)

<<languageDetails>>

Comments and additional details about the language used for the textual information of the resource.

(optional - data type: free text)
Language details

<<characterSet>>

Full name of the character coding standard used by the resource.

(mandatory - data type: OJ_CodeList)
Character-set

<<Used metadata standard>>

Name of the metadata standard specifications used. In FENIX framework this field would be pre-compiled by "*FENIX*".

(mandatory - data type: free text - No multilingual)
Used metadata standard

<<Version of metadata standard>>

Version of the metadata standard specifications used.

(option - data type: free text - No multilingual)
Version of metadata standard

<<Language(s) used for metadata>>

Language(s) used to compile the metadata record.

(mandatory - data type: OJ_CodeList)
Language(s) used for metadata

<<Value assigned to No-data>>

Value assigned to the cells to represent the absence of data. Missing values are usually highlight through apposite flags, however the data matrix does not report empty cells but a predefined combination of characters (such as "NA", "000" ...) indicating the absence of data.

(optional - data type: free text - No multilingual)
Value assigned to No-data

ME_Identification (M)

+uid: free text (M)
+version: free text (O)
+language: OJ_CodeList (M)
+languageDetails: free text (O)
+title: free text (M)
+creationDate: date (M)
+characterSet: OJ_CodeList (M)
+parentIdentifiers: sequence<free text> (O)
+metadataStandardName: free text (M)
+metadataStandardVersion: free text (O)
+metadataLanguage: OJ_Codelist (M)
+contacts: sequence<OJ_ResponsibleParty> (M)
+noDataValue: free text (O)

Figure A.1:

A.0.5 Contacts Information (ME_)

<<organization>>

Name of the responsible organization.

(mandatory - data type: free text)
Organization

<<organizationUnit>>

Addressable subdivision of an organization.

(optional - data type: free text)
Organization unit/division

<<pointOfContact>>

Responsible person-surname, given name, title separated by a delimiter. It contains information about the party who can be contacted for acquiring knowledge the resource.

(optional - data type: free text)
Point of contact

<<position>>

Role or position of the responsible person.

(optional - data type: free text)
Position

<<role>>

Function performed by the responsible party concerning the resource (e.g. owner, distributor, user...)

(optional - data type: multipleChoice[FENIX/tech/MD/02, **??**])
Role

<<specify>>

Textual metadata element that allows to specify the role performed by the responsible party. This field is conditional to the element <role>.

(conditional - data type: free text)
Specify
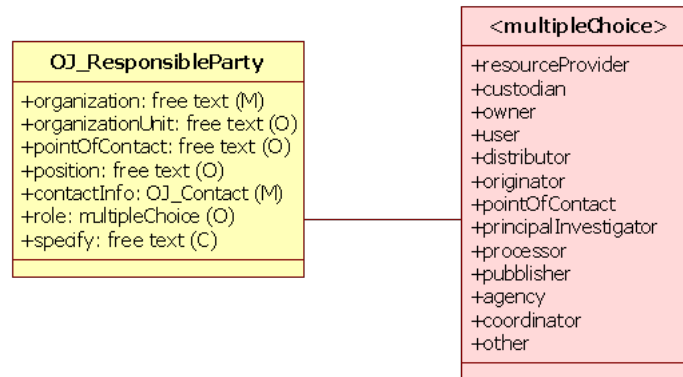
Figure A.2: Object - OJ_ResponsibleParty

<<**Contact information**>> The Object OJ_Contact mandatory requires at least one of the three conditional elements (phone/ address/ emailAddress).

<<**phone**>>

Telephone numbers at which the organization or individual may be contacted.

<div align="right">(conditional - data type: free text - No multilingual)
Telephone number</div>

<<**address**>>

Physical address at which the organization or individual may be contacted.

<div align="right">(conditional - data type: free text - No multilingual
Address</div>

<<**emailAddress**>>

E-mail address at which the organization or individual may be contacted.

<div align="right">(conditional - data type: free text - No multilingual
e-mail address</div>

<<**hoursOfService**>>

Time period (including time zone) when individuals can contact the organization or individual.

<div align="right">(optional - data type: free text)
Hour of service</div>

<<**contactInstruction**>>

Supplemental instructions on how or when to contact the individual or organization.

<div align="right">(optional - data type: free text)
Instruction</div>

OJ_Contact
+phone: free text (C) +address: free text (C) +emailAddress: free text (C) +hoursOfService: free text (O) +contactInstruction: free text (O)

Figure A.3: Object - OJ_Contact

A.0.6   Content (ME_Content)

<<Keywords>>

Commonly used word(s), formalized word(s) or phrase(s) used to describe the resource.

<div align="right">(mandatory - data type: sequence <free text - No multilingual>)
Keywords</div>

<<Abstract>>

Overview of the main characteristics of the resource and summary of the information contained in the resource, in an easily understandable manner, for technical and non-technical users.

<div align="right">(mandatory - data type: free text)
Abstract</div>

<<statisticalConceptsDefinitions>>

Definitions of the statistical concepts under measure (i.e. the statistical domain) and the main variables provided. The considered types of variables (e.g. raw figures, annual growth rates, index, flow or stock data, ...) should be defined and described in accordance with internationally accepted statistical standards, guidelines, or good practices.

<div align="right">(optional - data type: free text)
Statistical concepts / definitions</div>

<<SE_ReferencePopulation>>   Information on the statistical population the resource refers to. This sub-entity includes:

<<statisticalPopulation>>

Target statistical population (one or more) the resource refers to.

<div align="right">(mandatory - data type: free text)
Statistical population</div>

<<statisticalUnit>>

Simplest unit for which information is sought and for which statistics are ultimately compiled.

<div align="right">(optional - data type: free text)
Statistical unit</div>

<<referencePeriod>>

Specific time periods (e.g. a day, a week, a month, a fiscal year, a calendar year or several calendar years) the statistical variables refer to.

<div align="right">(mandatory - data type: OJ_CodeList)
Period of reference</div>

<<referenceArea>>

Type of geographical units the resource represents or refers to. Note that the spatial resolution must refer to the minimum mapping unit whose bounds are officially recognized indipendently from the measurement process of the phonomenon taken into account. Examples are: countries, administrative level 2, etc.

<div align="right">(mandatory - data type: OJ_CodeList))
Area of reference</div>

<<**SE_Coverage**>> Size and extent of the resource. The term "coverage" encompasses the descriptions of key dimensions delimiting the data, e.g. geographical, institutional, product, economic sector, etc., as well as relevant exceptions and exclusions.

<<**CoverageSectors**>>

Sector(s) the resource refers to as specified in the selected codelist. The word "Sector" indicates the subject area the resource refers to. These sectors can be institutional sectors, economic or other sectors (e.g. local government sector, agriculture, forestry, business services, etc.).

<div align="right">

(mandatory - data type: sequence<OJ_CodeList>)
Main sector(s) (coded)
</div>

<<**coverageSectorsDetails**>>

Textual element delimiting the statistical results with regard to the main sectors covered.

<div align="right">

(optional - data type: sequence<free text)
Main sector(s)
</div>

<<**coverageTime**>>

Information about the time period for which data are available. It requests to report the time window of reference (reporting the starting date and the ending date) even if it presents some lacks.

<div align="right">

(mandatory - data type: period)
Coverage period
</div>

<<**coverageGeographic**>>

Geographical coverage represented by the resource. It is highly recommended to make reference to officially recognized or easily identifiable macro-areas (e.g. South Saharan Africa, North America, OECD member countries..). The attributes "total" or "partial" may help specifying whether the resource refers to the whole cited area or just a part of it.

<div align="right">

(mandatory - data type: OJ_CodeList)
Geographic extent
</div>

A.0.7   Institutional mandate (ME_InstitutionalMandate)

<<**legalActsAgreements**>>

References (citations or website link) to legal acts or other formal or informal agreements that assign responsibility as well as authority to an agency for the collection, processing, and dissemination of the resource.

<div align="right">

(optional - data type: free text)
Legal acts/agreements
</div>

<<**institutionalMandateDataSharing**>>

References (citations or website link) to arrangements or procedures for data sharing and coordination.

<div align="right">

(optional - data type: free text)
Data sharing arrangements
</div>
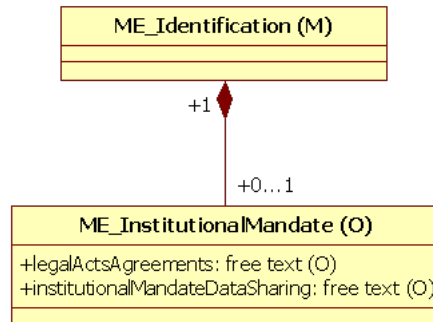
Figure A.4:

A.0.8 Processing (ME_Processing)

SE_StatisticalProcessing contains the data source and a set of elements that allow reporting methodological notes and detailed descriptions of the statistical process.

<<**SE_DataSource**>> process used to collect data. It includes a detailed description both of the primary data collection (e.g. type of collection, method to gather data from respondents, sampling procedures..) and the secondary data collection (information about data that have been already collected by another agency or institution).

<<**typeOfCollection**>>

Procedures used to collect data obtained directly from first-hand sources by means of surveys, observation or experimentation. In consists of a coded element which specifies the type of data collection method (e.g. census, random sampling, etc.).

(optional - data type: OJ_CodeList)
Type of collection

<<**samplingProcedure**>>

The type of sample design used to select the survey respondents to represent the population. It may refer to information on sample design, sample size, sample frame, sample updating etc.

(optional - data type: free text)
Sampling procedure

<<**dataCollection**>>

Methods used to gather data from the respondents (e.g. postal survey, CAPI, on-line survey, face-to-face interviews etc.) and description of data collection methods. This metadata element also includes more precise information about the kind of questionnaire (structured, unstructured etc.) and if necessary some noteworthy aspects of the data collection process.

(mandatory - data type: free text)
Data collection

<<collectionPeriodicity>>

Frequency with which the data are collected from the sources.

<div align="right">

(mandatory - data type: OJ_CodeList)
Periodicity of data collection

</div>

<<organization>>

Please refer to A.0.5 for clarifications

<div align="right">

(conditional - data type: OJ_CodeList)
Organization

</div>

<<originOfCollectedData>>

Coded element which allows to specify in a standard way the origin of the resource. It is filled when the agency compiling and publishing data does not coincide with the entity (subject, agency or institution) who has conducted the procedure of collecting data, otherwise it is assumed that the data source coincides with the responsible party already declared in ME_Identification.

<div align="right">

(optional - data type: OJ_CodeList)
Origin of collected data

</div>

<<rawDataDescription>>

Characteristics and components of the raw statistical data used for compiling statistical aggregates. It indicates if data set is based on a survey or on administrative data source. If administrative registers are used, the description of registers should be given (source, year, primary purpose, potential deficiencies. . .).

<div align="right">

(optional - data type: free text)
Description of raw data

</div>

<<dataCollection>>

Please refer to A.0.8 for clarifications

SE_DataCompilation   describes the main statistics actions operated on data (e.g. data editing, imputation, weighting, adjustment for non-response, model used etc.).

<<missingData>>

It describe under which circumstance missing data are estimated or imputed and when the cells are left empty. It also describe methodologies used to estimate/impute missing values.

<div align="right">

(optional - data type: free text)
Missing data

</div>

<<weights>>

Description of weights system (if any) used in order to produce accurate statistical results. This field reports the criteria for using weights in analysis of collection, the formulas and coefficients developed and how they are applied to data.

<div align="right">

(optional - data type: free text)
Weights

</div>

<<aggregationProcessing>>

Information about the methodology used to aggregate data.

<div align="right">(optional - data type: free text)
Process of aggregation</div>

<<aggregationFormula>>

Formula used to aggregate data.

<div align="right">(optional - data type: free text)
Aggregation formula</div>

<<dataAdjustment>>

Type of adjustment used represented by a code.

<div align="right">(optional - data type: OJ_CodeList)
Process of adjustment</div>

<<dataAdjustmentDetails>>

Set of procedures employed to modify statistical data to enable it to be conform with national or international standards (such as seasonal adjustment methods, time series decomposition, or other similar methods).

<div align="right">(optional - data type: free text)
Details on process of adjustment</div>

<<indexType>>

Type of index number used in the statistical production process.

<div align="right">(optional - data type: free text)
Type of index</div>

<<basePeriod>>

Period of time used as a base of an index number or to which a time series refers (e.g. base year 2000 for certain annual data).

<div align="right">(optional - data type: date)
Base period</div>

SE_DataValidation   methods and processes for assessing statistical data. It describes the process of monitoring data compilation progress and ensuring the quality of the statistical results.

<<dataValidationIntermediate>>

Assessment of the quality and correctness of intermediate calculations leading to statistical outputs.

<div align="right">(optional - data type: free text)
Data validation - intermediate</div>

<<dataValidationOutput>>

Assessment of discrepancies and/or inaccuracies observed in the statistical outputs.

<div align="right">

(optional - data type: free text)
Data validation - output

</div>

<<dataValidationSource>>

Assessment of discrepancies and/or inaccuracies inherent to the data source.

<div align="right">

(optional - data type: free text)
Data validation - source

</div>
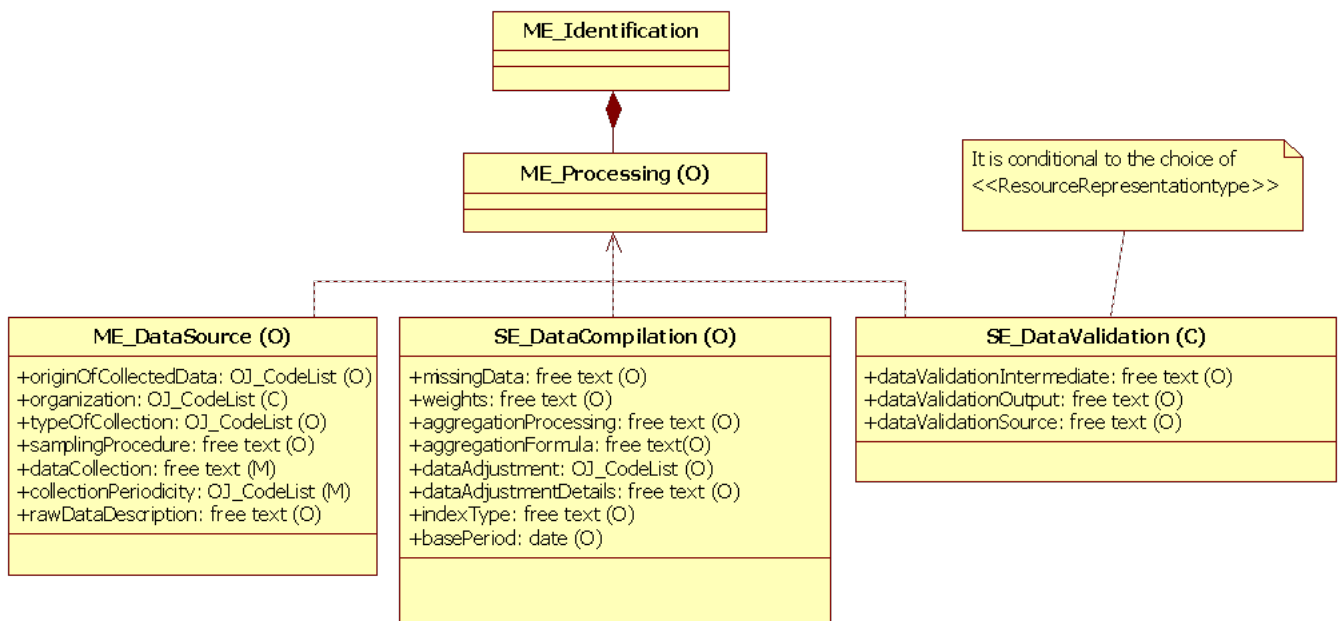
Figure A.5: Data quality

A.0.9   Data Quality (ME_DataQuality)

This section provides a description and evaluation of the data quality. It allows to describe the data quality assurance process, inclusive of data validation, completeness and accuracy standards. In addition an assessment of the comparability and intern coherence of the resource is considered a quality dimension.

<<qualityManagement>>

Structure, responsibilities and procedures established for guaranteeing the quality of the data.

<div align="right">

(optional - data type: free text)
Quality management

</div>

<<qualityAssessment>>

Overall qualitative assessment of the quality of the statistical outputs.

<div align="right">

(optional - data type: free text)
Data quality assessment
</div>

<<qualityAssurance>>

Description of the process assuring that the data production processes conforms to the statistical quality standards.

<div align="right">

(optional - data type: free text)
Quality assurance
</div>

<<SE_Accuracy>> closeness of computations or estimates to the exact values that the statistics were intended to measure. Accuracy can contain either measures of numerical results of the methods for assessing the accuracy of data or qualitative assessment indicators. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g. sampling, non-response, response error).

<<accuracyNonSampling>>

Error in sample estimates which cannot be attributed to sampling fluctuations. (e.g. defects in the sampling frame, faulty demarcation of sample units, defects in the selection of sample units, mistakes in the collection of data due to personal variations, misunderstanding, bias, negligence ... etc.)

<div align="right">

(optional - data type: free text)
Accuracy - non sampling error
</div>

<<accuracySampling>>

If probability sampling is used, the accuracy is an evaluation of difference between a population value and an estimate thereof, derived from a random sample (so due to the fact that only a subset of the population is enumerate), normally in the form of coefficient of variation, standard error or confidence intervals. For non-probability sampling, random errors cannot be calculated without reference to some kind of model, in this case estimates of the accuracy, a motivation for the invoked model for this estimation and brief discussion of sampling bias should be provided.

<div align="right">

(optional - data type: free text)
Accuracy - sampling error
</div>

SE_DataRevision This section describes the policy and practice for identifying the revision status of the data, as well as the availability of revision studies and analysis.

<<revisionPolicy>>

Policy concerning the periodically revision of the resource and ensuring the transparency of disseminated data.

<div align="right">

(optional - data type: free text)
Revision policy
</div>

<<revisionPractice>>

Information concerning the revision of data in order to give compilers the possibility of incorporating new and more accurate information in the resource. It also describes the revision status of available data. Data may also be subject to regular or ad hoc revisions as a result of the introduction of new classification, compilation frameworks and methodologies in order to improve the accuracy of the resource.

(optional - data type: free text)
Revision practice

<<SE_Relevance>> evaluation of data-quality through user satisfaction involving also information about the resource-completeness.

<<userNeeds>>

Classification of users with some indication of their importance, an indication of the uses for which they want the statistical outputs and as well users and uses given special considerations. Unmet user needs and the reasons for not meeting them should be included as well.

(optional - data type: free text)
User needs

<<userSatisfaction>>

How the views and opinions of the users are collected. In addition the main results regarding the user satisfaction should be shown (in the form of a user satisfaction index if available) and the date of most recent user satisfaction survey.

(optional - data type: free text)
User satisfaction

<<completeness>>

State of completeness of the resource.

(optional - data type: free text)
Completeness

<<completenessPercentage>>

Percentage of the state of completeness of the resource.

(optional - data type: real)
Percentage of completeness

SE_ComparabilityCoherence

<<comparabilityGeographical>>

Degree of data comparability across the geographic areas or regions referenced by the resource. Data might be derived from surveys that in general are conducted by different statistical agencies. These surveys often refer to populations of different geographical areas, sometimes based on different methodologies.

(optional - data type: free text)
Geographic comparability

<<comparabilityTime>>

Extent to which data are comparable or reconcilable over time. It refers to the degree of comparability between the measures of a time series (e.g. related to a country, a commodity and a variable) included in the resource.

(optional - data type: free text)
Time comparability

<<coherenceIntern>>

General estimate of the extent to which data are consistent within the resource.

(optional - data type: free text)
Internal coherence

<SE_TimelinessPunctuality> Evaluation of the timeliness of the resource dissemination with respect to the phenomenon it describes. In addition this section take in to account the punctuality of data dissemination.

<<timeliness>>

It refers to the speed of data availability, length of time between data availability and the event or phenomenon they describe.

(optional - data type: free text)
Timeliness

<<punctuality>>

Time lag between the release date of data and the target date announced in some official release calendar.

(optional - data type: free text)
Punctuality
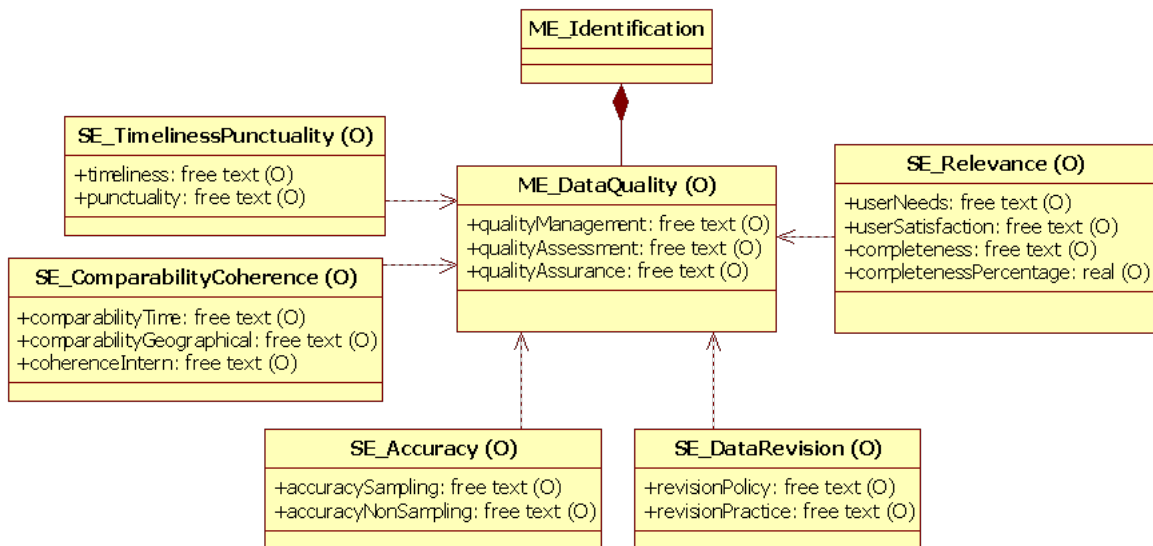
Figure A.6: Data quality

A.0.10 Accessibility (ME_Accessibility)

<<**SE_DataDissemination**>> This section reports the mode of distribution of the resource with a focus on how to access the resource, the supported formats.

<<**Distribution**>>

<<**onlineResource**>>

Link to the on-line resource. It is conditional to the policy governing distribution and sharing mechanism. For "restricted" resources it is not available.

(conditional - data type: free text - No multilingual)
Link to the on-line resource

<<**disseminationFormats**>>

Formats available for downloading the resources (e.g. excel, csv, pdf, etc.)...It is conditional to the policy governing distribution and sharing mechanism. For "restricted" resources it is not available.

(conditional - data type: sequence<free text - No multilingual>)
Dissemination formats

<<**Releasepolicy**>>

<<**releaseCalendar**>>

Policy regarding the release of the resource in accordance with the pre-announced schedule. It also provides information on the availability of the release calendar.

(optional - data type: free text)
Release calendar

<<releaseCalendarAccess>>

Link or references to the release calendar.

<div align="right">

(optional - data type: free text - No multilingual)
Access to the release calendar

</div>

<<disseminationPeriodicity>>

Frequency of data dissemination (e.g. daily, monthly, quarterly, yearly).

<div align="right">

(optional - data type: OJ_CodeList)
Dissemination periodicity

</div>

<<embargoTime>>

Time span between the completion of the production process of statistical data and their publication.

<div align="right">

(optional - data type: <free text>)
Dissemination periodicity

</div>

SE_Clarity   This section gives information about the availability of additional information (documentation, further metadata. . .) linked to the resource.

<<clarity>>

Extent to which easily comprehensible metadata are available. It indicates whether a resource is accompanied by appropriate metadata and other relevant documentation.

<div align="right">

(optional - data type: free text)
Clarity

</div>

<<metadataCompletenessRate>>

The percentage of completeness of metadata offers a numerical evaluation of the extent to which the resource is documented.

<div align="right">

(optional - data type: integer)
Metadata completeness rate

</div>

<<SE_Confidentiality>>   This section gives information on the level of confidentiality and the applied policy for releasing the resource. This metadata sub-entity concerns legislation (or any other formal provision) related to statistical confidentiality applied to the resource as well as the actual confidentiality data treatment applied (also with regard to the aggregated data disseminated).

<<confidentialityPolicy>>

Legislative measures or other formal procedures which prevent unauthorized disclosure of data that identify a person or economic entity either directly or indirectly. It consists in textual description and references to legislation or other rules related to statistical confidentiality.

<div align="right">

(optional - data type: free text)
Confidentiality - Policy

</div>

<<confidentialityDataTreatment>>

Rules applied for treating the resource to ensure confidentiality and prevent unauthorized disclosure.

<div align="right">

(optional - data type: free text)
Confidentiality - Data treatment

</div>

<<confidentialityStatus>>

Coded information describing the degree of confidentiality of the resource.

<div align="right">

(mandatory - data type: OJ_CodeList)
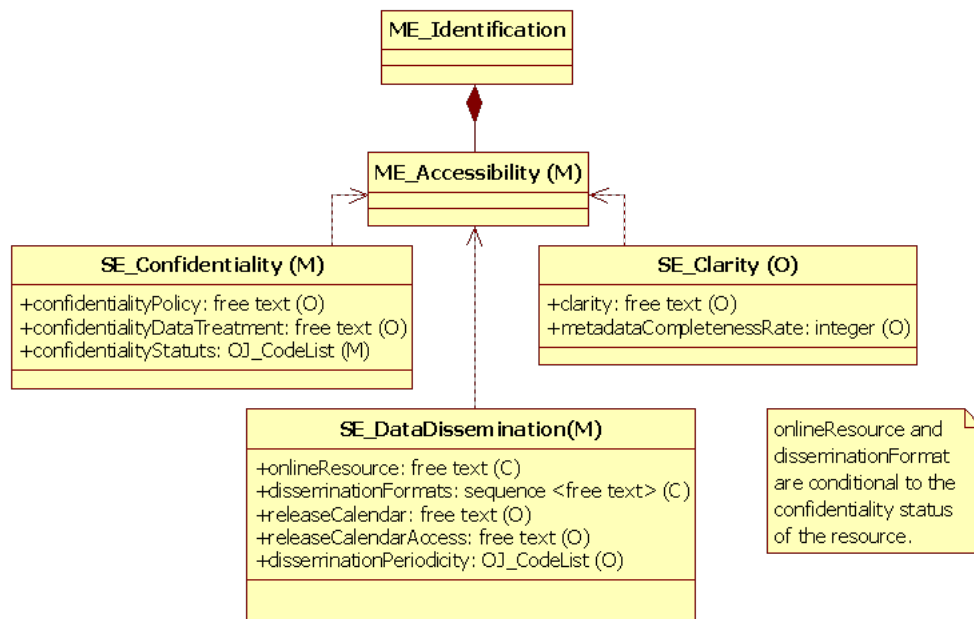**Status of confidentiality**

</div>

Figure A.7: Accessibility

A.0.11   Maintenance (ME_Maintenance)

<<maintenanceAgency>>

This section provides information about the frequency of resource upgrade and metadata maintenance.

<div align="right">

(optional - data type: free text)
Maintenance agency

</div>

<<SE_Update>>   involves maintenance operations concerning the periodic update of the resource.

<<updateDate>>

Last physical update date.

<div align="right">

(optional - data type: date)
Last update date

</div>

<<updatePeriodicity>>

Time span between the completion of the production process of statistical data and their publication.

<div align="right">

(optional - data type: OJ_CodeList)
Frequency of update

</div>

<<SE_MetadataMaintenace>>   This section involves maintenance operations concerning the periodic update of metadata to ensure that the resource is properly described.

<<metadataLastCertified>>

Latest date of certification of the metadata.

<div align="right">

(optional - data type: date)
Metadata last certified

</div>

<<metadataLastPosted>>

Latest date of publication of the metadata. It is usually automatically updated by the metadata production system.

<div align="right">

(optional - data type: date)
Metadata last posted

</div>

<<metadataLastUpdate>>

Most recent date of update of the metadata.
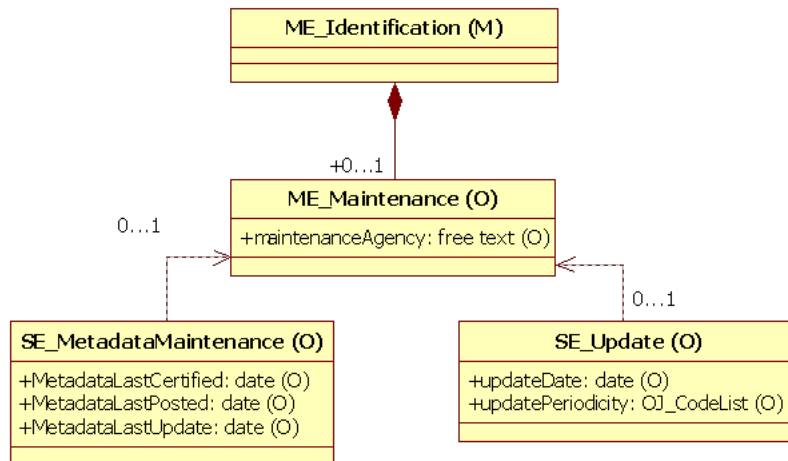
<div align="right">

(optional - data type: date)
Metadata last update

</div>

Figure A.8:

A.0.12   Documents (ME_Documents)

<<document>>

Regular or ad-hoc publications linked to the resource.

<div align="right">

(mandatory - data type: OJ_Citation A.0.15)

Documents linked to the data

</div>

<<referenceEntity>>

The metadata entity(ies) the document refers to. This element allows users to contextualize the document by making reference to specific metadata areas.

<div align="right">

(optional - data type: multipleChoice[FENIX/tech/MD/02])

Reference subject

</div>

<<referenceElement>>

The metadata element(s)the document refers to. This element allows users to contextualize the document by making reference to specific metadata element(s). The element(s) selected must belong to the metadata entity just specified in <<referenceEntity>>.

<div align="right">

(conditional - data type: multipleChoice[FENIX/tech/MD/02])
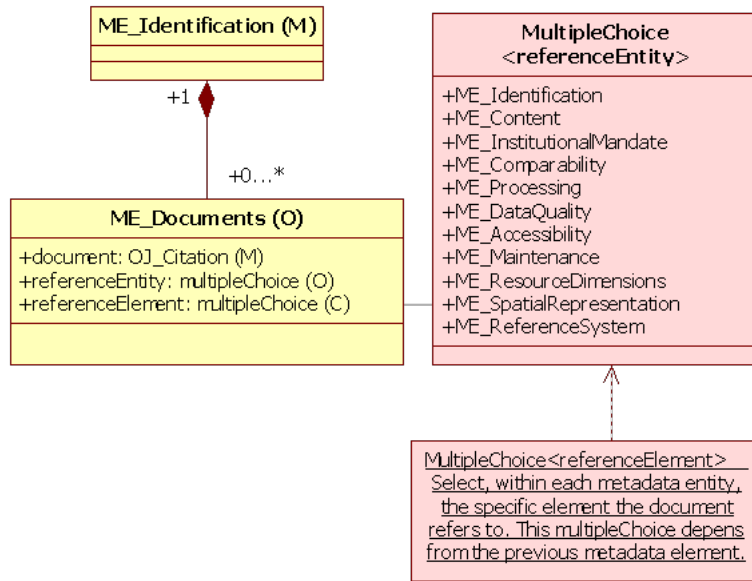
Reference subject

</div>

Figure A.9:

A.0.13 <<OJ_CodeList>>

Information that allows to uniquely identify a code list and to interpret the meaning of the codes reported. To identify a code list, it is necessary to have a name, a version number and a reference to a maintenance agency.

<<codes>>

Code corresponding to the element requested by the field it refers to.

(mandatory - data type: sequence<OJ_Code> A.0.14)

Code(s)

<<idCodeList>>

Codelist identifier. It is an alphanumeric acronyme that identify the codelist.

(mandatory - data type: free text)

Codelist identification

<<extendedName>>

Full name of the codelist.

(mandatory - data type: free text)

Extended name

<<version>>

Information about the version and the state of upgrade of the code list used.

(mandatory - data type: free text)

Version of the codelist

<<contactInfo>>

Individual/organizational contact points for a code list including information about the person(s)/maintenance-agency and how to reach the contact points.

(mandatory - data type:sequence< OJ_ResponsibleParty **??**>)
Contact(s)

<<link>>

Web site link allowing the user access to the information resource.

(mandatory - data type: free text
Codelist hyperlink

<<codeListResources>>

It allows to attach documents to the codelist or simply to cite a webpage linked to the code list to which it refers to.

(optional - data type:sequence< OJ_Citation> A.0.15)
Resource(s) linked to the codelist


```
            ┌──────────────────────────────────────────────┐
            │                 OJ_CodeList                  │
            ├──────────────────────────────────────────────┤
            │ +codes: OJ_Code (M)                          │
            │ +IDcodeList: free text (M)                   │
            │ +extendedName: free text (M)                 │
            │ +version: free text (M)                      │
            │ +contactInfo: sequence <OJ_ResponsibleParty> (M) │
            │ +link: free text (O)                         │
            │ +codeListResources: sequence<OJ_Citation> (O) │
            └──────────────────────────────────────────────┘
```

Figure A.10: Object - OJ_Codelist

A.0.14 <<OJ_Code>>

<<code>>

Alphanumeric sequence of characters that uniquely identify an attribute within a codelist.

(optional - data type: free text - No multilingual
code

<<label>>

Descriptive term associated to each code within a codelist.

(optional - data type: free text)
Label

A.0.15 <<OJ_Citation>>

Information necessary to report a reference of an attached document linked to data.

<<documentKind>>

Kind of resource attached to data.

(optional - data type: multipleChoice[FENIX/tech/MD/02, **??**])
Kind of document

<<title>>

Title by which the cited resource is known which offers a quick information about its content.

(mandatory - data type: free text)
Title

<<date>>

Reference date for the cited document.

(optional - data type: date)
Date

<<documentContact>>

Name and position information for an individual or organization that is responsible for the resource.

(optional - data type: sequence<OJ_ResponsibleParty **??**>)
Contact

<<link>>

Web site link allowing the user access to the information resource.

(optional - data type: free text)
Resource hyperlink

<<notes>>

Some additional details about the resource mainly concerning the domain to which the resource refers.

(optional - data type: free text)
Notes

<<ISBN>>

If available, the International Standard Book Number can be reported in order to uniquely identify the resource cited.

(optional[1] - data type: free text - No multilingual)

ISBN code

<<ISSN>>

If available, the International Standard Serial Number (used to identify periodical publications such as reviews and scientific journals) can be reported in order to uniquely identify the resource cited.

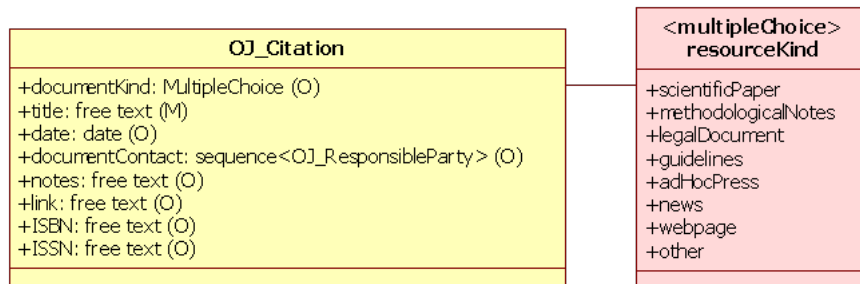(optional - data type: free text - No multilingual)

ISNN code

Figure A.11: Object - OJ_Citation

---

[1]Just one between <<ISBN>> and <<ISSN>> can be compiled