Long-term data for Europe

# EURHISFIRM

## D5.5: Report on process for extendable data models

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

**AUTHOR(S):**

Lukas Manuel Ranft (Goethe Universität Frankfurt)[1]

Jefferson Braswell (with Goethe Universität Frankfurt)

Wolfgang König (Goethe Universität Frankfurt)

**ABSTRACT:**

The fifth report of Work Package 5 describes how the common data model will be extended. The report is based on practically proven methods and processes for managing and generating standards. First, triggers are explained, which require a change to the common data model. Using the example of the development of standards within the project, the process of how the common data model will continue to be developed is explained by the example of the EURHISFIRM Legal Entity Standard.

**APPROVED IN 2021 BY:**

Jan Annaert (Universiteit Antwerpen)

Wolfgang König (Goethe Universität Frankfurt)

Angelo Riva (École d'Économie de Paris)

[1] I would like to thank Pantelis Karapanagiotis and the Leibniz Institute for Financial Research SAFE for their contributions to the report and support during the project. Furthermore, I would like to thank the Working Group of Identification and Standards (WGIS) for the discussions and constructive cooperation in the preparation of the initial standards, which are the starting point for our system.

## Table of Contents

1 Introduction ..... 6
2 Justification for Changes to the CDM ..... 6
3 The three dimensions of the Common Data Model ..... 7
3.1 Federated System Architecture ..... 8
3.2 Data Staging ..... 9
3.3 Core Data Model of Firms and Securities and Derived Entities ..... 11
4 Overarching Semantic Equivalence ..... 12
5 Metadata Management ..... 13
5.1 Participants of the change process ..... 13
5.1.1 Stakeholders: Organisation/Person ..... 14
5.1.2 Metadata Management Operational Unit ..... 14
5.1.3 Mission Group ..... 14
5.1.4 Maintenance Group ..... 15
5.1.5 EURHISFIRM Governance ..... 15
5.2 Metadata Management Process ..... 15
6 Evolving Common Data Model ..... 15
7 Conclusion and Summary ..... 18
8 References ..... 22
9 Appendix ..... 23
9.1 Legal Entity Data Standard 1.01 ..... 23
9.2 Legal Entity Data Standard 1.05 ..... 48
9.3 Legal Entity Data Standard 1.09 ..... 78
9.4 Financial Instrument Identification Data Standard 1.05 ..... 113
9.5 Legal Entity Data Artifact 1.05 ..... 143

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## List of Figures

Figure 1: EURHISFIRM Common Data Access Service Data and Technology Standard Zones (by Jefferson Braswell) ..... 8
Figure 2: EURHISFIRM Data Staging (by Jefferson Braswell) ..... 10
Figure 3: EURHISFIRM Central Entities (by Jefferson Braswell) ..... 11
Figure 4: Overarching Semantic Equivalence ..... 12
Figure 5: Metadata Management Visualization ..... 14
Figure 6: EURHISFIRM Central Entities with Sources (by Jefferson Braswell) ..... 17

## List of Tables

Table 1: CDM Change Scenarios ..... 6
Table 2: EURHISFIRM CDM standards development schedule ..... 16

## Abbreviations

| | |
|---|---|
| A | Data Admin |
| API | Application Programming Interface |
| C | Data Consumer |
| CDAS | Common Data Access Service |
| CDM | Common Data Model |
| CDMG | CDM Compliant Gateway |
| CD | Continuous Deployment |
| CI | Continuous Integration |
| D | Data Standards |
| DCU | Data Collection Unit |
| DSU | Data Submission Unit |
| EFII | EURHISFIRM Financial Instrument Identifier |
| ELEI | EURHISFIRM Legal Entity Identifier |
| ELF | Entity Legal Form |
| FAIR | Findable, Accessible, Interoperable, and Reusable |
| G | Data Gateway |

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| | |
|---|---|
| GLEIF | Global Legal Entity Identifier Foundation |
| ISO | International Organisation of Standardization |
| L | Legacy Database |
| MSWG | Mission-specific Working Group |
| NCA | National Competent Authority |
| NIC | Network Integration Centre |
| SMG | Standards Maintenance Group |
| S | Legacy Source |
| SME | Subject Matter Expert |
| T | Technology Standards |
| WGIS | Working Group of Identification and Standards |
| WP5 | Work Package 5 |

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 1 Introduction

The European data sets for historical financial and firm data are heterogeneous due to different laws and regulations, languages, and currencies of the individual countries (Karapanagiotis, 2019). Nevertheless, EURHISFIRM brings countires with different data sets to agree on flexible standards for the said data, allowing users to analyse such heterogeneous source data based on an extendable common data model (CDM) with, amongst other means, a uniform identification scheme across the countries.

All data points relevant to EURHISFIRM remain in the past. However, data models and respective standards need to evolve as the retrieval of the data has not been finalised yet. Chapter 2 presents various triggers for necessary changes to the CDM after the initial implementation. The CDM consists of several dimensions, all of which are affected by changes. We explain these dimensions and the changes to them in chapter 3. Chapter 4 presents the semantic equivalence model to support varying underlying technologies. The changes require EURHISFIRM to design an initial CDM and develop a process to reflect in the future newly collected data so that this model fulfils the users' underlying data requirements. This process is explained in chapter 5. It has already been applied practically within the project and in chapter 6 will be elaborated how to continue after the project. The different steps within metadata management in EURHISFIRM are explained and illustrated using the EURHISFIRM Legal Entity Standard for the indentification of legal entities (e.g. corporations). Finally, in chapter 7, we will summarise the results and place them in the project's overall context.

# 2 Justification for Changes to the CDM

In the later implementation of the EURHISFIRM system, a set of initial standards for different types of data shall be available, which allow the compatibility of data on a European and a national level. If further data with different characteristics were integrated into the system, the existing standards for these data would have to be revised. For new subject areas for which no standard has yet been defined, data cannot be integrated with other data until required standards are defined and implemented.

Revising the standards may be necessary due to the appearance of new data or data characteristics and changes in user requirements.

Changes can therefore be classified into two dimensions. They are either data- or end-user-driven, and they are either changes to an existing data type or require new ones.

This two-dimensional matrix illustrates how change can be initiated.

| CDM Change Scenarios | Covered Subject Areas | New Subject Areas |
|---|---|---|
| **Data-driven** | Use Case 1 | Use Case 3 |
| **End-user-driven** | Use Case 2 | Use Case 4 |

*Table 1: CDM Change Scenarios*

The following examples illustrate potential changes within said 2x2 dimensions.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

Use Case 1: The CDM has to be adapted to a new collection of of financial and corporate data. For instance, data from a European country that has not yet added any data sets. Since historical financial and firm data can differ significantly between countries, it is possible that the regional specifics of the data set are not considered in the existing standards. Thus, it must then be decided to what extent the source's additional features should be reflected in the CDM.

Use Case 2: Due to EURHISFIRM's focus on data up to 1980, users may wish to link data to another database with more recent data to conduct cross-cutting analyses. The other databases may not use the same identification scheme as EURHISFIRM. This may lead to the consideration of adding a link table to the CDM to enable the user's analysis.

Use Case 3: EURHISFIRM mainly deals with financial and firm data collected by various institutions and individual researchers. Hypothetically, the latter may have investigated the link between stock prices and a subject area that was not yet part of EURHISFIRM, such as results of a national election, which is why the researcher's data set contains both types of data. Therefore, it must be decided whether, for instance, election results should become an integral part of the CDM and whether their standards should be defined.

Use Case 4: The focal point of EURHISFIRM is on financial and firm data. However, users might pursue broadening the scope and extending to other data from new subject areas, such as socio-technical (e.g. patents) and sociological (e.g. staff careers, shareholders, and partners) data, which would require an extension of the CDM to new subject areas. In contrast to the previous example, the request here is made by the user of the system and not by data that is already available.

The four use cases show that it is necessary to define a process for further developing the CDM after its implementation.

# 3 The three dimensions of the Common Data Model

The CDM development in EURHISFIRM takes place in three dimensions: the federated system architecture, the stages of processing of data from raw source to harmonised end-user content, and the CDM that provides semantic definitions and relationships. These dimensions will be explained in this chapter.

A standardisation committee, named Working Group of Identification and Standards (WGIS), consisting of experts from all work packages and external consultants, plays a central role. Over 18 months, the WGIS met every 14 days to discuss and when necessary to revise the CDM's standards. The process of this revision will be explained in more detail in the following chapter 5.2. The entire CDM is too complex to generate all versions in one run. Instead, it was developed incrementally within the project's framework to have more communicable and thus manageable sub-packages, which can also be revised again in the ongoing process. We expect such revision processes more often at the beginning of such a work schedule and have these occurrences be substantially reduced over time.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 3.1 Federated System Architecture

First, the intended development process and data architecture were discussed in the project. A Federated System Architecture was already mentioned in our proposal to the European Commission and during the project's meetings identified as the most suitable solution, based on which the following diagram was created. The federation comprises actors who share equal rights and have agreed to common subject standards and communication standards. The federation of research infrastructures based on a National Competence Centres network is invisible to the European data model user. Figure 1 visualises that users only see the Common Data Access Service (CDAS), but not the underlying gateways, even if they simultaneously access all their data. This leads to enhanced usability for the users but requires the data to be transformed from raw to standardised.

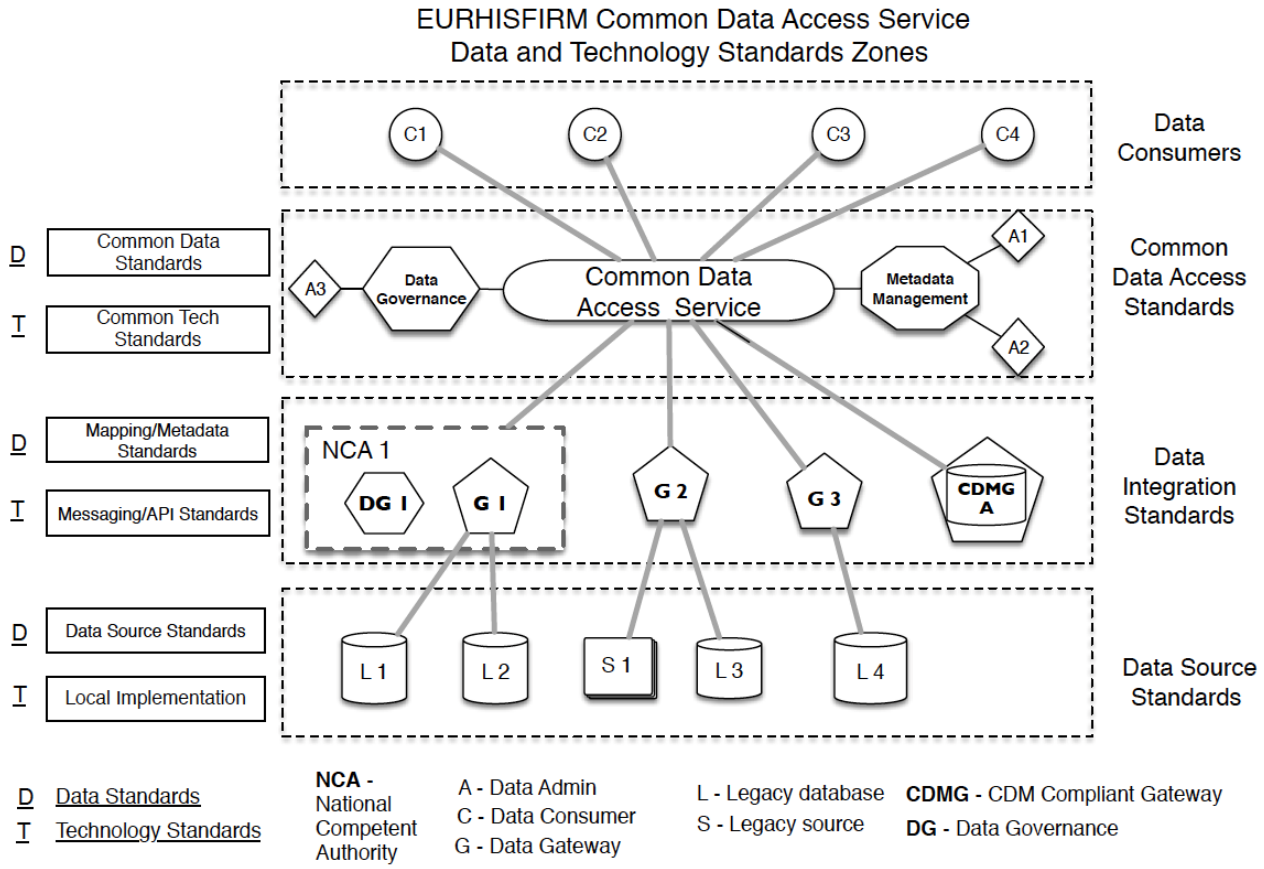



*Figure 1: EURHISFIRM Common Data Access Service Data and Technology Standard Zones (by Jefferson Braswell)*

This diagram was the first idea of a later design, revised in the project's subsequent developments. First, the diagram's components will be explained in more detail, so that their relationship to one another can then be explained.

A Legacy Database (L) is a currently existing database containing relevant data in the scope of EURHISIFRM (for instance the French D-FIH database, https://dfih.fr). In the drawing, we have, for instance, L1 and L2. This data is in a digital but non-harmonised format compatible with the CDM. In contrast, Legacy Source (S) is a scan of a newspaper or a spreadsheet (i.e., OCRed), again not yet harmonised. The initial hypothesis

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

of Data Gateways (G) tasks was to dynamically access local data and transform that data to support the common data access service. So G1 would dynamically access and transform L1 and L2 into a CDM compliant format. This dynamic virtual access was intended behind this gateway, and the design has been evolved during the project to process stages explained in the next chapter. NCA 1 is a source contribution to EURHISFIRM by a National Competent Authority in which the CMD standards have been realised employing their own Data Governance (DG1). A CDM Compliant Gateway (CDMG) is an implementation that complies internally with the CDM specifications. This allows the database to be readily accessed. Consumers (C) are users that interact with the system in various forms, via e. g. a web browser or an application programming interface (API) connection. Data Admins (A) are individuals, groups, or organisations that perform metadata management or data governance functions. The Common Data Access Service (CDAS) is the service that provides an initial point of access and facilitates user requests to access data in the EURHISFIRM federated architecture. Metadata Management is the process by which a data model is maintained and enhanced, and Data Governance (DG) is the process by which data quality is ensured (Bernstein, 2003, Khatri and Brown, 2010).

The layers in the diagram above represent a vision of the system architecture based upon the objective to translate user requests using a CDM to the representations of sources. Successive work on this vision has led us to realise that the gateway layer on this diagram is best realised by a series of process stages described in Figure 2. In this diagram, two primary types of standards have been identified: Data standards and technology standards. The need for data standards is being addressed with the CDM. The need for technology standards is being addressed in another dimension which has to do with data staging and operational system governance (Work Package 9). The different layers in the original vision diagram describe different types of needed standards and are relevant at that architecture level. These standards are objectives which must be developed for the implementation of a EURHISFIRM infrastructure. That is, the standards at the top levels involve the CDM. The standards at the lower levels involve the conventions for acquiring and cataloguing source data in a form that will allow the data to be harmonised and integrated with the higher-level CDM.

Furthermore, users submit queries to the CDAS. After receiving the data, the user may find mistakes or gaps in the data provided and may wish to correct or contribute additional data updates or revisions. As a part of this process, the end-user can drill down to the sources to validate their feedback or contributions. To this end, raw sources shall be preserved, and mechanisms for accessing them will be provided. This offers a mechanism to allow the correction of data errors in EURHISFIRM.

## 3.2 Data Staging

The 'gateway' function in the original diagram has been replaced by a series of process stages already described in Report D5.2 (Karapanagiotis, 2020) that accomplish the objectives of transformation and integration of sources into the CDM representation. These steps are described in Figure 2.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

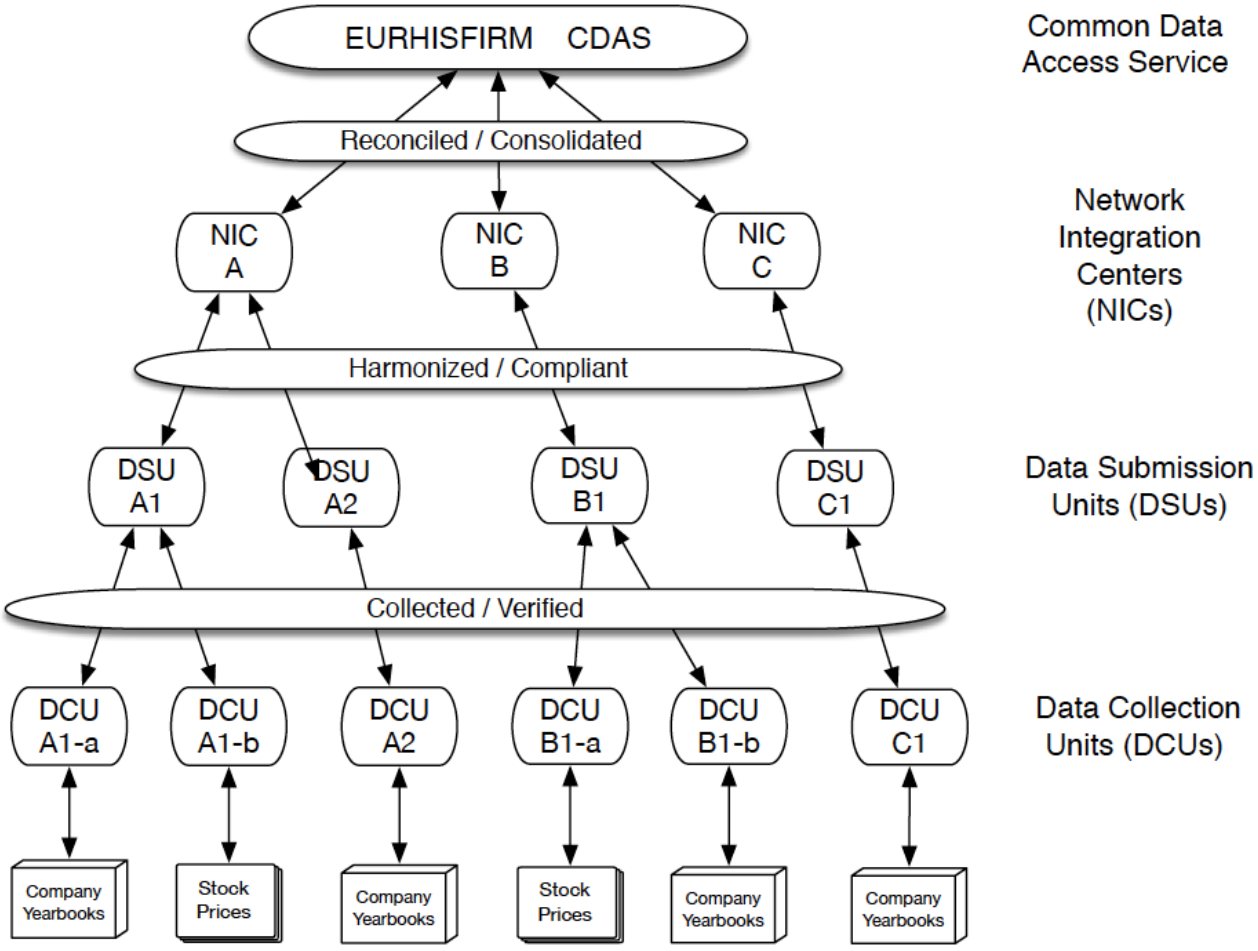



*Figure 2: EURHISFIRM Data Staging (by Jefferson Braswell)*

This diagram shows the process stages to perform the previously mentioned transformation of original data to CDM-compliant data to provide data to end-users with the CDAS. It visualises the sources out of which data is collected by Data Collection Units (DCUs). The Data Submission Units (DSUs) have the objective of harmonising data formats and labels collected by the DCUs to submit harmonised data to the Network Integration Centres (NICs). Harmonisation means to transform the data using the same metadata description as tags, labels and field formats. The NICs process the data submitted by DSUs to provide common identification of – as first examples - entities and financial instruments to support CDM access by end-users. The NICs can then integrate data and provide it to end-users in the CDAS.

The coordinated operation and coordination of the different process stages are essential to provide high-quality, consistent data to end-users. The harmonisation, identification, and consolidation of data in this fashion is crucial to achieving the EURHISFIRM research infrastructure's objectives.

This process does not prevent users from examining multiple layers of CDM-compliant data to source data as transformation steps should be transparent. The harmonisation process between DCU and DSU is also consistent with the Data Documentation Initiative (DDI) metadata definitions of variables for local data

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

and conceptual variables for harmonised data. Local language labels for financial terms will be replaced with common terms for the same data elements.

## 3.3 Core Data Model of Firms and Securities and Derived Entities

Content-wise, the very core of the future EURHISFIRM data structure concerns the "financial realm" of listed corporations (as an initial legal entity type). This approach results in financial data on firms (for instance as contained in the profit and loss statements) and in their stock data on (as an initial financial instrument type) e.g. share price. Both entities enable the modelling of the stock market as an initial market type, with currency as an attribute. These three subject matters constitute the centre three columns in Figure 3. Our wording "initial" implies that this very core data structure later can be and has to be incrementally extended, for instance by additional entity types (e.g. natural persons) and by additional financial instruments.

A common identification of the very core entities enables common access of end-users to data. Thus, we need identification standards for legal entities, financial Instruments and markets.

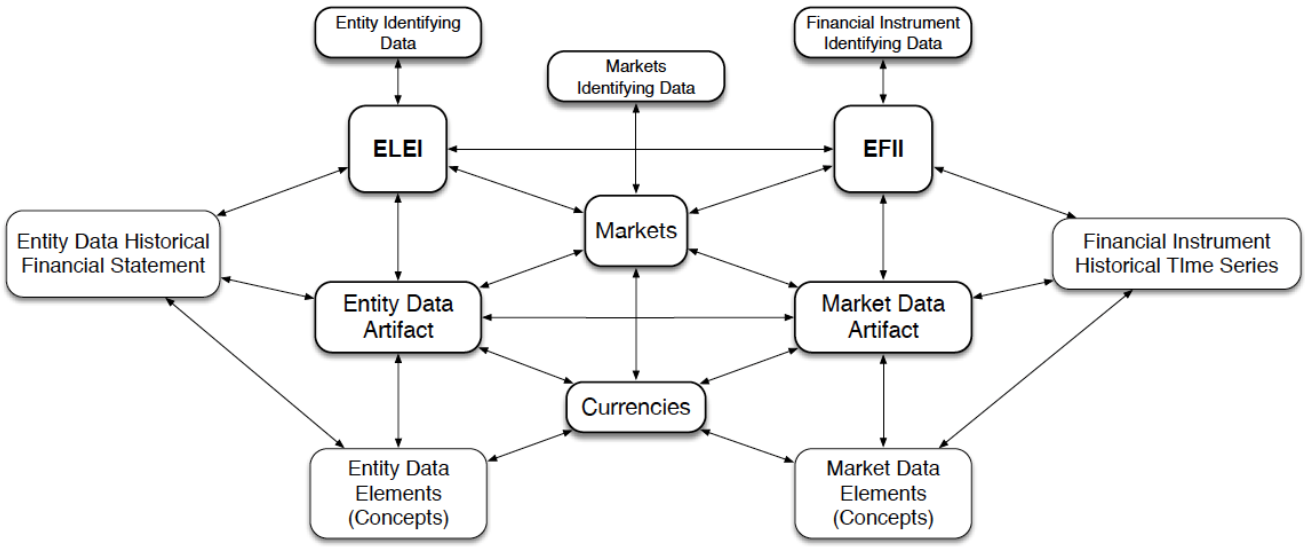



*Figure 3: EURHISFIRM Central Entities (by Jefferson Braswell)*

Specifications that elaborate aspects of this semantic model have been developed using a widely accepted standards development methodology (Enterprise Architecture (see D5.4)). by a standards development team in EURHISFRM (the WGIS). The development of these standard specifications evolved throughout the project by leveraging specific existing industry standards that pertain to the identification of organisational entities and financial instruments. That is, the ELEI (EURHISFIRM Legal Entity Identifier) is derived from the reference data standards for the LEI (Legal Entity Identifier) as published by the Global Legal Entity Identifier Foundation (GLEIF). The ELEI extends the reference data of the LEI by enabling historical snapshots of changes in entity reference data (something which is not supported in the GLEIF standards). The EFII (EURHISFIRM Financial Instrument Identifier) likewise is derived from the Object Management Group (OMG) Financial Instrument Global Identifier (FIGI) standard, which provides for the identification of securities at different hierarchical levels, namely trading venue, sovereign jurisdiction, and

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

the issuing corporation. These two identifiers, ELEI and EFII, establish the initial foundation for organising and integrating data from various sources related to these two fundamental classes. The double arrows in the diagram indicate that these entities have a relationship to all various data elements that a user can see when accessing the system.

Below the two core entities, there are the corresponding artefacts. In the diagram, artefacts are a mechanism to collect data samples in flexible containers that do not have rigid structures, which will be further explained in the subsequent chapter 7.

Current work is ongoing regarding CDM standards for collecting and identifying financial data associated with these two fundamental classes.

# 4 Overarching Semantic Equivalence

In light of the different platforms that may be most appropriate to use at various stages of a data collection, data harmonisation, identity-matching and end-user common access layers of functionality that exist in EURHISFIRM – as well as the data formats that would be used to transfer data between stages -- a higher-level, the semantic approach is needed.

This "semantic equivalence" approach allows for the interaction and integration of data on collaborating platforms in a distributed, federated architecture that can be a hybrid of technological platforms best suited for use in different areas of the system. The figure below shows this compatibility with varying types of sources. This approach enables historical data to be collected at one level and propagated through the system to other user access levels that may involve different access and storage forms.

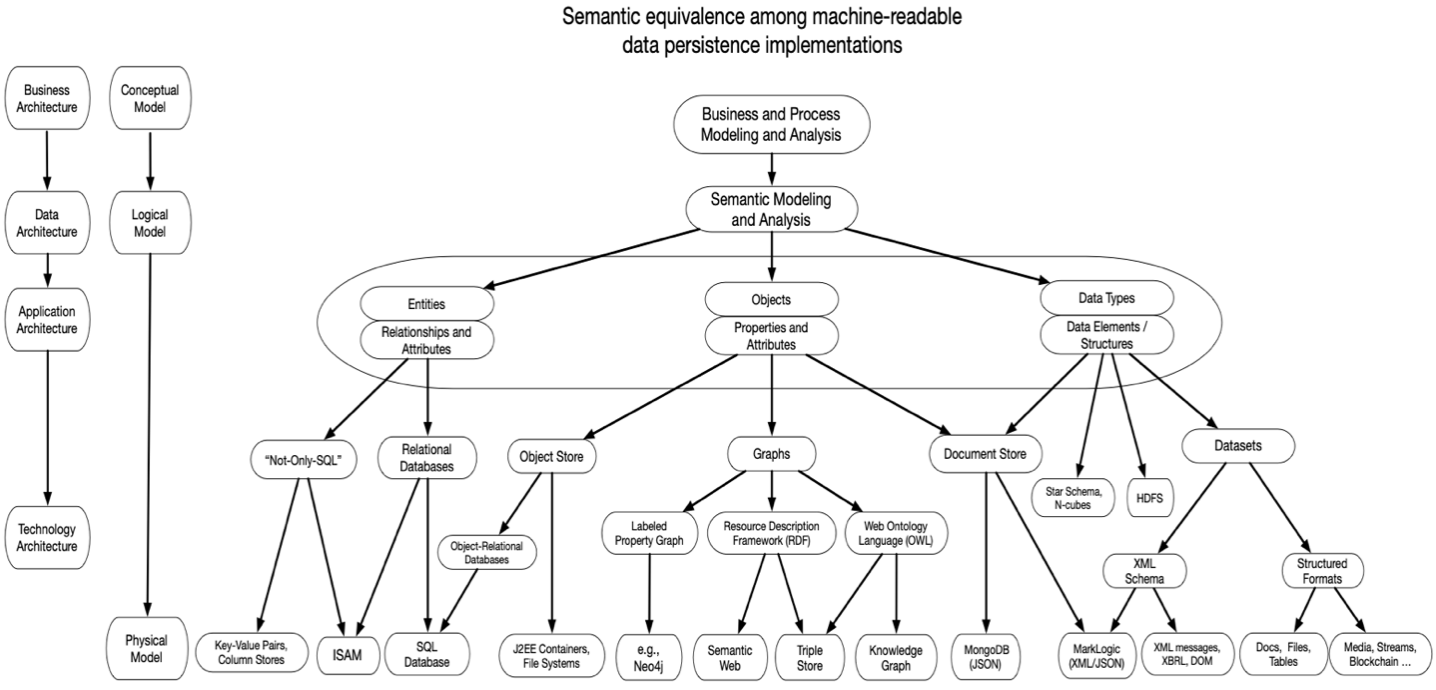



*Figure 4: Overarching Semantic Equivalence*

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

On the bottom are the different possible technologies in which source data is stored. Depending on the type, they can be grouped into more general classes such as relational data, object data or graph data. The semantic equivalence model allows the user to accommodate these different formats to account for the needs and requirements. These needs and requirements exist – as defined in Enterprise Architecture – on the different architectural levels of business, data, application, and technology.

# 5 Metadata Management

Metadata management is the end-to-end process that facilitates the integration and evolution of a schema describing the underlying data (Bernstein, 2003). With the higher complexity due to heterogeneous and unstructured data, a semantic model is used instead of a schema, because it is more flexible (Rajput and Khoja, 2018). It describes the characteristics of the existing values, the different expressions of the values and their origin. An example of the DDI metadata description describing the sources can be obtained from the data and sources documentation provided by Work Package 4 (Poukens, 2019). As a result, only the described values can be stored in the system. Like the staging system displayed in Figure 2, there are various layers of metadata on each stage. We have metadata on the bottom layer describing the sources and a different metadata level to build the relationship between the sources and the harmonisation layer. On top, we have the metadata that the end-user is utilising to query the system. All these levels need to be maintained and managed.

To illustrate the governance of these changes to the CDM, we describe a procedure based on the model of GS1, an organisation that has been implementing a process for generating and updating standards for logistics data management for many decades (GS1, 2019). This process of self-binding cooperative decisions on standards (a concept of committed consensus) has already been similarly applied by the Working Group of Identification and Standards (WGIS) within the project. For this purpose, after some ramp-up activities of all working packages, an agreed initial list of standardised objects was defined. This list, presented in Table 2 in chapter 6 of this report, was distributed to all consortium work packages. Subject matter experts of each work package were then delegated to the group, which developed and agreed on the standards together. These standards were designed to align with the FAIR principles, to provide a research infrastructure that facilitates findable, accessible, interoperable and reusable data (Wilkinson *et al.*, 2016).

## 5.1 Participants of the change process

We explain the different roles depicted in the metadata management process diagram in Figure 5 and the individuals and organisations who fill these roles to understand the metadata management process.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

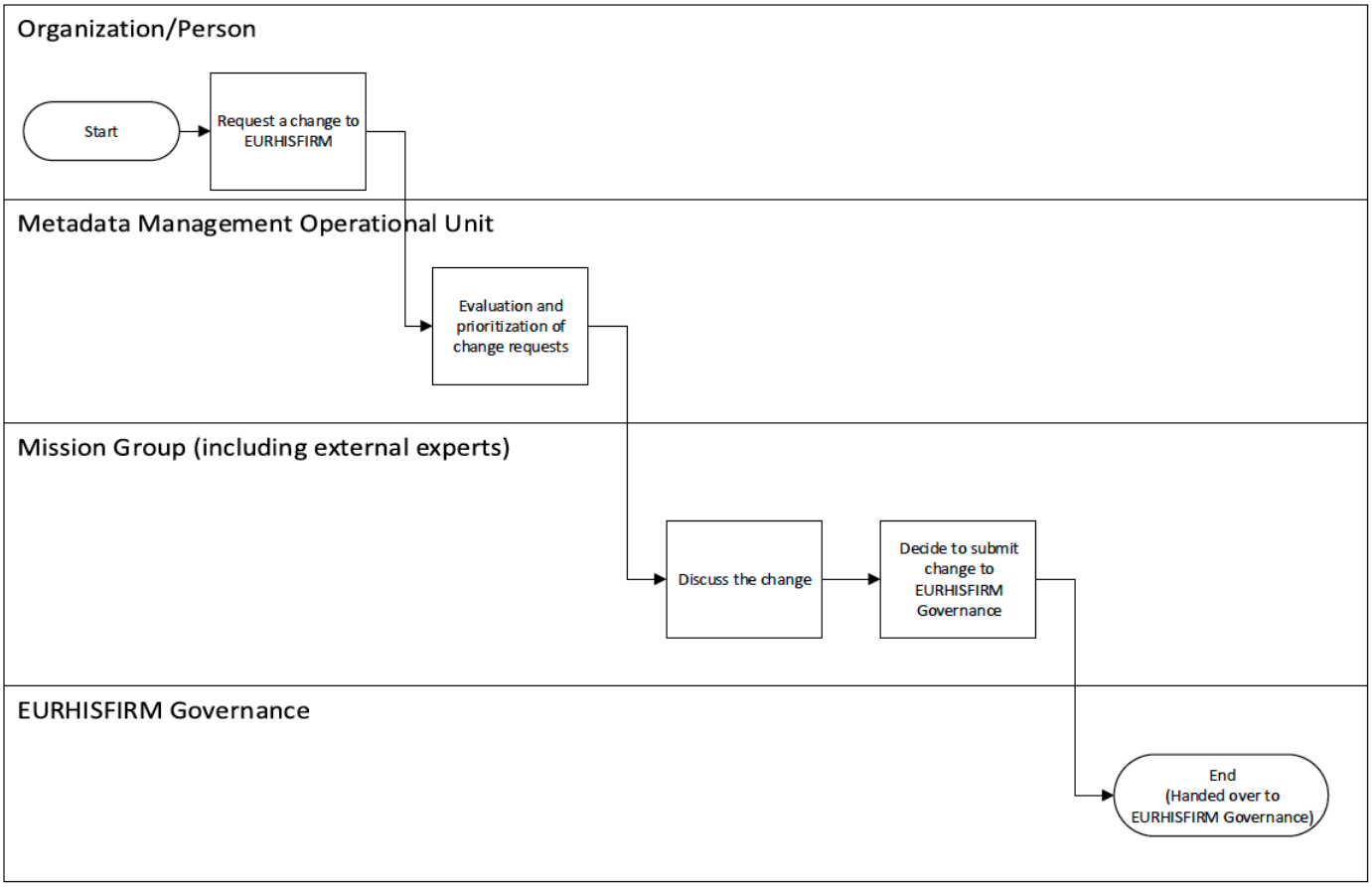



*Figure 5: Metadata Management Visualization*

### 5.1.1 Stakeholders: Organisation/Person

Who should be able to request changes to the EURHISFIRM CDM? Every stakeholder should have the possibility to request a change to the EURHISFIRM system if enough evidence is provided to justify the change application.

### 5.1.2 Metadata Management Operational Unit

To accompany the further development of the EURHISFIRM system and its CDM, an operational unit must deal with metadata management. It should be composed of specialists and representatives from the different members of the EURHISFIRM system. It is advisable to fill voids with external experts' competency.

### 5.1.3 Mission Group

The mission group deals with the definition and development of standards, which are not yet realised in the implementation of EURHISFIRM. This group is fulfilling the work that was performed by the WGIS during the EURHISFIRM design project. The group should consist of subject matter experts (SME) for metadata management from various countries. This means that they are also already part of the Metadata Operation Unit of EURHISFIRM described above. This group should be enriched with external experts for the topics to be discussed as necessary.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 5.1.4 Maintenance Group

This group is not visible in the process diagram and not operating until an initial implementation of the EURHISFIRM system has taken place. Like the mission group, the maintenance group is also part of the operational unit of metadata management, and a person can potentially work in both groups. This group is concerned with further developing existing standards, once they are decided to be implemented in the EURHISFIRM system. This group is closely connected to the EURHISFIRM data governance responsible for ensuring the various members' implementation. As long as the system is not in an implementation state, this group is not active.

### 5.1.5 EURHISFIRM Governance

The overarching governance of the EURHISFIRM system is manifested by the organisational structure and rules within the community. Governance comprises a set of processes aside from metadata management. Part of the EURHISFIRM governance is the data governance topic, ensuring that the mission group's decisions are enacted and maintains data quality (Khatri and Brown, 2010). These processes will not be discussed in this report, as the system governance and business model are part of the reports of work package 10. This includes, for example, the various processes for decision-making within EURHISFIRM, but also the monitoring of implementation and compliance with them.

## 5.2 Metadata Management Process

To update and extend standards within the EURHISFIRM system, the metadata management's underlying process will now be explained in detail.

Initially, any stakeholder is entitled to request changes to the metadata of EURHISFIRM. This is enabled regardless of the communication channels or the nature of the requested change. These submitted suggestions for changes are pre-filtered and prioritised by the Metadata Management Operations Unit of the EURHISFIRM project. In an already implemented system, it would also be evaluated whether the requested change concerns a standard already implemented or a standard not yet implemented, to evaluate a potential impact on the metadata and the governance unit.

The mission group should meet at regular intervals to discuss the various standards. Following the EURHISFIRM design project, the current set of standards will need to be revised, which the mission group should address. Once a standard has been revised, the mission group should submit it to the EURHISFIRM Governance Unit. The EURHISFIRM Governance Unit will debate and finally decide on the alteration or implementation of the standard. However, once a standard has been adopted, this should not mean that the mission group cannot develop it further. On the contrary, standards should be continuously reviewed and revised based on the model of Continuous Integration / Continuous Deployment (CI/CD), as it is applied in large software companies (Gallaba, 2019).

# 6 Evolving Common Data Model

We propose to develop and refine the CDM in the same way as the initial set of standards was designed in the project. This means that the CDM parts will be revised again and again based on the requirements

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

changing over time. In the second half of 2019, leaders from all project work packages met to prepare the roadmap of the WGIS in close collaboration with the architecture designed by work package 9.

Table 2 shows the development schedule for the various CDM (sub-)standards in 2020, where we also find the Data Staging topic elaborated in section 3.2. Each standardisation topic has been introduced in a first WGIS meeting, discussed in detail after 14 days and after another 14 days has been summarised and decided upon.

| EURHISFIRM CDM standards development schedule | | |
|---|---|---|
| Date | Meeting | Standardisation Topic |
| 22. Jan 2020 | WGIS 2020 | Standards Development Roadmap |
| 06. Feb 2020 | discussion | |
| 20. Feb 2020 | Feb WGIS | Legal Entity Data (Identifying) |
| 05. Mar 2020 | discussion | |
| 19. Mar 2020 | Mar WGIS | Legal Entity Data (Historical) |
| 02. Apr 2020 | discussion | |
| 16. Apr 2020 | Apr WGIS | Legal Entity Data (Staging/Integration/Access) |
| 30. Apr 2020 | discussion | |
| 13. May 2020 | May WGIS | Financial Instrument Data (Identifying) |
| 28. May 2020 | discussion | |
| 11. Jun 2020 | Jun WGIS | Financial Instrument Data (Historical) |
| 25. Jun 2020 | discussion | |
| 02. Jul 2020 | Jul WGIS | Re-evaluation of the First Five Increments |
| 09. Jul 2020 | discussion | |
| 23. Jul 2020 | Jul WGIS 2 | Financial Instrument Data (Staging/Integration/Access) |
| | Aug break | |
| 03. Sep 2020 | discussion | |
| 10. Sep 2020 | Sep WGIS | Financial Data (High-Level Taxonomy) |
| 24. Sep 2020 | discussion | |
| 08. Oct 2020 | Oct WGIS | Financial Data (Historical) |
| 22. Oct 2020 | discussion | |
| 05. Nov 2020 | Nov WGIS | Enumerated Codes and Categories |
| 19. Nov 2020 | discussion | |
| 03. Dec 2020 | Dec WGIS | Data Provenance Drill-Down |

*Table 2: EURHISFIRM CDM standards development schedule*

However, in this work sequence, a *decision* four weeks after its introduction means a decision on this 'increment' given our present state of knowledge. With the emergence of new findings, e.g. new data or use-cases, a discussion of a new or altered standard may become necessary. For instance, as visible in the table, identification topics have been treated repetitively to account for new insights, and in the middle of the schedule, a re-evaluation of the first five increments took place. The discussions were also in collaboration with the system architecture design created by work package 9. Other topics defined that did not fit in the plan for 2020 but still need to be discussed like the API design and access plan. These have been handled at the beginning of 2021.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

As a content-wise example of the incremental development combined with additional data structures, we see Figure 6, an updated version (after two revisions) of the central entities' standards elaborated earlier in Figure 3. The earlier version only shows the elements visible to the end-user; Figure 6 also shows the parts of the sources from where the harmonised data is derived. The figure also shows normalised versions of financial statements and indices. The additional lower section depicts data collection from sources described with appropriate DDI metadata definitions from the previously mentioned work package 4 for entity and market data.

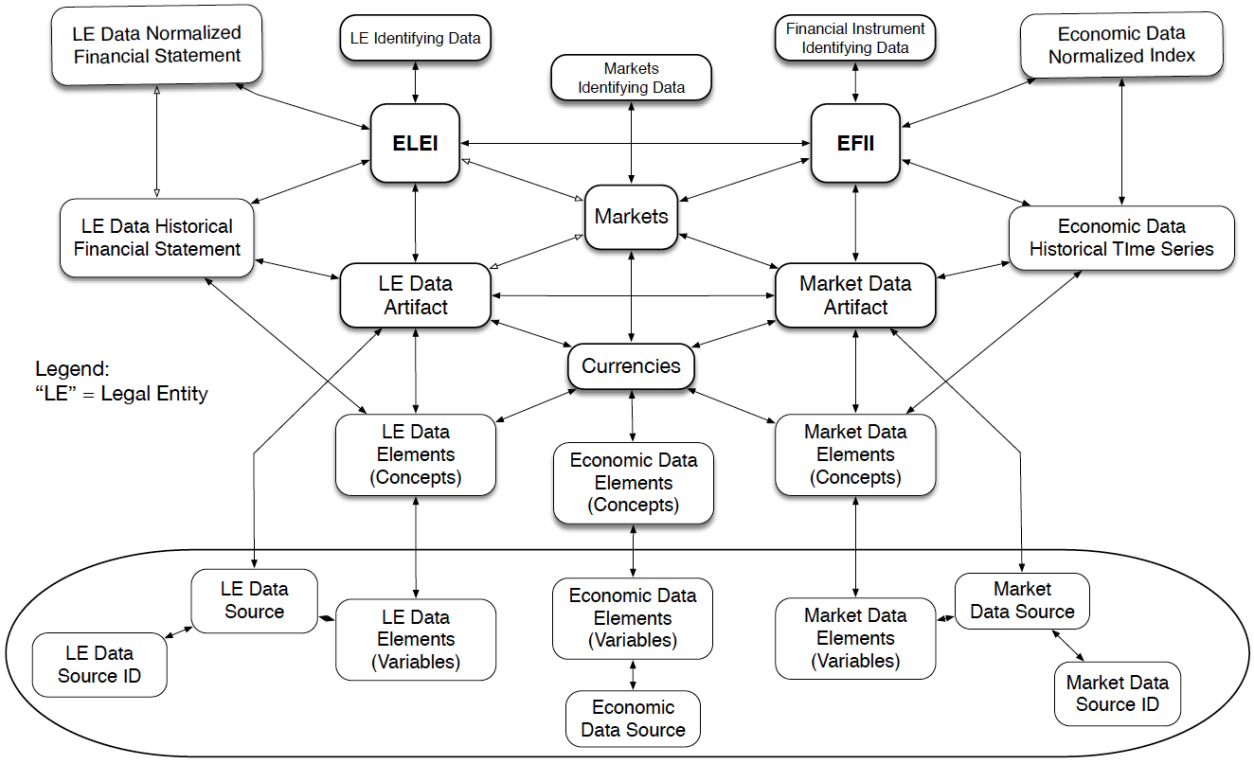



*Figure 6: EURHISFIRM Central Entities with Sources (by Jefferson Braswell)*

The additional top layer components, the top left and top right units, are showing normalised financial data, which allow comparability through 'discounting' by the inflation rate, and the definition of common labels, and terms.

In this report, the incremental development of a standard is exemplified by identifying organisations; we call these 'legal entities'. This identification scheme is a central part of the CDM, as already seen in Figure 3. Between January and December 2020, the identification of legal entities (and financial instruments) was a topic, and in July a composite decision was taken on this. Still, even in November we once again slightly altered the standard.

The Global Legal Entity Identifier (LEI) was already mentioned in the EURHISFIRM project proposal to the European Commission as the aspired identification scheme. The choice of the LEI is also underlined by literature and studies (Chan and Milne, 2019, Bottega and Powell, 2012). The initial design document was set up in March 2020 and can be found in appendix section 9.1 Legal Entity Data Standard 1.01.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

In the systemic course of discussions, it was identified that the LEI specification needs to be extended to meet historical data requirements within EURHISFIRM. GLEIF issues LEIs to contemporary active players in the market since 2012 and legal entities that were no longer active at that time have not and will not receive any LEIs from GLEIF (Global Legal Entity Identifier Foundations, 2020). For EURHISFIRM, an incremental revision we call EURHISFIRM LEI (ELEI) has been developed, which is listed in section 9.2. and comprises the first revisions for the identification of *historical* legal entities. GLEIF signals its willingness to grant EURHISFIRM the status of a special Local Operating Unit (LOU) for historical legal entities and provide a cross-tabulation of contemporary and historical legal entities.

In the course of discussions and the development of other standards, new findings were made, which again influenced the standard for identifying legal entities. Therefore, the standard for identifying legal entities was revised again to account for governmental organisations and government-sponsored entities. Section 9.3. shows the current version of the standard.

The development of a standard may require the revision of an existing standard, and it might be necessary that an entirely new standard must be defined for EURHISFIRM to extend the existing CDM. Moreover, also the controlled vocabularies – whenever logical – should follow existing standards. An example is section 3.3.8. There the Entity Legal Form data type (ELF) is defined, which is also a standard based on a norm of the International Organization of Standardization (ISO)(Global Legal Identifier Foundations, 2020). This describes the possible 'values' of the attribute Legal Form, which shows the close relation between standards and how they influence each other. This stepwise specifying and partly recursive process ensures an improved adaptation to the changing requirements of representing historical financial and firm data in EURHISFIRM.

# 7 Conclusion and Summary

This report concerns the process for extendable data models, i.e. the extendable CDM with its important interactions, including the respective processes and other data structures that may be beneficially integrated.

Thus, given a substantial amount of yet unknown data sources or data structures, we regard flexibility as a high data structuring priority. On the data object level, the report highlights the following important means of extendibility:

- We developed a **core data structure** comprising the following three legs:
  - A core class of legal entities as a linchpin of firm data collections with the ELEI as advanced identification scheme
  - A core class of securities as a linchpin of financial instruments with the EFII as advanced identification scheme
  - A core class of financial markets.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

This *core data structure* is envisioned to stay invariant for an extended period concerning the data structure as the EURHISFIRM data structure linchpin.

Somewhat similar to the ease of "growing "relational data structures, this core data structure is extendable in various ways, for instance:

- One may add new attributes to the entity classes of the core data structure, for instance spell-out market attributes
- One may add new entity classes (e. g. election results)
- We introduce the concept of an "artefact" as a form of a container of data elements gathered at various points and collection stages. This concept supports the ease of extending the data model's semantic definitions to include various flexible data elements expressed as key-value pairs. For example, different line items of companies' financial statements can be collected without requiring that they adhere to a rigid financial statement structure.
- Combining the core class identifier standards combined with the artefact data structures enables the extensibility and flexibility of the CDM and interoperability with other forms of databases and repositories (also, for instance, with repositories of linked data that use Wikibase).
- Following extensive discussions within EURHISFIRM between
  - facilitators of a further meliorated classically *structured* data model that supports the efficient automation of querying long time series and
  - facilitators of an approach based on knowledge graphs which have advantages in the discovery phase of data exploration.
- We developed an overarching **semantic equivalence concept** of data (see Figure 4) that supports the transformation of information and data from one type of representation to another (say: from a relational form to a graph database – and vice versa). Again, this concept enhances the future continuity, and thus the longevity of the EURHISFIRM data structure.
- We developed a **staging model of harmonising data** from raw data to EURHISFIRM standards-compliant data to implement the gateway functions (see Figure 2) necessary to enable a Common Data Access Service (CDAS). This staging also enables the federated processing of distributed data stores – likewise the federation of data management tasks – across Europe. Such an architecture also allows for the number of national Network Integration Centres to vary between 0, 1 and many.

How did these results evolve, and how should we systematically realise such types of solutions? We generalise our **development process** to the following guidelines:

- In such a high-dimensional design complexity, one cannot prescribe from the "top-down" this kind of a data structure all at once from the conceptual model down to the physical model. Thus, after

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

an initial tutorial phase of all EURHISFIRM participants, we chose to exercise an **incremental data model design methodology**.

- We initially tutored all work packages – participating in the Work Group on Identification and Standards (WGIS) – on the Enterprise Architecture methodology for specifying the system design. The Enterprise Architecture methodology first identifies the system's requirements in the form of a Business Architecture that takes stock of all of the system's major business process domains.

- This level of analysis is required to be completed (in order not to leave out use cases or processes that should be included) before moving on to the next level of analysis, which is that of the Data Architecture that is required to support all of the processes and requirements identified and catalogued in the Business Architecture layer.

- A full Enterprise Architecture methodology would then move on to the Application Architecture level, and only subsequently, finally, address the Technology Architecture, which specifies the implementation.

- As in EURHISFIRM, the pace of progress was slower than desirable (for instance due to limited excess capacity of the WPs parallel to their "own tasks "), we chose to "pivot "on an approach in which, based on prior familiarity and subject matter expertise of the participants, we chose the **concept of committed consensus** and developed an as a starting point an initial raw draft design, comprising the corner points – see chapters 3 and 4, which was subject-wise partitioned and sequenced in a systematic and time-wise dense repetition of discussion, evaluation, and alteration and was – then incrementally – decided.

- We employed the concept of committed consensus for all participants. Structurally, all incremental decisions are subject to an overall evaluation and possible re-specification of already agreed-upon decisions. Our experience confirms the practitioner's propositions that at the beginning of this work sequence the desire to reformulate earlier design decisions is a bit higher, but over the time the participants – probably given the approaching deadline – more and more agree to a stable common view of the subject matters.

- In the course of the incremental commitment, intermediate "consensus-based agreement "sign-off discussions and decisions have been interspersed in order to provide these decisions (for instance the identification architecture across several entity types) longer stability. This transforms later into the distinction of "mission-type decisions "(strategic) which are altered only rather seldom and – within the corner-points of the mission – operative details which may be altered more often.

- To hedge the incremental and consensus-derived decisions, we asked after each decision the EURHISFIRM ExCom to approve the WGIS decisions. The ExCom also periodically involved all WP leaders herein. The socio-technical dimension of this process again is highlighted (aside from the configuration of a European network of EURHISFIRM expert centres). In the end, we need a political body to make knowledgeable decisions which enjoy acceptance with all important stakeholders.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- In the course of the sequence of these sessions, unresolved issues are retained. These will constitute an initial starting point for a successor organisation of WGIS once phase 2 actors commence their work.

Finally, we condense some lessons learnt – crystallised out of our WP 5 and WGIS experience on extendible data models:

- DO Take stock of the comprehensive, and "wide", business requirements of a planned initiative before undertaking more detailed levels of analysis or definition
- Identify the core data structure and the main classes of information objects that support these business requirements, which are central to realising the initiative's goals and express them in a conceptual business model that provides a high-level overview of the data landscape and information ecosystem of the target solution.
- DO Select an open, consensus-based standards development methodology based on an agreement and commitment on the part of members to support the group's decisions, which is "fit for purpose" and appropriate for the background and availability of team resources.
- DO Take into account that data enters a system as input from external sources, is processed or transformed by a system, and is accessed, or output, from a system by end-users for whose use a system is designed to serve.
  - Corollary: DO NOT Assume that data models or standards for data contained or managed by a system are sufficient to describe data that has not been internalised by a system
- DO Consider that more than one physical representation, data storage platform, or data transmission channel is likely to exist in a large system, especially one that comprises a variety of cooperating consortium partners that are typical of EU initiatives.

All in all, this design is "not completed ". Rather, as the process of extendibility is specified, it is an ongoing process, and the result remains "open "– however, our specifications do not limit possible future development directions if the members of the system agree to such development.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 8 References

Bernstein, P.A. (2003) Published. 'Applying Model Management to Classical Meta Data Problems'. *CIDR,* 2003. Citeseer, pp.209-220.

Bottega, J.A. and Powell, L.F. (2012) 'Creating a linchpin for financial data: Toward a universal legal entity identifier'. *Journal of Economics and Business,* 64 (1), pp. 105-115.

Chan, K.K. and Milne, A. (2019) 'The global legal entity identifier system: How can it deliver?'. *Journal of Risk and Financial Management,* 12 (1), pp. 39.

Gallaba, K. (2019) Published. 'Improving the Robustness and Efficiency of Continuous Integration and Deployment'. *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME),* 2019. IEEE, pp.619-623.

Global Legal Entity Identifier Foundations (2020) *About GLEIF: History of the Global LEI System.* Available at: https://www.gleif.org/en/about/history (Accessed: 08.12.2020).

Global Legal Identifier Foundations (2020) *About LEI Code Lists: ISO 20275: Entity Legal Forms Code List.* (Accessed: 09.12.2020).

GS1 (2019) 'Global Standards Management Process (GSMP) Q1 2019 Management Report'. Available at: http://www.gs1.org/gsmp/reports.

Karapanagiotis, P. (2019) 'EURHISFIRM D5.1: Technical document on national data models'. Available at: https://doi.org/10.5281/zenodo.3467926.

Karapanagiotis, P. (2020) 'EURHISFIRM D5.2: Technical Document on Preliminary Common Data Model'. Available at: https://doi.org/10.5281/zenodo.3686930.

Khatri, V. and Brown, C.V. (2010) 'Designing data governance'. *Communications of the ACM,* 53 (1), pp. 148-152.

Poukens, J. (2019) 'EURHISFIRM D4.4: Report on data and sources documentation and quality assessment'. Available at: https://doi.org/10.5281/zenodo.3246465.

Rajput, D. and Khoja, D. (2018) 'CSSO: Ontology-Based Framework for Semantic Interoperatability Between Heterogeneous Web Sources'.

Wilkinson, M.D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship'. *Scientific Data,* 3 (1), pp. 160018.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 9 Appendix

## 9.1 Legal Entity Data Standard 1.01

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# EURHISFIRM Common Data Model

## Legal Entity Data Standard 1.0

Version of 15 January 2020

| Date | Revision | Description |
|---|---|---|
| 2020-01-15 | ver 1.01 | First Draft to be circulated (formatting fixed) |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Abstract

First, the semantic content of these attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the EURHISFIRM Common Data Model.

## Status of this document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this document series will be maintained on SeaFile.

This draft is a **Working Draft** which can be circulated to any interested parties for review and comment. It is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to use Working Drafts as reference material or to cite them as other than "work in progress." This is work in progress and does not imply endorsement by the EURHISFIRM ExCo.

Comments on this document should be sent to [TBD: insert mailing list or URL].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Table of Contents

1 Introduction....5
2 Terminology and Typographical Conventions....6
3 Abstract Data Content....6
3.1 ELEI File Header....6
3.2 ELEI Data Record....7
3.2.1 Entity Section of ELEI Data Record....7
3.2.2 Registration Section of ELEI Data Record....10
3.2.3 Extension Section of ELEI Data record....11
3.3 Data Types....11
3.3.1 Address Data Type....11
3.3.2 CountryCode Data Type....12
3.3.3 DateTime Data Type....12
3.3.4 DSUID Data Type....13
3.3.5 ELEI Data Type....13
3.3.6 EntityCategory Data Type....13
3.3.7 LanguageCode Data Type....13
3.3.8 LegalForm Data Type....14
3.3.9 LegalFormCode Data Type....14
3.3.10 Name Data Type....14
3.3.11 OtherAddress Data Type....14
3.3.12 OtherEntityName Data Type....15
3.3.13 RegionCode Data Type....15
3.3.14 RegistrationAuthority Data Type....15
3.4 Enumerated Code Lists....16
3.4.1 AddressTypeEnum Code List....16
3.4.2 EntityCategoryTypeEnum....16
3.4.3 EntityExpirationReasonEnum Code List....17
3.4.4 EntityNameTypeEnum Code List....17
3.4.5 EntityStatusEnum Code List....18
3.4.6 LegalFormEnum Code List....18



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

3.4.7 RegistrationAuthorityEnum Code List ..... 18
3.4.8 RegistrationStatusEnum Code List ..... 19
3.4.9 ValidationSourcesEnum Code List ..... 20
4 Constraints and Data Validation ..... 21
5 Change Management ..... 21
6 Examples (non-normative) ..... 21
7 References ..... 21
8 Appendix: Character Codes Allowed in Romanized Names ..... 22
9 Appendix: ELEI Record Transition Diagram ..... 24



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 1 Introduction

Legal entities are clearly one of the core classes of objects in a historical financial Research Infrastructure such as EURHISFIRM, on which the majority of other EURHISFIRM financial data depend. As such, legal entities should be uniquely and unambiguously identified in EURHISFIRM, and this identification is accomplished based on the identifying attributes associated with each entity.

First, the semantic content of those attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the reference data attributes of legal entities in the EURHISFIRM Common Data Model

This document proposes initial standards for EURHISFIRM legal entity reference data. It is important that this reference data should uniquely identify the legal entities that are harvested from contributing sources and then collected and assimilated into the common EURHISFIRM platform.

A EURHISFIRM Legal Entity Identifier (ELEI) code that resolves to this entity-identifying reference data is also introduced.

- A partitioning scheme for the structure of this identifier is defined that allows the independent and concurrent minting and assignment of ELEI codes to legal entity reference data that is produced by multiple Data Submission Units (DSUs) in the federated EURHISFIRM Research Infrastructure network.

The standard set by this document is expected, among other things, to reduce the risk of duplicates stemming from differences in formats and conventions of locally sourced data provided by regional contributors, to ensure data quality in the EURHISFIRM system, and to enable the subsequent detection and resolution of multiple identifiers for the same legal entity to the extent that they occur. The standard is expected to be used as a format for reference data consolidated from all sources in order to be promoted to the level of common EURHISFIRM data published for end-user access.

The contents of this document are as follows:

- Section 2 defines terminology and typographical conventions.
- Section 3 specifies the abstract content of EURHISFIRM legal entity reference data conforming to this standard, including a detailed description of each element of reference data associated with legal entities in EURHISFIRM. Allowable values for data elements that are code lists will also be (subsequently) specified. The partitioning scheme for the structure of the EURHISFIRM Legal Enty Identifier (ELEI) code will also be subsequently described.
- Section 4 specifies data validation constraints and other considerations intended to lead to high-quality data content.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- Section 5 specifies a concrete realization of the data definitions above in XML syntax, by means of XML schema (XSD 1.0).
- Section 6 specifies how this file format may be changed in the future, providing for versioning, forward- and backward-compatibility, etc.
- Section 7 provides examples to illustrate the file format.

## 2 Terminology and Typographical Conventions

Within this specification, the terms SHALL, SHALL NOT, SHOULD, SHOULD NOT, MAY, NEED NOT, CAN, and CANNOT are to be interpreted as specified in Annex G of the ISO/IEC Directives, Part 2, 2001, 4th edition [ISODir2]. When used in this way, these terms will always be shown in ALL CAPS; when these words appear in ordinary typeface they are intended to have their ordinary English meaning.

All sections of this document, with the exception of Section 1 are normative, except where explicitly noted as non-normative.

The following typographical conventions are used throughout the document:

- ALL CAPS type is used for the special terms from [ISODir2] enumerated above.
- `Monospace` type is used to denote programming language, UML, and XML identifiers, as well as for the text of XML documents.

➢ Placeholders for open issues and/or changes that need to be made to this document prior to its reaching the final stage of approved Proposed Standard are prefixed by a rightward-facing arrowhead, as this paragraph is.

## 3 Abstract Data Content

This section specifies the abstract data content of a a data file conforming to this standard.

A data file conforming to this standard SHALL consist of:

- An optional ELEI_FileHeader, as specified in Section 3.1.
- Zero or more ELEI Data Records, as specified in Section 3.2.

### 3.1 ELEI File Header

➢ TBD Define elements to go into a "header" area for an ELEI file. The purpose of the header is to provide context about the file and its contained ELEI data records. However, the header will not contain anything necessary to interpret the meaning of any ELEI record; e.g., things like default values for ELEI data records will not be in the header (such things would mean that the meaning of an ELEI record could change if taken away from the header). Examples of things that might be useful to include in the header:

➢ A DateTime value indicating when the file was created and/or the "as of" date of the file's contents



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- If the file came from a Data Submission Unit (DSU) (e.g., a file published by a DSU), the DSUID of the originator
- If the file is a regularly published file from a DSU, a code to indicate whether the file is a "full" file, a "delta" file (along with the date to which the delta is relative), a response to a query for ELEI records matching specified criteria, etc.

## 3.2 ELEI Data Record

An ELEI Data Record describes a single ELEI. Each ELEI Data record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ELEI | ELEI | 1 | The 20-character ELEI of the legal entity described by this ELEI Data Record. |
| Entity | Entity (Section 3.2.1) | 1 | Attributes describing the entity itself |
| Registration | Registration (Section 3.2.2) | 1 | Attributes describing the registration of this ELEI. |
| Extension | Extension (Section 3.2.3) | 0,1 | An optional element for including data beyond the standard data elements in an ELEI data file. This may include data specific to an DSU, data specific to a publisher of ELEI data, and so on. |

### 3.2.1 Entity Section of ELEI Data Record

The Entity section of an ELEI Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| LegalName | Name | 1 | The Legal Name of the Entity.<br>If an Entity is in a jurisdiction with more than one Legal Name (e.g., in different languages), this is the Primary Legal Name (see otherEntityNames for other names). |
| OtherEntityNames | Other EntityName | 0..n | An optional list of other Name instances for the Entity. |
| LegalAddress | Address | 1 | The address of the Entity as recorded in the registration of the Entity in its legal jurisdiction |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| HeadquartersAddress | Address | 1 | The address of the headquarters of the Entity |
| OtherAddresses | OtherAddress | 0..n | An optional list of other Address instances for the Entity. This may be used to provide alternative language forms of legal address or headquarters address.. |
| RegistrationAuthority | RegistrationAuthorityEnum | 0,1 | A code that identifies the business register or other registration authority that supplied the value of RegistrationAuthorityEntityID. |
| RegistrationAuthorityEntityID | String | 0,1 | The identifier of the legal entity as maintained by a business registry in the jurisdiction of legal registration,<br>OR<br>If the entity is one that is not recorded in a business registry (e.g., one of the varieties of funds registered instead with financial regulators), the identifier of the entity in the appropriate registration authority.<br>The RegistrationAuthorityEntityID element SHALL be included if RegistrationAuthority is included, and SHALL be omitted if RegistrationAuthority is omitted. |
| LegalJurisdiction | RegionCode | 0,1 | The jurisdiction of legal formation and registration of the entity (and on which the legalForm data element is also dependent). |
| EntityCategory | EntityCategoryTypeEnum | 0,1 | Indicates the general category of the type of entity identified by this ELEI data record |
| LegalForm | LegalFormType | 0,1 | The legal form of the entity, taken from the ISO Entity Legal Form (ELF) code list maintained by GLEIF. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityStatus | EntityStatusEnum | 1 | The status of the legal entity. This is not to be confused with the status of the registration; see RegistrationStatus.<br>If this ELEI record contains a non-empty SuccessorELEI field, EntityStatus is the last status of the legal entity before the successor ELEI record superseded this one, which is not necessarily the current status of the legal entity. |
| EntityExpirationDate | DateTime | 0,1 | The date that the legal entity ceased to operate,, whether due to dissolution, merger or acquisition. Omitted if the legal entity has not ceased to operate, or if this ELEI record contains a non-empty SuccessorELEI field. |
| EntityExpiration Reason | Entity Expiration ReasonEnum | 0,1 | The reason that a legal entity ceased to operate. This element SHALL be present if EntityExpirationDate is present, and omitted otherwise. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| SuccessorELEI | ELEI | 0,1 | The ELEI of the ELEI registration that supersedes or subsumes this ELEI registration for the same legal entity.<br>If RegistrationStatus is DUPLICATE, then SuccessorELEI is the ELEI of the surviving ELEI Registration.<br>If RegistrationStatus is MERGED, then SuccessorELEI is the ELEI Registration of the new/acquiring entity.<br>Otherwise, SuccessorELEI is omitted.<br>When a successor ELEI is assigned to an ELEI registration, the ELEI registration will no longer be updated (since another ELEI registration has superseded the registration that has just been assigned a successor).<br>The ELEIRecordLastUpdate that is recorded with the update assigning a SuccessorELEI will be the last update performed on the superseded registration record.<br>As a consequence, other fields of the superseded registration record (address, entity status, etc) may no longer reflect the actual state of the legal entity. |

### 3.2.2 Registration Section of ELEI Data Record

The Registration section of an ELEI Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| InitialRegistrationDate | DateTime | 1 | Date/time the ELEI record was created |
| LastUpdateDate | DateTime | 1 | Date/time the ELEI record was most recently updated |
| RegistrationStatus | Registration StatusEnum | 1 | Status of the ELEI registration. This is not to be confused with the status of the legal entity itself; see EntityStatus. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| NextRenewalDate | DateTime | 1 | The next date by which the ELEI registration should be renewed and re-certified by the legal entity. |
| ResponsibleDSU | DSUID | 1 | The Identifier of the Data Submission Unit (DSU) that was responsible for producing this ELEI registration. |
| ValidationSources | ValidationSourcesEnum | 0,1 | The current validation status of this ELEI record, or omitted if the validation status is not known or not revealed. |

### 3.2.3 Extension Section of ELEI Data record

The Extension section of an ELEI record may be used to include additional data not defined in this standard. For example, an DSU may use Extension to publish additional data elements it collects as part of registration.

- TBD: include the details of how this works. Basically, the idea is to use an XSD schema wildcard with namespace ##other, permitting the inclusion of XML elements from other XML namespaces.

## 3.3 Data Types

This section specifies the data types referenced by the tables in Section 3.2, in alphabetical order.

### 3.3.1 Address Data Type

A value of type Address in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Cardinality | Description |
|---|---|---|---|
| lang | LanguageCode | 0,1 | The language in which all of the string-valued components of this address are expressed. |
| Line1 | String | 1 | The first line of the street address |
| Line2 | String | 0,1 | The second line of the street address |
| Line3 | String | 0,1 | The third line of the street address. This element SHALL be omitted if addressLine2 is omitted. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Cardinality | Description |
|---|---|---|---|
| Line4 | String | 0,1 | The fourth line of the street address. This element SHALL be omitted if addressLine3 is omitted. |
| City | String | 1 | The name of the city |
| Region | RegionCode | 0,1 | The code for a region (state, province, county, parish, etc) as used in postal addresses for the country of this address. |
| Country | CountryCode | 1 | The 2-character ISO 3166-1 country code of the country |
| PostalCode | String | 0,1 | The postal code of this address as specified by the local postal service. |

- TBD: should postal code be optional?

### 3.3.2 CountryCode Data Type

A value of type CountryCode in a file conforming to this standard SHALL be a 2-character country code conforming to ISO ISO 3166-1 alpha-2 [ISO3166]. Note that ISO 3166-1 alpha-2 codes are all uppercase.

### 3.3.3 DateTime Data Type

A value of type DateTime in a conforming to this standard SHALL be a point in time expressed as a string conforming to ISO 8601 having the following format:

*YYYY*-*MM*-*DD*T*hh*:*mm*:*ss.sss*TZ

where the components of the above string are as follows:

- *YYYY* is the year
- *MM* is the month (01 = January, …, 12 = December)
- *DD* is the day of the month (01 = first day of the month)
- T is the single character 'T'
- *hh* is the hour (00 – 23)
- *mm* is the minute
- *ss.sss* is the second and milliseconds. From one to three digits may be used for milliseconds, or omitted entirely along with the decimal point.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- *TZ* is the time zone specifier, which can be either:
  - `Z` the single character 'Z', denoting Coordinated Universal Time (UTC); or
  - *+hh:mm* denoting a positive offset from UTC; or
  - *-hh:mm* denoting a negative offset from UTC

In the XML representation specified in Section **Error! Reference source not found.**, the XSD t ype `xs:dateTime` is used; however, whereas `xs:dateTime` permits the time zone specifier to be omitted, `DateTime` values in files conforming to this standard SHALL always include a time zone specifier.

*Explanation (non-normative): milliseconds are hardly necessary for ELEI reference data, and likewise it might seem simpler to allow only "Z" as a time zone specifier; however, XML processing tools support the full syntax given above and it is not always possible to restrict such tools to avoid milliseconds or force the use of "Z" as the time zone specifier. The restriction that the time zone specifier must be present is equivalent to using XSD type `xs:timestamp`; however this was introduced in XSD 1.1 and not supported by the majority of XML processing tools which still only implement XSD 1.0.*

### 3.3.4 DSUID Data Type

A value of type `DSUID` in a file conforming to this standard SHALL be a [TBD]-character Data Submission Unit Identifier conforming to [TBD].

### 3.3.5 ELEI Data Type

A value of type `ELEI` in a file conforming to this standard SHALL be a 20-character Legal Entity Identifier conforming to [ISO17422].

### 3.3.6 EntityCategory Data Type

A value of type `EntityCategory` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `EntityCategory` | `EntityCategoryTypeEnum` | 0,1 | Indicates the general category of the type of entity identified by this ELEI data record |

### 3.3.7 LanguageCode Data Type

A value of type `LanguageCode` in a file conforming to this standard SHALL be a 2-character language code conforming to [ISO639-1]. Note that ISO 639-1 language codes are all lowercase.

➢ TBD: Alternatively, we could use IETF language tags (RFC 4646) instead. IETF language tags can distinguish between variations of the same language in different countries. For



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

example, the IETF language tags `fr-CA` and `fr-FR` denote Canadian French and French as spoken in France, respectively; in ISO 639-1 these would both be simply `fr`.

### 3.3.8 LegalForm Data Type

A value of type `LegalForm` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `LegalForm` | `LegalFormCode` | 1,1 | The legal form of the entity |
| `OtherLegalForm` | `String` | 0,1 | Interim free-text legal form information in the process of transition to an ELF standard code |

### 3.3.9 LegalFormCode Data Type

`LegalFormCode` is a code that is a normative enumeration from the ISO Entity Legal Form (ELF) code list maintained by GLEIF.

A value of type `LegalForm` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `LegalFormCode` | `LegalFormEnum` | 1,1 | The legal form of the entity |

### 3.3.10 Name Data Type

A `Name` is a string expressed in a natural language, including a code indicating which natural language is used.

A value of type `Name` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `lang` | `LanguageCode` | 0,1 | The language of `name` |
| `Name` | `String` | 1 | The name itself. |

### 3.3.11 OtherAddress Data Type

A value of type `OtherAddress` in a file conforming to this standard SHALL include data elements as specified below. Each `Address` element includes an optional language code, permitting `OtherAddress` to be repeated as many times as necessary to express the same address type in multiple languages. The purpose of the `AddressTypeEnum` code list is to



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

accommodate legal address and headquarters address in different languages, not to add other address types (which could conceivably be added in the future).

| Element Name | Type | Card | Description |
|---|---|---|---|
| type | AddressTypeEnum | 1 | The type of address represented by this OtherAddress instance. |
| Address | Address | 1 | The address |

### 3.3.12 OtherEntityName Data Type

A value of type OtherEntityName in a file conforming to this standard SHALL include data elements as specified below. Each Name element includes an optional language code, permitting OtherEntityName to be repeated as many times as necessary to express the same name type in multiple languages. When type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, the language code specifies the language of the name prior to Romanization.

| Element Name | Type | Card | Description |
|---|---|---|---|
| type | EntityNameTypeEnum | 1 | The type of name represented by this OtherEntityName instance. |
| Name | Name | 1 | The name. If type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, then this value SHALL only include characters from the character set specified in Section 8. |

### 3.3.13 RegionCode Data Type

A value of type RegionCode in a file conforming to this standard SHALL be a code conforming to ISO 3166-2. Note that ISO 3166-2 codes are all uppercase.

### 3.3.14 RegistrationAuthority Data Type

A value of type RegistrationAuthority in a file conforming to this standard SHALL include data elements as specified below.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| RegistrationAuthorityID | RegistrationAuthorityEnum | 0,1 | An identifier for the legal entity **registry** of the entity in the jurisdiction of legal registration, or in the appropriate registration authority. |
| RegistrationAuthorityEntityID | String | 0,1 | An identifier for the **legal entity** at the registry in the jurisdiction of legal registration, or in the appropriate registration authority. |

## 3.4 Enumerated Code Lists

This section specifies the enumerated code list data types (all having the suffix Enum) referenced by the tables in Sections 3.2 and 3.2.2, in alphabetical order.

### 3.4.1 AddressTypeEnum Code List

The AddressTypeEnum value in an OtherAddress instance specifies how the alternative address relates to the legal entity.

A value of type AddressTypeEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| LEGAL_ADDRESS | Registered address of the entity in the legal jurisdiction |
| HEADQUARTERS_ADDRESS | Address of the headquarters of the entity |

### 3.4.2 EntityCategoryTypeEnum

The EntityCategoryTypeEnum value in an ELEI record specifies the general classification category of the legal entity.

- TBD: the table below is just an example with some possible categories, and is incomplete, and needs to be elaborated.

A value of type EntityCategoryTypeEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| PUBLIC_CORPORATION | Public shareholder entity |
| PRIVATE_CORPORATION | Privately owned limited liability corporation |
| LLC | Privately owned limited liability company |
| SOLE_PROPRIETOR | Private business owned and operated by a single individual |
| LLP | Privately owned limited liability partnership |
| TRUST | Legal trust |
| FUND | Alternative Investment Scheme |
| GSE | Government sponsored entity |
| NGO | Non-Government Organization |
| NGO | Non-Government Organization |

### 3.4.3 EntityExpirationReasonEnum Code List

The `EntityExpirationReasonEnum` value in an ELEI record specifies the reason that the legal entity expired.

A value of type `EntityExpirationReasonEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| DISSOLVED | The entity ceased to operate |
| CORPORATE_ACTION | The entity was acquired or merged with another entity |
| OTHER | The reason for expiry is not one of the above |

### 3.4.4 EntityNameTypeEnum Code List

The `EntityNameTypeEnum` value in an `EntityName` specifies how the name relates to the legal entity.

A value of type `EntityNameTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| OTHER_LEGAL | Registered name of the entity in an alternate language in the legal jurisdiction in which the entity is registered |
| PREFERRED_ROMANIZED_LEGAL | Romanized form, preferred by the entity |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| AUTO_ROMANIZED_LEGAL | Romanized form, auto-transliterated |

### 3.4.5 EntityStatusEnum Code List

The EntityStatusEnum value in an ELEI record indicates the status of the legal entity itself. This is not to be confused with the status of the ELEI registration, which is specified by RegistrationStatusEnum (Section 3.4.8). See also Section 9, which illustrates how the EntityStatusEnum value changes over the lifecycle of an ELEI registration.

A value of type EntityStatusEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| ACTIVE | As of the last report or update, the legal entity reported is legally registered and operating. |
| INACTIVE | It has been determined that the entity that was assigned the ELEI is no longer legally registered and/or operating, whether as a result of:<br>1. Business closure<br>2. Acquisition by or merger with another (or new) entity<br>3. Determination of illegitimacy |
| NOT_SPECIFIED | The ELEI record is in a state in which it does not provide information about whether the legal entity is legally registered and operating. |

### 3.4.6 LegalFormEnum Code List

The value of LegalFormEnum SHALL be a code published in the ISO Entity Legal Form (ELF) code list maintained by GLEIF.

### 3.4.7 RegistrationAuthorityEnum Code List

The RegistrationAuthorityEnum value in an Entity instance specifies what business register provided the value of RegistrationAuthorityEntityID for the legal entity.

The value of RegistrationAuthorityEnum SHALL be a code published in a list of registration authority codes.

- TBD: the external code list for Registration Authorities needs to be developed, but will borrow from, incorporate or reference the Registration Authorities Code List published by the Global LEI Foundation.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.8 RegistrationStatusEnum Code List

The `RegistrationStatusEnum` value in an ELEI record indicates the status of the registration of the legal entity with an DSU. This is not to be confused with the status of the ELEI itself, which is specified by `EntityStatusEnum` (Section 3.4.5). See also Section 9, which illustrates how the `EntityStatusEnum` value changes over the lifecycle of an ELEI registration.

A value of type `RegistrationStatusEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `PENDING` | An application for an ELEI that has been submitted and which is being processed and validated.<br>*NOTE: ELEI registrations in the `PENDING` state are not intended for public release, but could be used internally between DSUs.* |
| `ISSUED` | An ELEI Registration that has been validated and issued, and which identifies an entity that was an operating legal entity as of the last update. |
| `DUPLICATE` | An ELEI Registration that has been determined to be a duplicate registration of the same legal entity as another ELEI Registration; the `DUPLICATE` status is assigned to the non-surviving registration (i.e., the ELEI that should no longer be used). Only one of the potential multiple identifiers will survive; for all other duplicate registrations:<br>1. The `RegistrationStatus` is set to `DUPLICATE`,<br>2. The ELEI of the surviving ELEI Registration is set in the `SuccessorELEI` data element of (each) duplicate ELEI registration;<br>3. The `LastUpdateDate` is set to reflect the date of this update, and<br>4. No further updates of the `DUPLICATE` registration record will occur. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| MERGED | An ELEI registration for an entity that has been merged into another legal entity, such that this legal entity no longer exists as an operating entity.<br>If<br>• After being issued an ELEI, the entity is acquired by, or merged with, another legal entity;<br>• Per agreements among the parties to the transaction, the ELEI of the acquired or merged entity will not be used to identify the surviving entity (or if a new entity is created that is issued a new ELEI)<br>Then<br>1. The ELEIRegistrationStatus is set to "MERGED",<br>2. The ELEI of the surviving/new legal entity is set in the successorELEI data element of (each) ELEI registration that is no longer to be used;<br>3. The ELEIRecordLastUpdate is set to reflect the date of this update, and<br>4. No further updates of the MERGED registration record(s) will occur. |
| RETIRED | An ELEI registration for an entity that has ceased operation, without having been merged into another entity.<br>If<br>• The responsible DSU determines by public sources that the legal entity has been dissolved or ceased to operate (and the DSU seeks to confirm this status through all available channels)<br>Then<br>1. The ELEIRegistrationStatus is set to "RETIRED";<br>2. The ELEIRecordLastUpdate is set to reflect the date of this update;<br>3. The EntityExpirationDate is also set to the date of this update;<br>4. The EntityExpirationReason is set;<br>5. No further updates of the RETIRED registration record will occur. |

### 3.4.9 ValidationSourcesEnum Code List

A value of type ValidationSourcesEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| PENDING | The validation of the reference data for the entity has not yet occurred. |
| SUBMITTED | Based on the validation procedures in use by the DSU responsible for the record, the information associated with this record has significant reliance on the information that a submitter provided due to the unavailability of corroborating information. |
| PARTIALLY_CORROBORATED | Based on the validation procedures in use by the DSU responsible for the record, the information supplied for the entity can be partially corroborated by available sources, while some of the record is dependent upon information that the submitter collected, either due to conflicts with authoritative information, or due to data unavailability. |
| FULLY_CORROBORATED | Based on the validation procedures in use by the DSU responsible for the record, there is sufficient information contained in authoritative public sources to corroborate the information regarding the entity provided in the record. |

.

## 4 Constraints and Data Validation

All values of type `String` specified in Section 3 SHALL be 500 or fewer characters in length.

- TBD. Specify here any additional constraints not implicit in the data types or explicit in the definition of each reference data element

## 5 Change Management

- TBD. Use the <vnext> strategy for forward/backward compatibility, and/or include a version number in the XML. Needs to be explained

## 6 Examples (non-normative)

- TBD.

## 7 References

[ISO646] ISO, "Information technology -- ISO 7-bit coded character set for information interchange," ISO/IEC 646:1991.

[ISO3166-1] ISO, "Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes," ISO 3166-1:2013.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

[ISO3166-2] ISO, "Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code," ISO 3166-2:2013.

[ISO8601] ISO, "Data elements and interchange formats – Information interchange – Representation of dates and times," ISO 8601:2004.

[ISO10646] ISO, "Information technology -- Universal Coded Character Set (UCS)," ISO 10646:2012.

[ISO17422] ISO, "Financial Services - Legal Entity Identifier (LEI)," ISO/DIS 17442:2012.

[ISODir2] ISO, "Rules for the structure and drafting of International Standards (ISO/IEC Directives, Part 2, 2001, 4th edition)," July 2002.

[XSD1] H. Thompson, D. Beech, M. Maloney, N. Mendelsohn, "XML Schema Part 1: Structures," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-1/.

[XSD2] P. Biron, A. Malhotra, "XML Schema Part 2: Datatypes," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-2/.

# 8 Appendix: Character Codes Allowed in Romanized Names

When a `Name` instance is of type `PREFERRED_ROMANIZED` or `AUTO_ROMANIZED`, the value of the `name` field SHALL consist only of non-control characters drawn from the "invariant subset" of ISO 646. These characters are enumerated below. The "Hex Value" column indicates the code point value (expressed in hexadecimal) for each character in both ISO 646 and ISO 10646.

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| ! | Exclamation Mark | 21 | M | Capital Letter M | 4D |
| " | Quotation Mark | 22 | N | Capital Letter N | 4E |
| % | Percent Sign | 25 | O | Capital Letter O | 4F |
| & | Ampersand | 26 | P | Capital Letter P | 50 |
| ' | Apostrophe | 27 | Q | Capital Letter Q | 51 |
| ( | Left Parenthesis | 28 | R | Capital Letter R | 52 |
| ) | Right Parenthesis | 29 | S | Capital Letter S | 53 |
| * | Asterisk | 2A | T | Capital Letter T | 54 |
| + | Plus sign | 2B | U | Capital Letter U | 55 |
| , | Comma | 2C | V | Capital Letter V | 56 |
| - | Hyphen/ Minus | 2D | W | Capital Letter W | 57 |
| . | Full Stop | 2E | X | Capital Letter X | 58 |
| / | Solidus | 2F | Y | Capital Letter Y | 59 |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

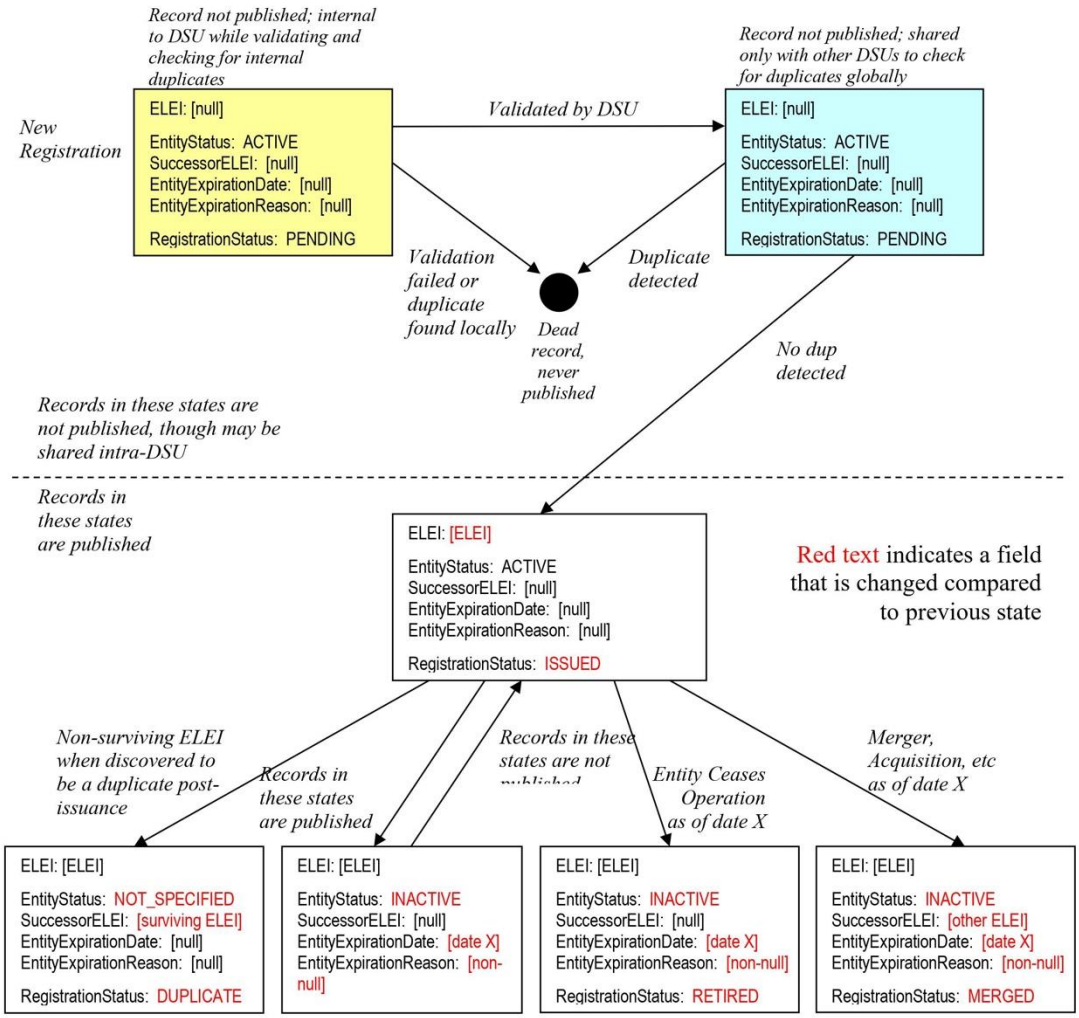
| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| 0 | Digit Zero | 30 | Z | Capital Letter Z | 5A |
| 1 | Digit One | 31 | _ | Low Line | 5F |
| 2 | Digit Two | 32 | a | Small Letter a | 61 |
| 3 | Digit Three | 33 | b | Small Letter b | 62 |
| 4 | Digit Four | 34 | c | Small Letter c | 63 |
| 5 | Digit Five | 35 | d | Small Letter d | 64 |
| 6 | Digit Six | 36 | e | Small Letter e | 65 |
| 7 | Digit Seven | 37 | f | Small Letter f | 66 |
| 8 | Digit Eight | 38 | g | Small Letter g | 67 |
| 9 | Digit Nine | 39 | h | Small Letter h | 68 |
| : | Colon | 3A | i | Small Letter i | 69 |
| ; | Semicolon | 3B | j | Small Letter j | 6A |
| < | Less-than Sign | 3C | k | Small Letter k | 6B |
| = | Equals Sign | 3D | l | Small Letter l | 6C |
| > | Greater-than Sign | 3E | m | Small Letter m | 6D |
| ? | Question Mark | 3F | n | Small Letter n | 6E |
| A | Capital Letter A | 41 | o | Small Letter o | 6F |
| B | Capital Letter B | 42 | p | Small Letter p | 70 |
| C | Capital Letter C | 43 | q | Small Letter q | 71 |
| D | Capital Letter D | 44 | r | Small Letter r | 72 |
| E | Capital Letter E | 45 | s | Small Letter s | 73 |
| F | Capital Letter F | 46 | t | Small Letter t | 74 |
| G | Capital Letter G | 47 | u | Small Letter u | 75 |
| H | Capital Letter H | 48 | v | Small Letter v | 76 |
| I | Capital Letter I | 49 | w | Small Letter w | 77 |
| J | Capital Letter J | 4A | x | Small Letter x | 78 |
| K | Capital Letter K | 4B | y | Small Letter y | 79 |
| L | Capital Letter L | 4C | z | Small Letter z | 7A |
| | Space | 20 | | | |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 9 Appendix: ELEI Record Transition Diagram

The following diagram illustrates the relationship between status fields of the ELEI record.




In rare circumstances, a record may pass directly from `PENDING` to `RETIRED` or `MERGED`; e.g., retroactive reporting of a failed or merged entity. In that case, the `EntityStatus` would be `INACTIVE` while the record is in the `PENDING` state.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 777489

## 9.2 Legal Entity Data Standard 1.05

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# EURHISFIRM Common Data Model

## Legal Entity Data Standard 1.0

Version of 04 March 2020

| Date | Revision | Description |
|---|---|---|
| 2020-01-15 | ver 1.01 | First Draft to be circulated (formatting fixed) |
| 2020-02-06 | ver 1.02 | Added partitioning scheme for ELEI in Appendix 11;<br>made **`Line1`** of the street address in the **`Address`** type optional;<br>qualified the term "entity" with "legal entity" throughout;<br>revised the legal entity record transition diagram appendix |
| 2020-02-19 | ver 1.03 | Added `LegalEntityCreationDate` start date for legal entity to complement the `LegalEntityExpirationDate`<br>Added `RegistrationUpToDate` and `RegistrationAsOfDate` to bracket the start and stop dates that this reference data is valid<br>Added `RegistrationSourceReference` in order to provide a reference to the definitive data source from which this data was collected<br>Added several revisions to element definitions |
| 2020-02-20 | ver 1.04 | Added ELEI Root record (needed to anchor the unique ELEI in order to support multiple (historical) ELEI Data records |
| 2020-03-04 | ver 1.05 | Revised the ELEI root record to include `EarliestRegistrationDate, LatestRegistrationDate. LegalEntityStatus`<br>Added `LegalEntityEvents` to ELEI data record<br>Revised `CountryCode` to be `ECountryCode` and `RegionCode` to be `ERegionCode` in order to include historical countries and regions |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Abstract

First, the semantic content of these attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the EURHISFIRM Common Data Model.

## Status of this document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this document series will be maintained on SeaFile.

This draft is a **Working Draft** which can be circulated to any interested parties for review and comment. It is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to use Working Drafts as reference material or to cite them as other than "work in progress." This is work in progress and does not imply endorsement by the EURHISFIRM ExCo.

Comments on this document should be sent to [TBD: insert mailing list or URL].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Table of Contents

1 Introduction 5
2 Terminology and Typographical Conventions 6
3 Abstract Data Content 6
3.1 ELEI File Header 6
3.2 ELEI Root Record 7
3.3 ELEI Data Record 8
3.3.1 Legal Entity Section of ELEI Data Record 9
3.3.2 Registration Section of ELEI Data Record 12
3.3.3 Extension Section of ELEI Data record 14
3.4 Data Types 14
3.4.1 Address Data Type 14
3.4.2 CountryCode Data Type 15
3.4.3 DateTime Data Type 15
3.4.4 DSUID Data Type 16
3.4.5 ELEI Data Type 16
3.4.6 LegalEntityCategory Data Type 16
3.4.7 LanguageCode Data Type 16
3.4.8 LegalForm Data Type 16
3.4.9 LegalFormCode Data Type 17
3.4.10 Name Data Type 17
3.4.11 OtherAddress Data Type 17
3.4.12 OtherLegalEntityName Data Type 17
3.4.13 RegionCode Data Type 18
3.4.14 RegistrationAuthority Data Type 18
3.5 Enumerated Code Lists 19
3.5.1 AddressTypeEnum Code List 19
3.5.2 LegalEntityCategoryTypeEnum 19
3.5.3 LegalEntityExpirationReasonEnum Code List 20
3.5.4 LegalEntityNameTypeEnum Code List 20
3.5.5 LegalEntityStatusEnum Code List 21



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

3.5.6 LegalFormEnum Code List ..... 21
3.5.7 RegistrationAuthorityEnum Code List ..... 21
3.5.8 RegistrationStatusEnum Code List ..... 22
3.5.9 ValidationSourcesEnum Code List ..... 23
4 Constraints and Data Validation ..... 24
5 XML Syntax ..... **Error! Bookmark not defined.**
5.1 XML Design Rules ..... **Error! Bookmark not defined.**
5.2 XML Schema ..... **Error! Bookmark not defined.**
6 Change Management ..... 24
7 Examples (non-normative) ..... 24
8 References ..... 24
9 Appendix: Character Codes Allowed in Romanized Names ..... 25
10 Appendix: ELEI Record Transition Diagram ..... 28
11 Appendix: ELEI Code Partitioning Scheme ..... 29



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 1 Introduction

Legal entities are clearly one of the core classes of objects in a historical financial Research Infrastructure such as EURHISFIRM, on which the majority of other EURHISFIRM financial data depend. As such, legal entities should be uniquely and unambiguously identified in EURHISFIRM, and this identification is accomplished based on the identifying attributes associated with each legal entity.

First, the semantic content of those attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the reference data attributes of legal entities in the EURHISFIRM Common Data Model

This document proposes initial standards for EURHISFIRM legal entity reference data. It is important that this reference data should uniquely identify the legal entities that are harvested from contributing sources and then collected and assimilated into the common EURHISFIRM platform.

A EURHISFIRM Legal Entity Identifier (ELEI) code that resolves to this legal entity-identifying reference data is also introduced.

- A partitioning scheme for the structure of this identifier is defined that allows the independent and concurrent minting and assignment of ELEI codes to legal entity reference data that is produced by multiple Data Submission Units (DSUs) in the federated EURHISFIRM Research Infrastructure network.

The standard set by this document is expected, among other things, to reduce the risk of duplicates stemming from differences in formats and conventions of locally sourced data provided by regional contributors, to ensure data quality in the EURHISFIRM system, and to enable the subsequent detection and resolution of multiple identifiers for the same legal entity to the extent that they occur. The standard is expected to be used as a format for reference data consolidated from all sources in order to be promoted to the level of common EURHISFIRM data published for end-user access.

The contents of this document are as follows:

- Section 2 defines terminology and typographical conventions.
- Section 3 specifies the abstract content of EURHISFIRM legal entity reference data conforming to this standard, including a detailed description of each element of reference data associated with legal entities in EURHISFIRM. Allowable values for data elements that are code lists will also be (subsequently) specified. The partitioning scheme for the structure of the EURHISFIRM Legal Enty Identifier (ELEI) code will also be subsequently described.
- Section 4 specifies data validation constraints and other considerations intended to lead to high-quality data content.
- Section 5 specifies a concrete realization of the data definitions above in XML syntax, by means of XML schema (XSD 1.0).



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- Section 6 specifies how this file format may be changed in the future, providing for versioning, forward- and backward-compatibility, etc.
- Section 7 provides examples to illustrate the file format.

## 2 Terminology and Typographical Conventions

Within this specification, the terms SHALL, SHALL NOT, SHOULD, SHOULD NOT, MAY, NEED NOT, CAN, and CANNOT are to be interpreted as specified in Annex G of the ISO/IEC Directives, Part 2, 2001, 4th edition [ISODir2]. When used in this way, these terms will always be shown in ALL CAPS; when these words appear in ordinary typeface they are intended to have their ordinary English meaning.

All sections of this document, with the exception of Section 1 are normative, except where explicitly noted as non-normative.

The following typographical conventions are used throughout the document:

- ALL CAPS type is used for the special terms from [ISODir2] enumerated above.
- `Monospace` type is used to denote programming language, UML, and XML identifiers, as well as for the text of XML documents.

➢ Placeholders for open issues and/or changes that need to be made to this document prior to its reaching the final stage of approved Proposed Standard are prefixed by a rightward-facing arrowhead, as this paragraph is.

## 3 Abstract Data Content

This section specifies the abstract data content of a a data file conforming to this standard.

A data file conforming to this standard SHALL consist of:

- An optional ELEI File Header, as specified in Section 3.1.
- Zero or more ELEI Data Records, as specified in Section 3.2.

### 3.1 ELEI File Header

➢ TBD Define elements to go into a "header" area for an ELEI file. The purpose of the header is to provide context about the file and its contained ELEI data records. However, the header will not contain anything necessary to interpret the meaning of any ELEI record; e.g., things like default values for ELEI data records will not be in the header (such things would mean that the meaning of an ELEI record could change if taken away from the header). Examples of things that might be useful to include in the header:

➢ A DateTime value indicating when the file was created and/or the "as of" date of the file's contents

➢ If the file came from a Data Submission Unit (DSU) (e.g., a file published by a DSU), the DSUID of the originator



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- If the file is a regularly published file from a DSU, a code to indicate whether the file is a "full" file, a "delta" file (along with the date to which the delta is relative), a response to a query for ELEI records matching specified criteria, etc.

## 3.2 ELEI Root Record

The ELEI Root Record anchors a single ELEI. This root record is needed in order to provide a unique ELEI anchor to support multiple (historical) ELEI Data Records. Each ELEI Root record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ELEI | ELEI | 1 | The 20-character ELEI of the legal entity described by this ELEI Data Record. |
| SuccessorELEI | ELEI | 0,1 | The ELEI of the ELEI registration that supersedes or subsumes this ELEI registration for the same legal entity.<br>If RegistrationStatus is DUPLICATE, then SuccessorELEI is the ELEI of the surviving ELEI Registration.<br>If RegistrationStatus is MERGED, then SuccessorELEI is the ELEI Registration of the new/acquiring legal entity.<br>Otherwise, SuccessorELEI is omitted.<br>When a successor ELEI is assigned to an ELEI registration, the ELEI registration will no longer be updated (since another ELEI registration has superseded the registration that has just been assigned a successor).<br>The ELEIRecordLastUpdate that is recorded with the update assigning a SuccessorELEI will be the last update performed on the superseded registration record.<br>As a consequence, other fields of the superseded registration record (address, legal entity status, etc) may no longer reflect the actual state of the legal entity. |
| ELEICreationDate | DateTime | 1 | Date/time the ELEI root record was initially created in the system |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| LegalEntityStatus | LegalEntityStatusEnum | 1 | The status of the legal entity. This is not to be confused with the status of the registration; see RegistrationStatus.<br>If this ELEI record contains a non-empty SuccessorELEI field, LegalEntityStatus is the last status of the legal entity before the successor ELEI record superseded this one, which is not necessarily the current status of the legal entity. |
| LegalEntityFormationDate | DateTime | 0,1 | Date/time the Legal Entity was formed, if known |
| LegalEntityExpirationDate | DateTime | 0,1 | The date that the legal entity ceased to operate, whether due to dissolution, merger or acquisition. Omitted if the legal entity has not ceased to operate, or if this ELEI record contains a non-empty SuccessorELEI field. |
| LegalEntityExpirationReason | LegalEntity Expiration ReasonEnum | 0,1 | he reason that a legal entity ceased to operate. This element SHALL be present if LegalEntityExpirationDate is present, and omitted otherwise. |
| EarliestRegistrationDate | DateTime | 1 | Date/time of the Registration record with the earliest historical date |
| LatestRegistrationDate | DateTime | 1 | Date/time of the Registration record with the most recent historical date |

## 3.3 ELEI Data Record

An ELEI Data Record describes a single ELEI. Each ELEI Data record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ELEI | ELEI | 1 | The 20-character ELEI of the legal entity described by this ELEI Data Record. |
| LegalEntity | LegalEntity (Section 3.3.1) | 1 | Attributes describing the legal entity itself |
| Registration | Registration (Section 0) | 1 | Attributes describing the registration of this ELEI. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| Extension | Extension (Section 3.3.3) | 0,1 | An optional element for including data beyond the standard data elements in an ELEI data file. This may include data specific to an DSU, data specific to a publisher of ELEI data, and so on. |

### 3.3.1 Legal Entity Section of ELEI Data Record

The Entity section of an ELEI Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| LegalName | Name | 1 | The Legal Name of the Entity.<br>If an Entity is in a jurisdiction with more than one Legal Name (e.g., in different languages), this is the Primary Legal Name (see otherEntityNames for other names). |
| OtherEntityNames | Other EntityName | 0..n | An optional list of other Name instances for the Entity. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| LegalAddress | Address | 1 | The address of the Entity as recorded in the registration of the Entity in its legal jurisdiction |
| HeadquartersAddress | Address | 0,1 | The address of the headquarters of the Entity |
| OtherAddresses | OtherAddress | 0..n | An optional list of other Address instances for the Entity. This may be used to provide alternative language forms of legal address or headquarters address.. |
| RegistrationAuthority | RegistrationAuthorityEnum | 0,1 | A code that identifies the business register, or other registration authority that supplied the value of RegistrationAuthorityEntityID. |
| RegistrationAuthorityEntityID | String | 0,1 | The identifier of the legal entity as maintained by a business registry in the jurisdiction of legal registration,<br>OR<br>If the legal entity is one that is not recorded in a business registry (e.g., one of the varieties of funds registered instead with financial regulators), the identifier of the legal entity in the appropriate registration authority.<br>The RegistrationAuthorityEntityID element SHALL be included if RegistrationAuthority is included, and SHALL be omitted if RegistrationAuthority is omitted. |
| LegalJurisdiction | RegionCode | 0,1 | The jurisdiction of legal formation and registration of the legal entity (and on which the legalForm data element is also dependent). |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityCategory | EntityCategoryTypeEnum | 0,1 | Indicates the general category of the type of legal entity identified by this ELEI data record |
| LegalForm | LegalFormType | 0,1 | The legal form of the legal entity, from an external code list that incorporates the ISO Entity Legal Form (ELF) code list mai |
| LegalEntityStatus | LegalEntityStatusEnum | 1 | The status of the legal entity. This is not to be confused with the status of the registration; see RegistrationStatus.<br>If this ELEI record contains a non-empty SuccessorELEI field, LegalEntityStatus is the last status of the legal entity before the successor ELEI record superseded this one, which is not necessarily the current status of the legal entity. |
| LegalEntityEvents | LegalEntityEvent | 0,n | Corporate events that occurred during this historical period |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.3.2 Registration Section of ELEI Data Record

The `Registration` section of an ELEI Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `RegistrationCreationDate` | `DateTime` | 1 | Date/time this ELEI record was initially created in the system |
| `RegistrationUpdateDate` | `DateTime` | 1 | Date/time that this historical ELEI record was most recently updated in the system. |
| `RegistrationSourceReference` | `String` | 0,1 | A locator of the source of the legal entity reference data in this registration record (mechanism TBD) |
| `RegistrationStatus` | `RegistrationStatusEnum` | 1 | Status of the ELEI registration. This is not to be confused with the status of the legal entity itself; see `LegalEntityStatus`. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| RegistrationUpToDate | DateTime | 0,1 | Should it exist and be known, the historical date up until which time the ELEI identifying reference data for this legal entity is valid. (Not the same thing as when the data was changed in the system). If present, a subsequent historical record for this ELEI may exist with the revised information. |
| RegistrationAsOfDate | DateTime | 0,1 | The date that this version of the reference data is known to be valid. This allows a record of historical changes to the legal entity identifying data to be recorded. This date would typically be derived from either:<br>1. The publication date of the source reference, or<br>2. Information contained in the source reference |
| ResponsibleDSU | DSUID | 1 | The Identifier of the Data Submission Unit (DSU) that produced and manages this ELEI registration. |
| ValidationSources | ValidationSourcesEnum | 0,1 | The current validation status of this ELEI record, or omitted if the validation status is not known or not revealed. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.3.3 Extension Section of ELEI Data record

The `Extension` section of an ELEI record may be used to include additional data not defined in this standard. For example, an DSU may use `Extension` to publish additional data elements it collects as part of registration.

- TBD: include the details of how this works. Basically, the idea is to use an XSD schema wildcard with namespace ##other, permitting the inclusion of XML elements from other XML namespaces.

## 3.4 Data Types

This section specifies the data types referenced by the tables in Section 3.2, in alphabetical order.

### 3.4.1 Address Data Type

A value of type `Address` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Cardinality | Description |
|---|---|---|---|
| `lang` | `LanguageCode` | 0,1 | The language in which all of the string-valued components of this address are expressed. |
| `Line1` | `String` | 0,1 | The first line of the street address |
| `Line2` | `String` | 0,1 | The second line of the street address |
| `Line3` | `String` | 0,1 | The third line of the street address. This element SHALL be omitted if `addressLine2` is omitted. |
| `Line4` | `String` | 0,1 | The fourth line of the street address. This element SHALL be omitted if `addressLine3` is omitted. |
| `City` | `String` | 1 | The name of the city |
| `Region` | `ERegionCode` | 0,1 | The "EURHISFIRM" region code for a region (state, province, county, parish, etc) that extends the ISO 3166-2 region code with historical additions |
| `Country` | `ECountryCode` | 1 | The "EURHISFIRM" country code that extends the 2-character ISO 3166-1 country code with historical additions |
| `PostalCode` | `String` | 0,1 | The postal code of this address as specified by the local postal service. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- TBD: should postal code be optional? [YES]

### 3.4.2 CountryCode Data Type

A value of type `CountryCode` in a file conforming to this standard SHALL be a 2-character country code conforming to ISO ISO 3166-1 alpha-2 [ISO3166]. Note that ISO 3166-1 alpha-2 codes are all uppercase.

### 3.4.3 DateTime Data Type

A value of type `DateTime` in a conforming to this standard SHALL be a point in time expressed as a string conforming to ISO 8601 having the following format:

*YYYY*-*MM*-*DD*T*hh*:*mm*:*ss.sss*TZ

where the components of the above string are as follows:

- *YYYY* is the year
- *MM* is the month (01 = January, …, 12 = December)
- *DD* is the day of the month (01 = first day of the month)
- T is the single character 'T'
- *hh* is the hour (00 – 23)
- *mm* is the minute
- *ss.sss* is the second and milliseconds. From one to three digits may be used for milliseconds, or omitted entirely along with the decimal point.
- *TZ* is the time zone specifier, which can be either:
  - Z the single character 'Z', denoting Coordinated Universal Time (UTC); or
  - +*hh*:*mm* denoting a positive offset from UTC; or
  - -*hh*:*mm* denoting a negative offset from UTC

In the XML representation specified in Section **Error! Reference source not found.**, the XSD type `xs:dateTime` is used; however, whereas `xs:dateTime` permits the time zone specifier to be omitted, `DateTime` values in files conforming to this standard SHALL always include a time zone specifier.

*Explanation (non-normative): milliseconds are hardly necessary for ELEI reference data, and likewise it might seem simpler to allow only "Z" as a time zone specifier; however, XML processing tools support the full syntax given above and it is not always possible to restrict such tools to avoid milliseconds or force the use of "Z" as the time zone specifier. The restriction that the time zone specifier must be present is equivalent to using XSD type `xs:timestamp`; however this was introduced in XSD 1.1 and not supported by the majority of XML processing tools which still only implement XSD 1.0.*



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.4 DSUID Data Type

A value of type `DSUID` in a file conforming to this standard SHALL be a [TBD]-character Data Submission Unit Identifier conforming to [TBD].

### 3.4.5 ELEI Data Type

A value of type `ELEI` in a file conforming to this standard SHALL be a 20-character Legal Entity Identifier conforming to [ISO17422]. (See Appendix 11: ELEI Code Partitioning Scheme for details on the partitioning of the ELEI code.)

### 3.4.6 LegalEntityCategory Data Type

A value of type `LegalEntityCategory` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `LegalEntityCategory` | `LegalEntityCategoryTypeEnum` | 0,1 | Indicates the general category of the type of legal entity identified by this ELEI data record |

### 3.4.7 LanguageCode Data Type

A value of type `LanguageCode` in a file conforming to this standard SHALL be a 2-character language code conforming to [ISO639-1]. Note that ISO 639-1 language codes are all lowercase.

- TBD: Alternatively, we could use IETF language tags (RFC 4646) instead. IETF language tags can distinguish between variations of the same language in different countries. For example, the IETF language tags `fr-CA` and `fr-FR` denote Canadian French and French as spoken in France, respectively; in ISO 639-1 these would both be simply `fr`.

### 3.4.8 LegalForm Data Type

A value of type `LegalForm` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `LegalForm` | `LegalFormCode` | 1,1 | The legal form of the legal entity |
| `OtherLegalForm` | `String` | 0,1 | Interim free-text legal form information in the process of transition to an ELF standard code |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.9 LegalFormCode Data Type

`LegalFormCode` is a code that is a normative enumeration that subsumes the ISO Entity Legal Form (ELF) code list maintained by GLEIF, but which incorporates additional codes for historical legal forms that are no longer in use.

A value of type `LegalForm` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `LegalFormCode` | `LegalFormEnum` | 1,1 | The legal form of the legal entity |

### 3.4.10 Name Data Type

A `Name` is a string expressed in a natural language, including a code indicating which natural language is used.

A value of type `Name` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `lang` | `LanguageCode` | 0,1 | The language of `name` |
| `Name` | `String` | 1 | The name itself. |

### 3.4.11 OtherAddress Data Type

A value of type `OtherAddress` in a file conforming to this standard SHALL include data elements as specified below. Each `Address` element includes an optional language code, permitting `OtherAddress` to be repeated as many times as necessary to express the same address type in multiple languages. The purpose of the `AddressTypeEnum` code list is to accommodate legal address and headquarters address in different languages, not to add other address types (which could conceivably be added in the future).

| Element Name | Type | Card | Description |
|---|---|---|---|
| `type` | `AddressTypeEnum` | 1 | The type of address represented by this `OtherAddress` instance. |
| `Address` | `Address` | 1 | The address |

### 3.4.12 OtherLegalEntityName Data Type

A value of type `OtherLegalEntityName` in a file conforming to this standard SHALL include data elements as specified below. Each `Name` element includes an optional language code, permitting `OtherLegalEntityName` to be repeated as many times as necessary to express the same name type in multiple languages. When `type` is



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, the language code specifies the language of the name prior to Romanization.

| Element Name | Type | Card | Description |
|---|---|---|---|
| type | LegalEntityNameTypeEnum | 1 | The type of name represented by this OtherLegalEntityName instance. |
| Name | Name | 1 | The name. If type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, then this value SHALL only include characters from the character set specified in Section 8. |

### 3.4.13 RegionCode Data Type

A value of type RegionCode in a file conforming to this standard SHALL be a code conforming to ISO 3166-2. Note that ISO 3166-2 codes are all uppercase.

### 3.4.14 RegistrationAuthority Data Type

A value of type RegistrationAuthority in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| RegistrationAuthorityID | RegistrationAuthorityEnum | 0,1 | An identifier for the legal entity **registry** of the legal entity in the jurisdiction of legal registration, or in the appropriate registration authority. |
| RegistrationAuthorityEntityID | String | 0,1 | An identifier for the **legal entity** at the registry in the jurisdiction of legal registration, or in the appropriate registration authority. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 3.5 Enumerated Code Lists

This section specifies the enumerated code list data types (all having the suffix `Enum`) referenced by the tables in Sections 3.2 and 0, in alphabetical order.

### 3.5.1 AddressTypeEnum Code List

The `AddressTypeEnum` value in an `OtherAddress` instance specifies how the alternative address relates to the legal entity.

A value of type `AddressTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| LEGAL_ADDRESS | Registered address of the legal entity in the legal jurisdiction |
| HEADQUARTERS_ADDRESS | Address of the headquarters of the legal entity |

### 3.5.2 LegalEntityCategoryTypeEnum

The `LegalEntityCategoryTypeEnum` value in an ELEI record specifies the general classification category of the legal entity.

- TBD: the table below is just an example with some possible categories, and is incomplete, and needs to be elaborated.

A value of type `LegalEntityCategoryTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| PUBLIC_CORPORATION | Public shareholder entity |
| PRIVATE_CORPORATION | Privately owned limited liability corporation |
| LLC | Privately owned limited liability company |
| SOLE_PROPRIETOR | Private business owned and operated by a single individual |
| LLP | Privately owned limited liability partnership |
| TRUST | Legal trust |
| FUND | Alternative Investment Scheme |
| GSE | Government sponsored entity |
| NGO | Non-Government Organization |
| NGO | Non-Government Organization |
| | Partnership limited by shares |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| | Limited partnership |
| | Unlimited partnership |
| | Public limited company (e.g. S.A., N.V., PLC, …) |
| | Private limited company (e.g. GmbH) |

There actually is a convenient list on Wikipedia:
https://en.wikipedia.org/wiki/List_of_legal_entity_types_by_country

### 3.5.3 LegalEntityExpirationReasonEnum Code List

The `LegalEntityExpirationReasonEnum` value in an ELEI record specifies the reason that the legal entity expired.

A value of type `LegalEntityExpirationReasonEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `DISSOLVED` | The legal entity ceased to ~~operate~~ exist, and was legally shuttered. |
| `CORPORATE_ACTION` | The legal entity was acquired or merged with another legal entity |
| `OTHER` | The reason for expiry is not one of the above |

### 3.5.4 LegalEntityNameTypeEnum Code List

The `LegalEntityNameTypeEnum` value in an `LegalEntityName` specifies how the name relates to the legal entity.

A value of type `LegalEntityNameTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `OTHER_LEGAL` | Registered name of the legal entity in an alternate language in the legal jurisdiction in which the legal entity is registered |
| `PREFERRED_ROMANIZED_LEGAL` | Romanized form, preferred by the legal entity |
| `AUTO_ROMANIZED_LEGAL` | Romanized form, auto-transliterated |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.5.5 LegalEntityStatusEnum Code List

The `LegalEntityStatusEnum` value in an ELEI record indicates the status of the legal entity itself. This is not to be confused with the status of the ELEI registration, which is specified by `RegistrationStatusEnum` (Section 3.4.8). See also Section 9, which illustrates how the `LegalEntityStatusEnum` value changes over the lifecycle of an ELEI registration.

A value of type `LegalEntityStatusEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `ACTIVE` | As of the last report or update, the legal entity reported is legally registered and operating. |
| `INACTIVE` | It has been determined that the legal entity that was assigned the ELEI is no longer legally registered and/or operating, whether as a result of:<br>1. Business closure<br>2. Acquisition by or merger with another (or new) legal entity<br>3. Determination of illegitimacy |
| `NOT_SPECIFIED` | The ELEI record is in a state in which it does not provide information about whether the legal entity is legally registered and operating. |

### 3.5.6 LegalFormEnum Code List

The value of `LegalFormEnum` SHALL be a code published in the ISO Entity Legal Form (ELF) code list maintained by GLEIF.

### 3.5.7 RegistrationAuthorityEnum Code List

The `RegistrationAuthorityEnum` value in an `LegalEntity` instance specifies what business register provided the value of `RegistrationAuthorityEntityID` for the legal entity.

The value of `RegistrationAuthorityEnum` SHALL be a code published in a list of registration authority codes.

- TBD: the external code list for Registration Authorities needs to be developed, but will borrow from, incorporate or reference the Registration Authorities Code List published by the Global LEI Foundation.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.5.8 RegistrationStatusEnum Code List

The `RegistrationStatusEnum` value in an ELEI record indicates the status of the registration of the legal entity with an DSU. This is not to be confused with the status of the ELEI itself, which is specified by `LegalEntityStatusEnum` (Section 3.5.5). See also Section 9, which illustrates how the `LegalEntityStatusEnum` value changes over the lifecycle of an ELEI registration.

A value of type `RegistrationStatusEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `PENDING` | An application for an ELEI that has been submitted and which is being processed and validated.<br>*NOTE: ELEI registrations in the `PENDING` state are not intended for public release, but could be used internally between DSUs.* |
| `ISSUED` | An ELEI Registration that has been validated and issued, and which identifies ggal entity that was an operating legal entity as of the last update. |
| `DUPLICATE` | An ELEI Registration that has been determined to be a duplicate registration of the same legal entity as another ELEI Registration; the `DUPLICATE` status is assigned to the non-surviving registration (i.e., the ELEI that should no longer be used). Only one of the potential multiple identifiers will survive; for all other duplicate registrations:<br>1. The `RegistrationStatus` is set to `DUPLICATE`,<br>2. The ELEI of the surviving ELEI Registration is set in the `SuccessorELEI` data element of (each) duplicate ELEI registration;<br>3. The `LastUpdateDate` is set to reflect the date of this update, and<br>4. No further updates of the `DUPLICATE` registration record will occur. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| MERGED | An ELEI registration for a legal entity that has been merged into another legal entity, such that this legal entity no longer exists as an operating entity.<br>If<br>• After being issued an ELEI, the legal entity is acquired by, or merged with, another legal entity;<br>• Per agreements among the parties to the transaction, the ELEI of the acquired or merged legal entity will not be used to identify the surviving legal entity (or if a new legal entity is created that is issued a new ELEI)<br>Then<br>1. The ELEIRegistrationStatus is set to "MERGED",<br>2. The ELEI of the surviving/new legal entity is set in the successorELEI data element of (each) ELEI registration that is no longer to be used;<br>3. The ELEIRecordLastUpdate is set to reflect the date of this update, and<br>4. No further updates of the MERGED registration record(s) will occur. |
| RETIRED | An ELEI registration for a legal entity that has ceased operation, without having been merged into another legal entity.<br>If<br>• The responsible DSU determines by public sources that the legal entity has been dissolved or ceased to operate (and the DSU seeks to confirm this status through all available channels)<br>Then<br>1. The ELEIRegistrationStatus is set to "RETIRED";<br>2. The ELEIRecordLastUpdate is set to reflect the date of this update;<br>3. The LegalEntityExpirationDate is also set to the date of this update;<br>4. The LegalEntityExpirationReason is set;<br>5. No further updates of the RETIRED registration record will occur. |

### 3.5.9 ValidationSourcesEnum Code List

A value of type ValidationSourcesEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| PENDING | The validation of the reference data for the legal entity has not yet occurred. |
| SUBMITTED | Based on the validation procedures in use by the DSU responsible for the record, the information associated with this record has significant reliance on the information that a submitter provided due to the unavailability of corroborating information. |
| PARTIALLY_CORROBORATED | Based on the validation procedures in use by the DSU responsible for the record, the information supplied for the legal entity can be partially corroborated by available sources, while some of the record is dependent upon information that the submitter collected, either due to conflicts with authoritative information, or due to data unavailability. |
| FULLY_CORROBORATED | Based on the validation procedures in use by the DSU responsible for the record, there is sufficient information contained in authoritative public sources to corroborate the information regarding the legal entity provided in the record. |

.

## 4 Constraints and Data Validation

All values of type `String` specified in Section 3 SHALL be 500 or fewer characters in length.

- TBD. Specify here any additional constraints not implicit in the data types or explicit in the definition of each reference data element

## 5 Change Management

- TBD. Use the <vnext> strategy for forward/backward compatibility, and/or include a version number in the XML. Needs to be explained

## 6 Examples (non-normative)

- TBD.

## 7 References

[ISO646] ISO, "Information technology -- ISO 7-bit coded character set for information interchange," ISO/IEC 646:1991.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

[ISO3166-1] ISO, "Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes," ISO 3166-1:2013.

[ISO3166-2] ISO, "Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code," ISO 3166-2:2013.

[ISO8601] ISO, "Data elements and interchange formats – Information interchange – Representation of dates and times," ISO 8601:2004.

[ISO10646] ISO, "Information technology -- Universal Coded Character Set (UCS)," ISO 10646:2012.

[ISO17422] ISO, "Financial Services - Legal Entity Identifier (LEI)," ISO/DIS 17442:2012.

[ISODir2] ISO, "Rules for the structure and drafting of International Standards (ISO/IEC Directives, Part 2, 2001, 4th edition)," July 2002.

[XSD1] H. Thompson, D. Beech, M. Maloney, N. Mendelsohn, "XML Schema Part 1: Structures," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-1/.

[XSD2] P. Biron, A. Malhotra, "XML Schema Part 2: Datatypes," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-2/.

# 8 Appendix: Character Codes Allowed in Romanized Names

When a `Name` instance is of type `PREFERRED_ROMANIZED` or `AUTO_ROMANIZED`, the value of the `name` field SHALL consist only of non-control characters drawn from the "invariant subset" of ISO 646. These characters are enumerated below. The "Hex Value" column indicates the code point value (expressed in hexadecimal) for each character in both ISO 646 and ISO 10646.

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| ! | Exclamation Mark | 21 | M | Capital Letter M | 4D |
| " | Quotation Mark | 22 | N | Capital Letter N | 4E |
| % | Percent Sign | 25 | O | Capital Letter O | 4F |
| & | Ampersand | 26 | P | Capital Letter P | 50 |
| ' | Apostrophe | 27 | Q | Capital Letter Q | 51 |
| ( | Left Parenthesis | 28 | R | Capital Letter R | 52 |
| ) | Right Parenthesis | 29 | S | Capital Letter S | 53 |
| * | Asterisk | 2A | T | Capital Letter T | 54 |
| + | Plus sign | 2B | U | Capital Letter U | 55 |
| , | Comma | 2C | V | Capital Letter V | 56 |
| - | Hyphen/ Minus | 2D | W | Capital Letter W | 57 |



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 777489

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| . | Full Stop | 2E | X | Capital Letter X | 58 |
| / | Solidus | 2F | Y | Capital Letter Y | 59 |
| 0 | Digit Zero | 30 | Z | Capital Letter Z | 5A |
| 1 | Digit One | 31 | _ | Low Line | 5F |
| 2 | Digit Two | 32 | a | Small Letter a | 61 |
| 3 | Digit Three | 33 | b | Small Letter b | 62 |
| 4 | Digit Four | 34 | c | Small Letter c | 63 |
| 5 | Digit Five | 35 | d | Small Letter d | 64 |
| 6 | Digit Six | 36 | e | Small Letter e | 65 |
| 7 | Digit Seven | 37 | f | Small Letter f | 66 |
| 8 | Digit Eight | 38 | g | Small Letter g | 67 |
| 9 | Digit Nine | 39 | h | Small Letter h | 68 |
| : | Colon | 3A | i | Small Letter i | 69 |
| ; | Semicolon | 3B | j | Small Letter j | 6A |
| < | Less-than Sign | 3C | k | Small Letter k | 6B |
| = | Equals Sign | 3D | l | Small Letter l | 6C |
| > | Greater-than Sign | 3E | m | Small Letter m | 6D |
| ? | Question Mark | 3F | n | Small Letter n | 6E |
| A | Capital Letter A | 41 | o | Small Letter o | 6F |
| B | Capital Letter B | 42 | p | Small Letter p | 70 |
| C | Capital Letter C | 43 | q | Small Letter q | 71 |
| D | Capital Letter D | 44 | r | Small Letter r | 72 |
| E | Capital Letter E | 45 | s | Small Letter s | 73 |
| F | Capital Letter F | 46 | t | Small Letter t | 74 |
| G | Capital Letter G | 47 | u | Small Letter u | 75 |
| H | Capital Letter H | 48 | v | Small Letter v | 76 |
| I | Capital Letter I | 49 | w | Small Letter w | 77 |
| J | Capital Letter J | 4A | x | Small Letter x | 78 |
| K | Capital Letter K | 4B | y | Small Letter y | 79 |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

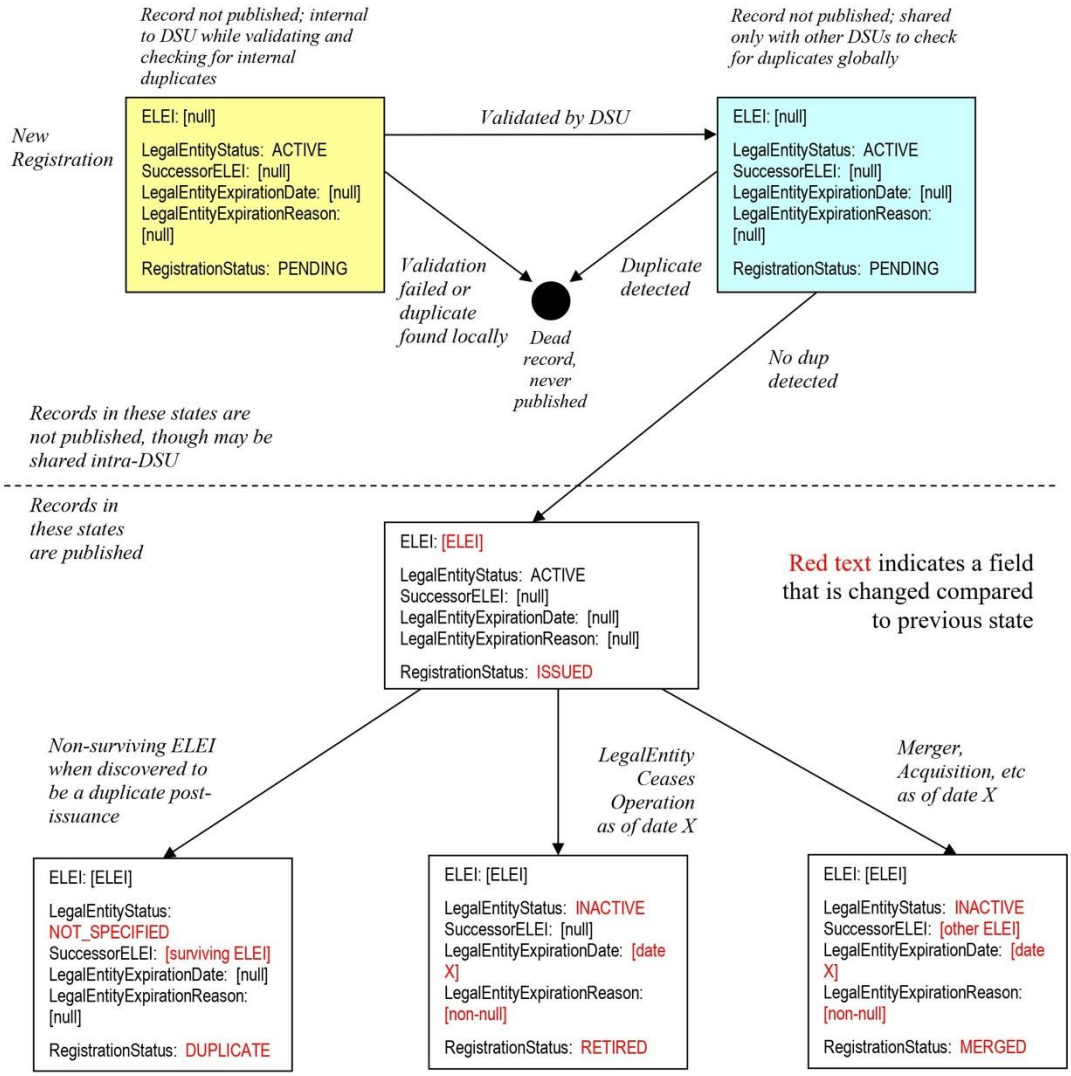
| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| L | Capital Letter L | 4C | z | Small Letter z | 7A |
| | Space | 20 | | | |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 9 Appendix: ELEI Record Transition Diagram

The following diagram illustrates the relationship between status fields of the ELEI record.




In rare circumstances, a record may pass directly from `PENDING` to `RETIRED` or `MERGED`; e.g., retroactive reporting of a failed or merged legal g. In that case, the `LegalEntityStatus` would be `INACTIVE` while the record is in the `PENDING` state.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 777489

## 10 Appendix: ELEI Code Partitioning Scheme

The following diagram illustrates the partitioning scheme of the ELEI code. The ELEI code is an ISO 17442-compliant, opaque identifier that resolves to the EURHISFIRM legal entity reference data that uniquely identifies a legal entity in the EURHISFIRM network.

The ELEI is minted by a Data Submission Unit (DSU) as part of the process that would add the ELEI -- and the unique-identifying reference data -- for a legal entity to the EURHISFIRM network under the EURHISFIRM Common Data Model (CDM) data standard.

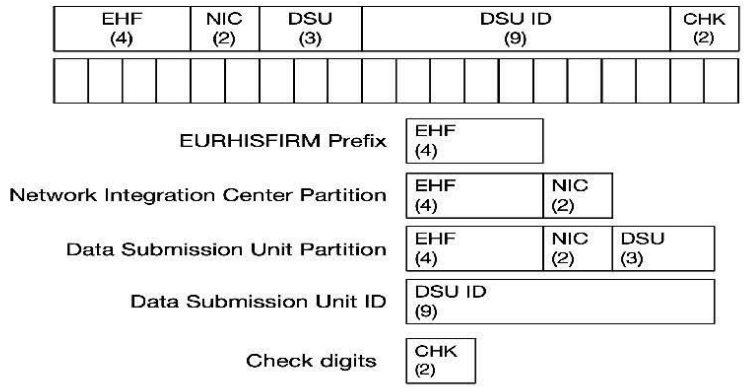
The DSU should, if possible, perform a check with the consolidated database of existing EURHISFIRM ELEI identifier reference data to verify that a legal entity has not already been identified and been assigned an ELEI before submitting a newly minted ELEI to a Network Integration Center (NIC). Otherwise, the Network Integration Center would need to perform this check for the existence of an ELEI that had already been assigned to the legal entity in question in order to maintain the uniqueness of the ELEI assigned to the legal entity.

In cases where the NIC may not be able to do this, or in which a duplicative entry existed but was not found, remediation and cleanup of multiple ELEIs for the same legal entity could be detected and rectified by background data quality processes that periodically sweep the ELEI reference data or by challenges submitted from the EURHISFIRM community.

Note that the partitioning scheme does not add intelligence or actually identify any DSU or NIC, as multiple prefixes could be issued to a single DSU under one of multiple prefixes that may be issued to a NIC. The partitioning scheme is only for the purpose of guaranteeing the ability to concurrently create unique ELEI codes in a federated and decentralized legal entity data "supply chain" in the EURHISFIRM network. (The EHF is the prefix that would be assigned to EURHISFIRM by the Global LEI Foundation in order to be globally unique in the Global LEI System )

EURHISFIRM Legal Entity Identifier ( ELEI )
*(ISO 17742 Compliant)*

Code Partitioning Scheme






This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 9.3 Legal Entity Data Standard 1.09

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# EURHISFIRM Common Data Model

## Legal Entity Data Standard 1.0

Version of 28 November 2020

| Date | Revision | Description |
|---|---|---|
| 2020-01-15 | ver 1.01 | First Draft to be circulated (formatting fixed) |
| 2020-02-06 | ver 1.02 | Added partitioning scheme for ELEI in Appendix 11;<br>made `Line1` of the street address in the **`Address`** type optional;<br>qualified the term "entity" with "legal entity" throughout;<br>revised the legal entity record transition diagram appendix |
| 2020-02-19 | ver 1.03 | Added `EntityCreationDate` start date for legal entity to complement the `EntityExpirationDate`<br>Added `RegistrationUpToDate` and `RegistrationAsOfDate` to bracket the start and stop dates that this reference data is valid<br>Added `RegistrationSourceReference` in order to provide a reference to the definitive data source from which this data was collected<br>Added several revisions to element definitions |
| 2020-02-20 | ver 1.04 | Added ELEI Root record (needed to anchor the unique ELEI in order to support multiple (historical) ELEI Data records |
| 2020-03-04 | ver 1.05 | Revised the ELEI root record to include `EarliestRegistrationDate, LatestRegistrationDate. EntityStatus`<br>Added `EntityEvents` to ELEI data record<br>Revised `CountryCode` to be `ECountryCode` and `RegionCode` to be `ERegionCode` in order to include historical countries and regions |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| | | |
|---|---|---|
| 2020-03-18 | ver 1.06 | Consolidated comments and accepted revisions<br>Revised `LegalForm` to be `ELegalForm`, `RegistrationAuthority` to be `ERegistrationAuthority` and `RegistrationAuthorityEnum` to be `ERegistrationAuthorityEnum` in order to be able to accommodate historical forms |
| 2020-04-02 | ver 1.07 | Converted the ERegistrationAuthority elements into a data type<br>Added OtherRegistrationAuthority element to allow for historical government legal entity authorities that are not yet included in the RegistrationAuthority external code list to be recorded;<br>Elaborated on the `RegistrationSource` element in the Registration Record to make it a source metadata reference structure containing `RegistrationSourceLocatorKey`, `RegistrationSourceLocatorAnchor`, `DCUID`, and `RegistrationSourceEntityID` elements.<br>Added a `DCUID` data element in order to identify Data Collection Units as well as Data Submission Units;<br>Added preliminary elements for the Header structure; |
| 2020-04-14 | ver 1.08 | Renamed `RegistrationSourceLocator` to `RegistrationSource`.<br>Added `RegistrationSourceCitation` to `RegistrationSource` Data Type |
| 2020-11-26 | ver 1.09 | Changed references to "legal entity" to "entity" in order to allow the inclusion of public sector organizations ( sovereign governments, nationalized cororations, etc ) |

## Abstract

First, the semantic content of these attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the EURHISFIRM Common Data Model.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Status of this document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this document series will be maintained on SeaFile.

This draft is a **Working Draft** which can be circulated to any interested parties for review and comment. It is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to use Working Drafts as reference material or to cite them as other than "work in progress." This is work in progress and does not imply endorsement by the EURHISFIRM ExCo.

Comments on this document should be sent to [TBD: insert mailing list or URL].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Table of Contents

1 Introduction 6
2 Terminology and Typographical Conventions 7
3 Abstract Data Content 7
3.1 ELEI File Header 8
3.2 ELEI Root Record 8
3.3 ELEI Data Record 10
3.3.1 Entity Section of ELEI Data Record 10
3.3.2 Registration Section of ELEI Data Record 14
3.3.3 Extension Section of ELEI Data record 15
3.4 Data Types 15
3.4.1 Address Data Type 16
3.4.2 DateTime Data Type 16
3.4.3 DCUID Data Type 17
3.4.4 DSUID Data Type 17
3.4.5 ECountryCode Data Type 18
3.4.6 ELegalForm Data Type 18
3.4.7 ELegalFormCode Data Type 18
3.4.8 ERegionCode Data Type 18
3.4.9 ERegistrationAuthorityType Data Type 19
3.4.10 ELEI Data Type 20
3.4.11 EntityCategory Data Type 20
3.4.12 LanguageCode Data Type 20
3.4.13 Name Data Type 20
3.4.14 OtherAddress Data Type 20
3.4.15 OtherEntityName Data Type 21
3.4.16 RegistrationSource Data Type 22
3.4.17 RegistrationSourceCitation Data Type 23
3.4.18 RegistrationSourceLocatorAnchorType Data Type 24
3.4.19 RegistrationSourceLocatorKeyType Data Type 24
3.5 Enumerated Code Lists 24



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

3.5.1 AddressTypeEnum Code List....24
3.5.2 ELegalFormEnum Code List....24
3.5.3 ERegistrationAuthorityEnum Code List....25
3.5.4 EntityCategoryTypeEnum....25
3.5.5 EntityExpirationReasonEnum Code List....26
3.5.6 EntityNameTypeEnum Code List....26
3.5.7 EntityStatusEnum Code List....26
3.5.8 RegistrationStatusEnum Code List....27
3.5.9 ValidationSourcesEnum Code List....28
4 Constraints and Data Validation....29
5 XML Syntax....29
5.1 XML Design Rules....29
6 Change Management....30
7 Examples (non-normative)....30
8 References....30
9 Appendix: Character Codes Allowed in Romanized Names....31
10 Appendix: ELEI Record Transition Diagram....33
11 Appendix: ELEI Code Partitioning Scheme....34



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 1 Introduction

Legal entities are clearly one of the core classes of objects in a historical financial Research Infrastructure such as EURHISFIRM, on which the majority of other EURHISFIRM financial data depend. As such, legal entities should be uniquely and unambiguously identified in EURHISFIRM, and this identification is accomplished based on the identifying attributes associated with each entity.

First, the semantic content of those attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the reference data attributes of legal entities in the EURHISFIRM Common Data Model

This document proposes initial standards for EURHISFIRM entity reference data. It is important that this reference data should uniquely identify the legal entities that are harvested from contributing sources and then collected and assimilated into the common EURHISFIRM platform.

A EURHISFIRM Entity Identifier (ELEI) code that resolves to this entity-identifying reference data is also introduced.

- A partitioning scheme for the structure of this identifier is defined that allows the independent and concurrent minting and assignment of ELEI codes to entity reference data that is produced by multiple Data Submission Units (DSUs) in the federated EURHISFIRM Research Infrastructure network.

The standard set by this document is expected, among other things, to reduce the risk of duplicates stemming from differences in formats and conventions of locally sourced data provided by regional contributors, to ensure data quality in the EURHISFIRM system, and to enable the subsequent detection and resolution of multiple identifiers for the same entity to the extent that they occur. The standard is expected to be used as a format for reference data consolidated from all sources in order to be promoted to the level of common EURHISFIRM data published for end-user access.

The contents of this document are as follows:

- Section 2 defines terminology and typographical conventions.
- Section 3 specifies the abstract content of EURHISFIRM entity reference data conforming to this standard, including a detailed description of each element of reference data associated with legal entities in EURHISFIRM. Allowable values for data elements that are code lists will also be (subsequently) specified. The partitioning scheme for the structure of the EURHISFIRM Legal Enty Identifier (ELEI) code will also be subsequently described.
- Section 4 specifies data validation constraints and other considerations intended to lead to high-quality data content.
- Section 5 specifies a concrete realization of the data definitions above in XML syntax, by means of XML schema (XSD 1.0).



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- Section 6 specifies how this file format may be changed in the future, providing for versioning, forward- and backward-compatibility, etc.
- Section 7 provides examples to illustrate the file format.

## 2 Terminology and Typographical Conventions

Within this specification, the terms SHALL, SHALL NOT, SHOULD, SHOULD NOT, MAY, NEED NOT, CAN, and CANNOT are to be interpreted as specified in Annex G of the ISO/IEC Directives, Part 2, 2001, 4th edition [ISODir2]. When used in this way, these terms will always be shown in ALL CAPS; when these words appear in ordinary typeface they are intended to have their ordinary English meaning.

All sections of this document, with the exception of Section 1 are normative, except where explicitly noted as non-normative.

The following typographical conventions are used throughout the document:

- ALL CAPS type is used for the special terms from [ISODir2] enumerated above.
- `Monospace` type is used to denote programming language, UML, and XML identifiers, as well as for the text of XML documents.

➢ Placeholders for open issues and/or changes that need to be made to this document prior to its reaching the final stage of approved Proposed Standard are prefixed by a rightward-facing arrowhead, as this paragraph is.

## 3 Abstract Data Content

This section specifies the abstract data content of a a data file conforming to this standard.

A data file conforming to this standard SHALL consist of:

- An optional ELEI File Header, as specified in Section 3.1.
- An ELEI Root Record as specified in Section 3.2.
- Zero or more ELEI Data Records, as specified in Section 3.3.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.1 ELEI File Header

- TBD Define elements to go into a "header" area for an ELEI file. The purpose of the header is to provide context about the file and its contained ELEI data records. However, the header will not contain anything necessary to interpret the meaning of any ELEI record; e.g., things like default values for ELEI data records will not be in the header (such things would mean that the meaning of an ELEI record could change if taken away from the header). Examples of things that might be useful to include in the header:

| Element Name | Type | Card | Description |
|---|---|---|---|
| ContentDate | DateTime | 1,1 | The date and time of generation of the data |
| Originator | | 0,1 | The identifier of the creator of the content of this file |
| FileContent | | 1,1 | A code describing the content of this data file. |
| ProcessStage | | 0,1 | A code indicating the stage of this file in the EURHISFIRM workflow |
| RecordCount | | 1,1 | The number of data records in the file. Can be a positive whole (integer) number, or zero (0). |

### 3.2 ELEI Root Record

The ELEI Root Record anchors a single ELEI. This root record is needed in order to provide a unique ELEI anchor to support multiple (historical) ELEI Data Records. Each ELEI Root record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ELEI | ELEI | 1 | The 20-character ELEI of the entity described by this ELEI Data Record. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| SuccessorELEI | ELEI | 0,1 | The ELEI of the ELEI registration that supersedes or subsumes this ELEI registration for the same entity.<br>If RegistrationStatus is DUPLICATE, then SuccessorELEI is the ELEI of the surviving ELEI Registration.<br>If RegistrationStatus is MERGED, then SuccessorELEI is the ELEI Registration of the new/acquiring entity.<br>Otherwise, SuccessorELEI is omitted.<br>When a successor ELEI is assigned to an ELEI registration, the ELEI registration will no longer be updated (since another ELEI registration has superseded the registration that has just been assigned a successor).<br>The ELEIRecordLastUpdate that is recorded with the update assigning a SuccessorELEI will be the last update performed on the superseded registration record.<br>As a consequence, other fields of the superseded registration record (address, entity status, etc) may no longer reflect the actual state of the entity. |
| ELEICreationDate | DateTime | 1 | Date/time the ELEI root record was initially created in the system |
| EntityStatus | EntityStatusEnum | 1 | The status of the entity. This is not to be confused with the status of the registration; see RegistrationStatus.<br>If this ELEI record contains a non-empty SuccessorELEI field, EntityStatus is the last status of the entity before the successor ELEI record superseded this one, which is not necessarily the current status of the entity. |
| EntityFormationDate | DateTime | 0,1 | Date/time the Entity was formed, if known |
| EntityExpirationDate | DateTime | 0,1 | The date that the entity ceased to operate, whether due to dissolution, merger or acquisition. Omitted if the entity has not ceased to operate, or if this ELEI record contains a non-empty SuccessorELEI field. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityExpirationReason | EntityExpirationReasonEnum | 0,1 | he reason that a entity ceased to operate. This element SHALL be present if EntityExpirationDate is present, and omitted otherwise. |
| EarliestRegistrationDate | DateTime | 1 | Date/time of the Registration record with the earliest historical date |
| LatestRegistrationDate | DateTime | 1 | Date/time of the Registration record with the most recent historical date |

### 3.3 ELEI Data Record

An ELEI Data Record describes a single ELEI. Each ELEI Data record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ELEI | ELEI | 1 | The 20-character ELEI of the entity described by this ELEI Data Record. |
| Entity | Entity (Section 3.3.1) | 1 | Attributes describing the entity itself |
| Registration | Registration (Section 0) | 1 | Attributes describing the registration of this ELEI. |
| Extension | Extension (Section 0) | 0,1 | An optional element for including data beyond the standard data elements in an ELEI data file. This may include data specific to an DSU, data specific to a publisher of ELEI data, and so on. |

#### 3.3.1 Entity Section of ELEI Data Record

The Entity section of an ELEI Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityName | Name | 1 | The preferred name of the Entity.<br>If an Entity is in a jurisdiction with more than one Legal Name (e.g., in different languages), this is the Primary Legal Name (see otherEntityNames for other names). |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| LegalName | Name | 0,1 | The Legal Name of the Entity.<br>If an Entity is in a jurisdiction with more than one Legal Name (e.g., in different languages), this is the Primary Legal Name (see otherEntityNames for other names). |
| OtherEntityNames | Other EntityName | 0..n | An optional list of other Name instances for the Entity. |
| EntityLocation | Address | 0,1 | The address of the Entity |
| LegalAddress | Address | 1 | The address of the Entity as recorded in the registration of the Entity in its legal jurisdiction |
| HeadquartersAddress | Address | 0,1 | The address of the headquarters of the Entity |
| OtherAddresses | OtherAddress | 0..n | An optional list of other Address instances for the Entity. This may be used to provide alternative language forms of legal address or headquarters address.. |
| ERegistrationAuthority | ERegistration AuthorityType | 0,1 | A code that identifies the business register, or other registration authority that supplied the value of ERegistrationAuthorityEntityID. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| ERegistrationAuthority EntityID | String | 0,1 | The identifier of the entity as maintained by a business registry in the jurisdiction of legal registration,<br>OR<br>If the entity is one that is not recorded in a business registry (e.g., one of the varieties of funds registered instead with financial regulators), the identifier of the entity in the appropriate registration authority.<br>The RegistrationAuthorityEntityID element SHALL be included if RegistrationAuthority is included, and SHALL be omitted if RegistrationAuthority is omitted. |
| LegalJurisdiction | ERegionCode | 0,1 | The jurisdiction of legal formation and registration of the entity (and on which the legalForm data element is also dependent). |
| EntityCategory | EntityCategoryTypeEnum | 0,1 | Indicates the general category of the type of entity identified by this ELEI data record |
| LegalForm | ELegalFormType | 0,1 | The legal form of the entity, from an external code list that incorporates the ISO Entity Legal Form (ELF) code list mai |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityStatus | EntityStatusEnum | 1 | The status of the entity. This is not to be confused with the status of the registration; see RegistrationStatus.<br>If this ELEI record contains a non-empty SuccessorELEI field, EntityStatus is the last status of the entity before the successor ELEI record superseded this one, which is not necessarily the current status of the entity. |
| EntityEvents | EntityEvent | 0,n | Corporate events that occurred during this historical period |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.3.2 Registration Section of ELEI Data Record

The `Registration` section of an ELEI Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `RegistrationCreationDate` | `DateTime` | 1 | Date/time this ELEI record was initially created in the system |
| `RegistrationUpdateDate` | `DateTime` | 1 | Date/time that this historical ELEI record was most recently updated in the system. |
| `RegistrationSource` | `RegistrationSourceType` | 0,n | A locator of the source of the entity reference data in this registration record (mechanism TBD, typically company yearbooks) |
| `RegistrationStatus` | `RegistrationStatusEnum` | 1 | Status of the ELEI registration. This is not to be confused with the status of the entity itself; see `EntityStatus`. |
| `RegistrationUpToDate` | `DateTime` | 0,1 | Should it exist and be known, the historical date up until which time the ELEI identifying reference data for this entity is valid. (Not the same thing as when the data was changed in the system). If present, a subsequent historical record for this ELEI may exist with the revised information. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| RegistrationAsOfDate | DateTime | 0,1 | The date that this version of the reference data is known to be valid. This allows a record of historical changes to the entity identifying data to be recorded. This date would typically be derived from either:<br>1. The publication date of the source reference, or<br>2. Information contained in the source reference |
| ResponsibleDSU | DSUID | 1 | The Identifier of the Data Submission Unit (DSU) that produced and manages this ELEI registration. |
| ValidationSources | ValidationSourcesEnum | 0,1 | The current validation status of this ELEI record, or omitted if the validation status is not known or not revealed. |

### 3.3.3 Extension Section of ELEI Data record

The Extension section of an ELEI record may be used to include additional data not defined in this standard. For example, an DSU may use Extension to publish additional data elements it collects as part of registration.

- TBD: include the details of how this works. Basically, the idea is to use an XSD schema wildcard with namespace ##other, permitting the inclusion of XML elements from other XML namespaces.

## 3.4 Data Types

This section specifies the data types referenced by the tables in Section 3.2, in alphabetical order.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.1 Address Data Type

A value of type `Address` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Cardinality | Description |
|---|---|---|---|
| `lang` | `LanguageCode` | 0,1 | The language in which all of the string-valued components of this address are expressed. |
| `Line1` | `String` | 0,1 | The first line of the street address |
| `Line2` | `String` | 0,1 | The second line of the street address |
| `Line3` | `String` | 0,1 | The third line of the street address. This element SHALL be omitted if `addressLine2` is omitted. |
| `Line4` | `String` | 0,1 | The fourth line of the street address. This element SHALL be omitted if `addressLine3` is omitted. |
| `City` | `String` | 1 | The name of the city |
| `Region` | `ERegionCode` | 0,1 | The "EURHISFIRM" region code for a region (state, province, county, parish, etc) that extends the ISO 3166-2 region code with historical additions |
| `Country` | `ECountryCode` | 1 | The "EURHISFIRM" country code that extends the 2-character ISO 3166-1 country code with historical additions [ e.g., https://en.wikipedia.org/wiki/Lists_of_sovereign_states_by_year ] |
| `PostalCode` | `String` | 0,1 | The postal code of this address as specified by the local postal service. |

- TBD: should postal code be optional? [YES]

### 3.4.2 DateTime Data Type

A value of type `DateTime` in a conforming to this standard SHALL be a point in time expressed as a string conforming to ISO 8601 having the following format:

- TBD: Provision for different calendars ? (i.e., Gregorian, Julian, etc. )



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

*YYYY*–*MM*–*DD*T*hh*:*mm*:*ss.sss*TZ

where the components of the above string are as follows:

- *YYYY* is the year
- *MM* is the month (01 = January, …, 12 = December)
- *DD* is the day of the month (01 = first day of the month)
- T is the single character 'T'
- *hh* is the hour (00 – 23)
- *mm* is the minute
- *ss.sss* is the second and milliseconds. From one to three digits may be used for milliseconds, or omitted entirely along with the decimal point.
- *TZ* is the time zone specifier, which can be either:
  - Z the single character 'Z', denoting Coordinated Universal Time (UTC); or
  - +*hh*:*mm* denoting a positive offset from UTC; or
  - –*hh*:*mm* denoting a negative offset from UTC

In the XML representation specified in Section 5, the XSD type xs:dateTime is used; however, whereas xs:dateTime permits the time zone specifier to be omitted, DateTime values in files conforming to this standard SHALL always include a time zone specifier.

*Explanation (non-normative): milliseconds are hardly necessary for ELEI reference data, and likewise it might seem simpler to allow only "Z" as a time zone specifier; however, XML processing tools support the full syntax given above and it is not always possible to restrict such tools to avoid milliseconds or force the use of "Z" as the time zone specifier. The restriction that the time zone specifier must be present is equivalent to using XSD type xs:timestamp; however this was introduced in XSD 1.1 and not supported by the majority of XML processing tools which still only implement XSD 1.0.*

### 3.4.3 DCUID Data Type

A value of type DCUID in a file conforming to this standard SHALL be a [TBD]-character Data Collection Unit Identifier conforming to [TBD].

### 3.4.4 DSUID Data Type

A value of type DSUID in a file conforming to this standard SHALL be a [TBD]-character Data Submission Unit Identifier conforming to [TBD].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.5 ECountryCode Data Type

A value of type `CountryCode` in a file conforming to this standard SHALL be a 2-character country code conforming to ISO ISO 3166-1 alpha-2 [ISO3166]. Note that ISO 3166-1 alpha-2 codes are all uppercase.

- TBD: CountryCode external code list must extend ISO country codes with start/stop dates and historical changes

### 3.4.6 ELegalForm Data Type

A value of type `LegalForm` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `LegalForm` | `ELegalFormCode` | 1,1 | The legal form of the entity |
| `OtherLegalForm` | `String` | 0,1 | Interim free-text legal form information in the process of transition to an ELF standard code |

### 3.4.7 ELegalFormCode Data Type

`LegalFormCode` is a code that is a normative enumeration that subsumes the ISO Entity Legal Form (ELF) code list maintained by GLEIF, but which incorporates additional codes for historical legal forms that are no longer in use.

A value of type `LegalForm` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `ELegalFormCode` | `ELegalFormEnum` | 1,1 | The legal form of the entity |

- TBD: Research on additions and extensions to ELF code list is needed, including start-stop date periods and historical jurisdictions (referencing extensions to the Country Code external code list)

### 3.4.8 ERegionCode Data Type

A value of type `RegionCode` in a file conforming to this standard SHALL be a code conforming to ISO 3166-2. Note that ISO 3166-2 codes are all uppercase.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- TBD: Like CountryCode code list, RegionCode external code list needs historical extensions and additions as well.

### 3.4.9 ERegistrationAuthorityType Data Type

A value of type `ERegistrationAuthority` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `ERegistrationAuthorityID` | `ERegistrationAuthorityEnum` | 0,1 | An identifier for the entity **registry** of the entity in the jurisdiction of legal registration, or in the appropriate registration authority. |
| `EOtherRegistrationAuthority` | `String` | 0,1 | A legacy / historical reference code of a registration authority which is not yet entered in the ERegistration Authorities List (RAL), or the designation of an interim register until such time as an entry from RAL can be delivered . |
| `ERegistrationAuthorityEntityID` | `String` | 0,1 | An identifier for the **entity** at the registry in the jurisdiction of legal registration, or in the appropriate registration authority. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.10 ELEI Data Type

A value of type `ELEI` in a file conforming to this standard SHALL be a 20-character Entity Identifier conforming to [ISO17422]. (See Appendix 11: ELEI Code Partitioning Scheme for details on the partitioning of the ELEI code.)

### 3.4.11 EntityCategory Data Type

A value of type `EntityCategory` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `EntityCategory` | `EntityCategoryTypeEnum` | 0,1 | Indicates the general category of the type of entity identified by this ELEI data record |

### 3.4.12 LanguageCode Data Type

A value of type `LanguageCode` in a file conforming to this standard SHALL be a 2-character language code conforming to [ISO639-1]. Note that ISO 639-1 language codes are all lowercase.

- TBD: Alternatively, we could use IETF language tags (RFC 4646) instead. IETF language tags can distinguish between variations of the same language in different countries. For example, the IETF language tags `fr-CA` and `fr-FR` denote Canadian French and French as spoken in France, respectively; in ISO 639-1 these would both be simply `fr`.

### 3.4.13 Name Data Type

A `Name` is a string expressed in a natural language, including a code indicating which natural language is used.

A value of type `Name` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `lang` | `LanguageCode` | 0,1 | The language of `name` |
| `Name` | `String` | 1 | The name itself. |

### 3.4.14 OtherAddress Data Type

A value of type `OtherAddress` in a file conforming to this standard SHALL include data elements as specified below. Each `Address` element includes an optional language code, permitting `OtherAddress` to be repeated as many times as necessary to express the same address type in multiple languages. The purpose of the `AddressTypeEnum` code list is to



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

accommodate legal address and headquarters address in different languages, not to add other address types (which could conceivably be added in the future).

| Element Name | Type | Card | Description |
|---|---|---|---|
| type | AddressTypeEnum | 1 | The type of address represented by this OtherAddress instance. |
| Address | Address | 1 | The address |

### 3.4.15 OtherEntityName Data Type

A value of type OtherEntityName in a file conforming to this standard SHALL include data elements as specified below. Each Name element includes an optional language code, permitting OtherEntityName to be repeated as many times as necessary to express the same name type in multiple languages. When type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, the language code specifies the language of the name prior to Romanization.

| Element Name | Type | Card | Description |
|---|---|---|---|
| type | EntityNameTypeEnum | 1 | The type of name represented by this OtherEntityName instance. The EntityNameType observes language, since 'Name' type has a language attribute. |
| Name | Name | 1 | The name. If type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, then this value SHALL only include characters from the character set specified in Section 9. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.16 RegistrationSource Data Type

| | | | |
|---|---|---|---|
| RegistrationSourceCitation | RegistrationSourceCitationType | 0,1 | A locator of the source of the entity reference data in this registration record (mechanism TBD, typically company yearbooks) |
| RegistrationSourceLocatorKey | RegistrationSourceLocatorKeyType | 0,1 | A locator of the source of the entity reference data in this registration record (mechanism TBD, typically company yearbooks) |
| RegistrationSourceLocatorAnchor | RegistrationSourceLocatorAnchorType | 0,1 | Additional information to allow pinpointing where in the RegistrationSource can the entity data be found |
| RegistrationSourceDCU | DCUID | 0,1 | The identifier of the Data Collection Unit that sourced the registration data |
| RegistrationSourceEntityID | String | 0,1 | An identifier of the entity that may have been locally used or assigned by the DCU at collection time |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.17 RegistrationSourceCitation Data Type

An associative locator (link or key) that resolves to a DDI metadata block regarding the source of the entity reference data in this registration record (mechanism TBD, typically company yearbooks -)

DDI 3.2 uses the `CitationType` Element for referencing publications (e.g. yearbooks). It contains 11 elements which are described in detail in the *DDI Lifecycle XML Schema*. The table below lists the suggested cardinality for each element and some notes on special uses. The name of each element is a link to its description in the DDI XML Schema.

| Element Name | Type | Card | Description |
|---|---|---|---|
| Title | DDI Title | 1,1 | |
| SubTitle | DDI SubTitle | 0,1 | |
| AlternateTitle | DDI AlternateTitle | 0,n | |
| Creator | DDI Creator | 0,n | |
| Publisher | DDI Publisher | 0,1 | |
| Contributor | DDI Contributor | 0,n | |
| PublicationDate | DDI PublicationDate | 0,1 | For monographs: SimpleDate<br>For serials (e.g. yearbooks): StartDate and EndDate (the date/year when the first and the last volume of a serial were published – the specific volume and page numbers of the year from which information about the entity was taken can then be recorded in the RegistrationSourceLocatorAnchor element) |
| Language | DDI Language | 0,n | |
| InternationalIdentifier | DDI InternationalIdentifier | 0,n | |
| Copyright | DDI Copyright | 0,1 | |
| dc:isPartof | DDI dc:isPartof | 0,1 | Use for instance to record the title of the newspaper in which a stock exchange price list is published |

The Citation as we would want to use it in the Entity Reference Data would be considered as a DataSource in DDI 3.2 (Origin element).



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.18 RegistrationSourceLocatorAnchorType Data Type

The `RegistrationSourceLocatorAnchorType` provides additional information to allow pinpointing where in the RegistrationSource referenced by the `RegistrationSourceLocatorKey` can the entity data be found

### 3.4.19 RegistrationSourceLocatorKeyType Data Type

An associative locator (link or key) that resolves to a DDI metadata block regarding the source of the entity reference data in this registration record (mechanism TBD, typically company yearbooks -)

## 3.5 Enumerated Code Lists

This section specifies the enumerated code list data types (all having the suffix `Enum`) referenced by the tables in Sections 3.2 and 0, in alphabetical order.

### 3.5.1 AddressTypeEnum Code List

The `AddressTypeEnum` value in an `OtherAddress` instance specifies how the alternative address relates to the entity.

A value of type `AddressTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `LEGAL_ADDRESS` | Registered address of the entity in the legal jurisdiction |
| `HEADQUARTERS_ADDRESS` | Address of the headquarters of the entity |

### 3.5.2 ELegalFormEnum Code List

The value of `ELegalFormEnum` SHALL include codes published in the ISO Entity Legal Form (ELF) code list maintained by GLEIF, but with additions needed for historical variations

- TBD: Research on additions and extensions to ELF code list is needed, including start-stop date periods and historical jurisdictions (referencing extensions to the Country Code external code list)



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.5.3 ERegistrationAuthorityEnum Code List

The `RegistrationAuthorityEnum` value in an `Entity` instance specifies what business register provided the value of `RegistrationAuthorityEntityID` for the entity.

The value of `ERegistrationAuthorityEnum` SHALL be a code published in a list of registration authority codes.

- TBD: the external code list for Registration Authorities needs to be developed, but will borrow from, incorporate or reference the Registration Authorities Code List published by the Global LEI Foundation.
- TBD: this may become a metadata reference to the source document (e.g., yearbook) that provided the information about the Entity

### 3.5.4 EntityCategoryTypeEnum

The `EntityCategoryTypeEnum` value in an ELEI record specifies the general classification category of the entity.

- TBD: the table below is just an example with some possible categories, and is incomplete, and needs to be elaborated.

A value of type `EntityCategoryTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

- TBD: Alternative to "Corporation" ? (Note: these enumerated codes are currently used by the GLOBAL LEI System – changing them creates differences with the current entity identification system) [ http://dx.doi.org/10.1080/00076791.2013.837893 ]

| Code | Definition |
|---|---|
| PUBLIC_CORPORATION | Public shareholder entity |
| PRIVATE_CORPORATION | Privately owned limited liability corporation |
| LLC | Privately owned limited liability company |
| SOLE_PROPRIETOR | Private business owned and operated by a single individual |
| LLP | Privately owned limited liability partnership |
| TRUST | Legal trust |
| FUND | Alternative Investment Scheme |
| GSE | Government sponsored entity |
| NGO | Non-Government Organization |
| GOV | Public Sector, Government Organization |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.5.5 EntityExpirationReasonEnum Code List

The `EntityExpirationReasonEnum` value in an ELEI record specifies the reason that the entity expired.

A value of type `EntityExpirationReasonEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `DISSOLVED` | The entity ceased to exist, and was legally shuttered. |
| `CORPORATE_ACTION` | The entity was acquired or merged with another entity |
| `OTHER` | The reason for expiry is not one of the above |

### 3.5.6 EntityNameTypeEnum Code List

The `EntityNameTypeEnum` value in an `EntityName` specifies how the name relates to the entity.

A value of type `EntityNameTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `OTHER_LEGAL` | Registered name of the entity in an alternate language in the legal jurisdiction in which the entity is registered |
| `PREFERRED_ROMANIZED_LEGAL` | Romanized form, preferred by the entity |
| `AUTO_ROMANIZED_LEGAL` | Romanized form, auto-transliterated |

### 3.5.7 EntityStatusEnum Code List

The `EntityStatusEnum` value in an ELEI record indicates the status of the entity itself. This is not to be confused with the status of the ELEI registration, which is specified by `RegistrationStatusEnum` (Section 3.4.8). See also Section 10, which illustrates how the `EntityStatusEnum` value changes over the lifecycle of an ELEI registration.

A value of type `EntityStatusEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `ACTIVE` | As of the last report or update, the entity reported is legally registered and operating. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| INACTIVE | It has been determined that the entity that was assigned the ELEI is no longer legally registered and/or operating, whether as a result of:<br>1. Business closure<br>2. Acquisition by or merger with another (or new) entity<br>3. Determination of illegitimacy [perhaps not required for historical entities] |
| NOT_SPECIFIED | The ELEI record is in a state in which it does not provide information about whether the entity is legally registered and operating. |

### 3.5.8 RegistrationStatusEnum Code List

The RegistrationStatusEnum value in an ELEI record indicates the status of the registration of the entity with an DSU. This is not to be confused with the status of the ELEI itself, which is specified by EntityStatusEnum (Section 3.5.7). See also Section 10, which illustrates how the EntityStatusEnum value changes over the lifecycle of an ELEI registration.

A value of type RegistrationStatusEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| PENDING | An application for an ELEI that has been submitted and which is being processed and validated.<br>*NOTE: ELEI registrations in the PENDING state are not intended for public release, but could be used internally between DSUs.* |
| ISSUED | An ELEI Registration that has been validated and issued, and which identifies ggal entity that was an operating entity as of the last update. |
| DUPLICATE | An ELEI Registration that has been determined to be a duplicate registration of the same entity as another ELEI Registration; the DUPLICATE status is assigned to the non-surviving registration (i.e., the ELEI that should no longer be used). Only one of the potential multiple identifiers will survive; for all other duplicate registrations:<br>1. The RegistrationStatus is set to DUPLICATE,<br>2. The ELEI of the surviving ELEI Registration is set in the SuccessorELEI data element of (each) duplicate ELEI registration;<br>3. The LastUpdateDate is set to reflect the date of this update, and<br>4. No further updates of the DUPLICATE registration record will occur. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| MERGED | An ELEI registration for a entity that has been merged into another entity, such that this entity no longer exists as an operating entity.<br>If<br>• After being issued an ELEI, the entity is acquired by, or merged with, another entity;<br>• Per agreements among the parties to the transaction, the ELEI of the acquired or merged entity will not be used to identify the surviving entity (or if a new entity is created that is issued a new ELEI)<br>Then<br>1. The ELEIRegistrationStatus is set to "MERGED",<br>2. The ELEI of the surviving/new entity is set in the successorELEI data element of (each) ELEI registration that is no longer to be used;<br>3. The ELEIRecordLastUpdate is set to reflect the date of this update, and<br>4. No further updates of the MERGED registration record(s) will occur. |
| RETIRED | An ELEI registration for a entity that has ceased operation, without having been merged into another entity.<br>If<br>• The responsible DSU determines by public sources that the entity has been dissolved or ceased to operate (and the DSU seeks to confirm this status through all available channels)<br>Then<br>1. The ELEIRegistrationStatus is set to "RETIRED";<br>2. The ELEIRecordLastUpdate is set to reflect the date of this update;<br>3. The EntityExpirationDate is also set to the date of this update;<br>4. The EntityExpirationReason is set;<br>5. No further updates of the RETIRED registration record will occur. |

### 3.5.9 ValidationSourcesEnum Code List

A value of type ValidationSourcesEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| PENDING | The validation of the reference data for the entity has not yet occurred. |
| SUBMITTED | Based on the validation procedures in use by the DSU responsible for the record, the information associated with this record has significant reliance on the information that a submitter provided due to the unavailability of corroborating information. |
| PARTIALLY_CORROBORATED | Based on the validation procedures in use by the DSU responsible for the record, the information supplied for the entity can be partially corroborated by available sources, while some of the record is dependent upon information that the submitter collected, either due to conflicts with authoritative information, or due to data unavailability. |
| FULLY_CORROBORATED | Based on the validation procedures in use by the DSU responsible for the record, there is sufficient information contained in authoritative public sources to corroborate the information regarding the entity provided in the record. |

.

# 4 Constraints and Data Validation

All values of type `String` specified in Section 3 SHALL be 500 or fewer characters in length.

- TBD. Specify here any additional constraints not implicit in the data types or explicit in the definition of each reference data element

# 5 XML Syntax

This section specifies the XML schema for an ELEI data file conforming to this standard.

## 5.1 XML Design Rules

- The XSD schema conforms to [XSD1,XSD2]
- The XML namespace is `http://www.EURHISFIRM.eu/schema/ELEIdata/1`
- All interior elements and attributes are namespace-qualified (element/attribute form = qualified)
- Element names are upper camel case
- Attribute name are lower camel case
- XSD type names are upper camel case
- Enumeration code list values are all caps with underscores



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- Elements are used in preference to attributes *except* for language and type qualifiers
- For a data element specified in Section 3 as having unbounded cardinality, the XML includes a single container element whose sub-elements are one or more instances of the data element whose cardinality is unbounded. The name of the container element is formed as the plural of the name of the contained elements.

➢ TBD: specify the approach to enumeration types, recognizing the need for change management. See https://www.ibm.com/developerworks/library/x-extenum/ for various possible approaches.

➢ TBD: finalize the XML namespace

➢ TBD: take a final decision on whether to use element/attribute form = qualified or unqualified

# 6 Change Management

➢ TBD. Use the <vnext> strategy for forward/backward compatibility, and/or include a version number in the XML. Needs to be explained

# 7 Examples (non-normative)

➢ TBD.

# 8 References

[ISO646] ISO, "Information technology -- ISO 7-bit coded character set for information interchange," ISO/IEC 646:1991.

[ISO3166-1] ISO, "Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes," ISO 3166-1:2013.

[ISO3166-2] ISO, "Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code," ISO 3166-2:2013.

[ISO8601] ISO, "Data elements and interchange formats – Information interchange – Representation of dates and times," ISO 8601:2004.

[ISO10646] ISO, "Information technology -- Universal Coded Character Set (UCS)," ISO 10646:2012.

[ISO17422] ISO, "Financial Services - Entity Identifier (LEI)," ISO/DIS 17442:2012.

[ISODir2] ISO, "Rules for the structure and drafting of International Standards (ISO/IEC Directives, Part 2, 2001, 4th edition)," July 2002.

[XSD1] H. Thompson, D. Beech, M. Maloney, N. Mendelsohn, "XML Schema Part 1: Structures," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-1/.

[XSD2] P. Biron, A. Malhotra, "XML Schema Part 2: Datatypes," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-2/.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 9 Appendix: Character Codes Allowed in Romanized Names

When a `Name` instance is of type `PREFERRED_ROMANIZED` or `AUTO_ROMANIZED`, the value of the `name` field SHALL consist only of non-control characters drawn from the "invariant subset" of ISO 646. These characters are enumerated below. The "Hex Value" column indicates the code point value (expressed in hexadecimal) for each character in both ISO 646 and ISO 10646.

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| ! | Exclamation Mark | 21 | M | Capital Letter M | 4D |
| " | Quotation Mark | 22 | N | Capital Letter N | 4E |
| % | Percent Sign | 25 | O | Capital Letter O | 4F |
| & | Ampersand | 26 | P | Capital Letter P | 50 |
| ' | Apostrophe | 27 | Q | Capital Letter Q | 51 |
| ( | Left Parenthesis | 28 | R | Capital Letter R | 52 |
| ) | Right Parenthesis | 29 | S | Capital Letter S | 53 |
| * | Asterisk | 2A | T | Capital Letter T | 54 |
| + | Plus sign | 2B | U | Capital Letter U | 55 |
| , | Comma | 2C | V | Capital Letter V | 56 |
| - | Hyphen/ Minus | 2D | W | Capital Letter W | 57 |
| . | Full Stop | 2E | X | Capital Letter X | 58 |
| / | Solidus | 2F | Y | Capital Letter Y | 59 |
| 0 | Digit Zero | 30 | Z | Capital Letter Z | 5A |
| 1 | Digit One | 31 | _ | Low Line | 5F |
| 2 | Digit Two | 32 | a | Small Letter a | 61 |
| 3 | Digit Three | 33 | b | Small Letter b | 62 |
| 4 | Digit Four | 34 | c | Small Letter c | 63 |
| 5 | Digit Five | 35 | d | Small Letter d | 64 |
| 6 | Digit Six | 36 | e | Small Letter e | 65 |
| 7 | Digit Seven | 37 | f | Small Letter f | 66 |
| 8 | Digit Eight | 38 | g | Small Letter g | 67 |
| 9 | Digit Nine | 39 | h | Small Letter h | 68 |
| : | Colon | 3A | i | Small Letter i | 69 |
| ; | Semicolon | 3B | j | Small Letter j | 6A |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

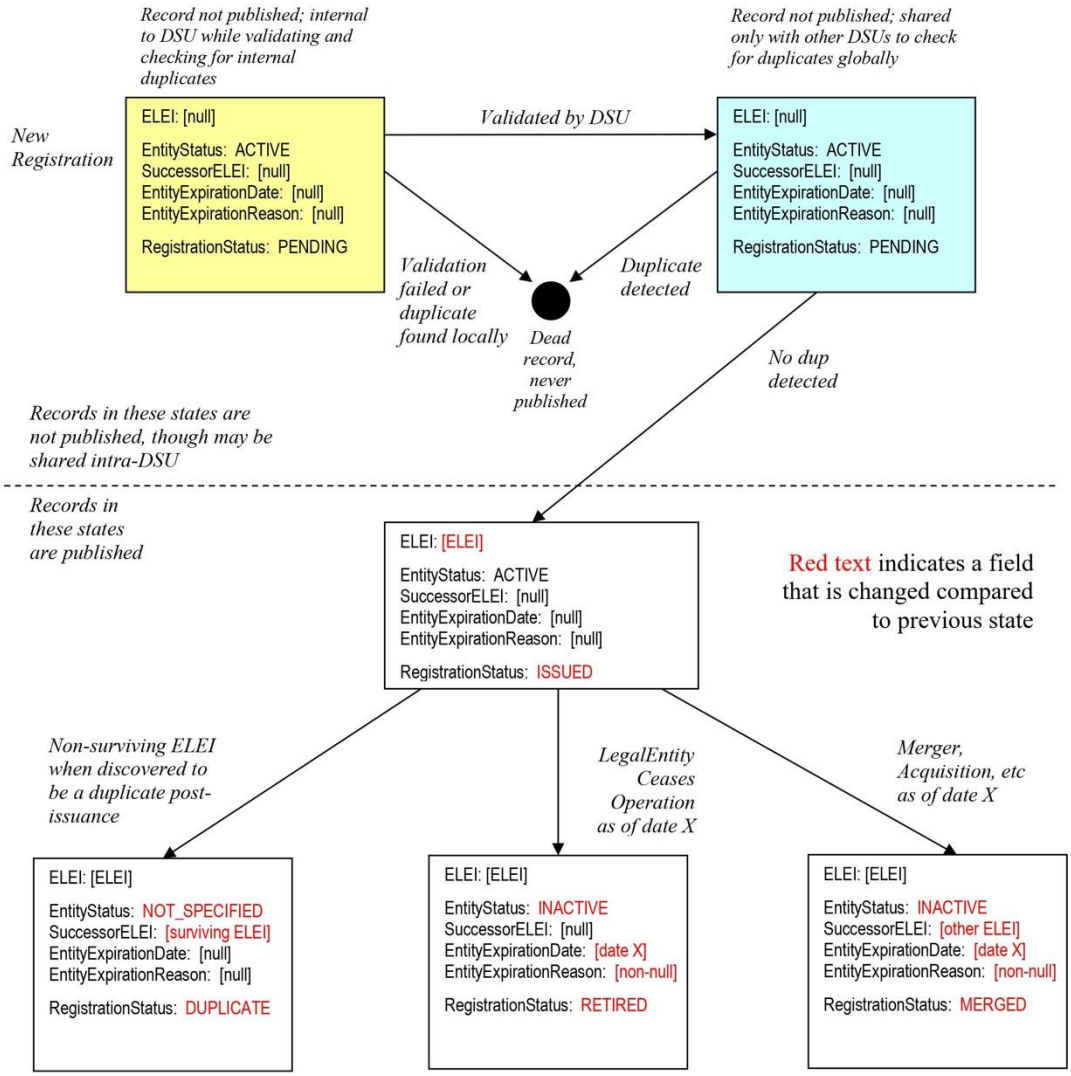
| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| < | Less-than Sign | 3C | k | Small Letter k | 6B |
| = | Equals Sign | 3D | l | Small Letter l | 6C |
| > | Greater-than Sign | 3E | m | Small Letter m | 6D |
| ? | Question Mark | 3F | n | Small Letter n | 6E |
| A | Capital Letter A | 41 | o | Small Letter o | 6F |
| B | Capital Letter B | 42 | p | Small Letter p | 70 |
| C | Capital Letter C | 43 | q | Small Letter q | 71 |
| D | Capital Letter D | 44 | r | Small Letter r | 72 |
| E | Capital Letter E | 45 | s | Small Letter s | 73 |
| F | Capital Letter F | 46 | t | Small Letter t | 74 |
| G | Capital Letter G | 47 | u | Small Letter u | 75 |
| H | Capital Letter H | 48 | v | Small Letter v | 76 |
| I | Capital Letter I | 49 | w | Small Letter w | 77 |
| J | Capital Letter J | 4A | x | Small Letter x | 78 |
| K | Capital Letter K | 4B | y | Small Letter y | 79 |
| L | Capital Letter L | 4C | z | Small Letter z | 7A |
| | Space | 20 | | | |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 10 Appendix: ELEI Record Transition Diagram

The following diagram illustrates the relationship between status fields of the ELEI record.




In rare circumstances, a record may pass directly from `PENDING` to `RETIRED` or `MERGED`; e.g., retroactive reporting of a failed or merged legal g. In that case, the `EntityStatus` would be `INACTIVE` while the record is in the `PENDING` state.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 777489

## 11 Appendix: ELEI Code Partitioning Scheme

The following diagram illustrates the partitioning scheme of the ELEI code. The ELEI code is an ISO 17442-compliant, opaque identifier that resolves to the EURHISFIRM entity reference data that uniquely identifies a entity in the EURHISFIRM network.

The ELEI is minted by a Data Submission Unit (DSU) as part of the process that would add the ELEI -- and the unique-identifying reference data -- for a entity to the EURHISFIRM network under the EURHISFIRM Common Data Model (CDM) data standard.

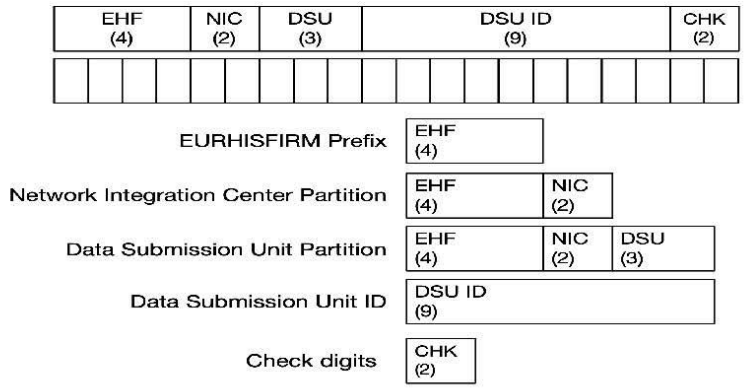
The DSU should, if possible, perform a check with the consolidated database of existing EURHISFIRM ELEI identifier reference data to verify that a entity has not already been identified and been assigned an ELEI before submitting a newly minted ELEI to a Network Integration Center (NIC). Otherwise, the Network Integration Center would need to perform this check for the existence of an ELEI that had already been assigned to the entity in question in order to maintain the uniqueness of the ELEI assigned to the entity.

In cases where the NIC may not be able to do this, or in which a duplicative entry existed but was not found, remediation and cleanup of multiple ELEIs for the same entity could be detected and rectified by background data quality processes that periodically sweep the ELEI reference data or by challenges submitted from the EURHISFIRM community.

Note that the partitioning scheme does not add intelligence or actually identify any DSU or NIC, as multiple prefixes could be issued to a single DSU under one of multiple prefixes that may be issued to a NIC. The partitioning scheme is only for the purpose of guaranteeing the ability to concurrently create unique ELEI codes in a federated and decentralized entity data "supply chain" in the EURHISFIRM network. (The EHF is the prefix that would be assigned to EURHISFIRM by the Global LEI Foundation in order to be globally unique in the Global LEI System )

EURHISFIRM Legal Entity Identifier ( ELEI )
*(ISO 17742 Compliant)*

Code Partitioning Scheme






This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 9.4 Financial Instrument Identification Data Standard 1.05

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# EURHISFIRM Common Data Model Standard 1.0

## Financial Instrument Identification Data

Version of 22 July 2020

| Date | Revision | Description |
|---|---|---|
| 2020-05-30 | ver 1.01 | First Draft to be circulated (formatting fixed) |
| 2020-06-20 | ver 1.02 | Multiple revisions and edits in response to comments |
| 2020-07-07 | ver 1.03 | Market Sector enumeration data type added.<br>DataMaturityStage enumeration type added<br>Additional revisions and updates |
| 2020-07-09 | ver 1.04 | Revisions to DataMaturityStage enumerations |
| 2020-07-20 | ver 1.05 | Acceptance of changes |


## Abstract

First, the semantic content of these attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the EURHISFIRM Common Data Model.


## Status of this document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this document series will be maintained on SeaFile.

This draft is a **Working Draft** which can be circulated to any interested parties for review and comment. It is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to use Working Drafts as reference material or to cite



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

them as other than "work in progress." This is work in progress and does not imply endorsement by the EURHISFIRM ExCo.

Comments on this document should be sent to [TBD: insert mailing list or URL].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Table of Contents

1 Introduction 5
2 Terminology and Typographical Conventions ....6
3 Abstract Data Content ....6
3.1 EFII File Header ....7
3.2 EFII Root Record ....7
3.3 EFII Data Record ....8
3.3.1 Financial Instrument Section of EFII Data Record ....9
3.3.2 Registration Section of EFII Data Record ....13
3.3.3 Extension Section of EFII Data record ....14
3.4 Data Types ....14
3.4.1 DateTime Data Type ....15
3.4.2 DCUID Data Type ....15
3.4.3 DSUID Data Type ....16
3.4.4 ECountryCode Data Type ....16
3.4.5 ERegionCode Data Type ....16
3.4.6 ERegistrationAuthorityType Data Type ....16
3.4.7 EFII Data Type ....17
3.4.8 FinancialInstrumentCategory Data Type ....17
3.4.9 LanguageCode Data Type ....17
3.4.10 Name Data Type ....18
3.4.11 OtherFinancialInstrumentName Data Type ....18
3.4.12 RegistrationSource Data Type ....19
3.4.13 RegistrationSourceCitation Data Type ....20
3.4.14 RegistrationSourceLocatorAnchorType Data Type ....21
3.4.15 RegistrationSourceLocatorKeyType Data Type ....21
3.5 Enumerated Code Lists ....21
3.5.1 FinancialInstrumentCategoryTypeEnum ....21
3.5.2 FinancialInstrumentExpirationReasonEnum Code List ....22
3.5.3 FinancialInstrumentNameTypeEnum Code List ....22
3.5.4 FinancialInstrumentStatusEnum Code List ....23



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

3.5.5 MarketSectorEnum Code List ..... 23
3.5.6 RegistrationStatusEnum Code List ..... 24
3.5.7 DataMaturityStageEnum Code List ..... 26
4 Change Management ..... 27
5 Examples (non-normative) ..... 27
6 References ..... 27
7 Appendix: Character Codes Allowed in Romanized Names ..... 27
8 ..... 29



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 1 Introduction

Financial Instruments, and in particular market securities that have been publicly traded on exchanges, are clearly one of the core classes of objects in a historical financial Research Infrastructure such as EURHISFIRM, on which a substantial amount of other EURHISFIRM financial data depend. Such financial instruments should be uniquely and unambiguously identified in EURHISFIRM, and this identification is accomplished based on the identifying attributes associated with each financial instrument.

First, the semantic content of those attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the reference data attributes of financial instruments in the EURHISFIRM Common Data Model

This document proposes initial standards for EURHISFIRM financial instrument reference data. It is important that this reference data should uniquely identify the financial instruments that are harvested from contributing sources and then collected and assimilated into the common EURHISFIRM platform.

A EURHISFIRM Financial Instrument Identifier (EFII) code that resolves to this financial instrument-identifying reference data is also introduced.

- A classification hierarchy for the structure of this identifier, modeled after the Object Management Group Financial Instrument Global Identifier (FIGI) is defined that provides for the identification of publicly traded market instruments at three levels: a unique identifier of a financial instrument traded at a given exchange, an identifier of a financial instrument traded within a country (national jurisdiction), and an identifier for a financial instrument issued at a share class level[1] by the institution that created the instrument. This allows for the independent and concurrent minting and assignment of EFII codes to financial instrument reference data that is collected by multiple Data Collection Units (DCUs) working at the exchange level, and the subsequent global identification of the financial instrument in the federated EURHISFIRM Research Infrastructure network.

The standard set by this document is expected, among other things, to reduce the risk of duplicates stemming from differences in formats and conventions of locally sourced data provided by regional contributors (i.e., Data Collection Units, or DCUs), to ensure data quality in the EURHISFIRM system, and to enable the subsequent detection and resolution of multiple identifiers for the same financial instrument traded on different national exchanges. The standard is expected to be used as a format for reference data consolidated from all sources in order to be promoted to the level of common EURHISFIRM data published for end-user access.

The contents of this document are as follows:

[1] "Share class" should not be interpreted as limited only to securities with "shares" – it is instead intended to distinguish among different unique securities offerings made by issuers. This is just the terminology that FIGI uses, but we could provide clarifying language or terms that would not be misinterpreted.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

- Section 2 defines terminology and typographical conventions.
- Section 3 specifies the abstract content of EURHISFIRM financial instrument reference data conforming to this standard, including a detailed description of each element of reference data associated with financial instruments in EURHISFIRM. Allowable values for data elements that are code lists will also be (subsequently) specified. The partitioning scheme for the structure of the EURHISFIRM Financial Instrument Identifier (EFII) code will also be subsequently described.

## 2 Terminology and Typographical Conventions

Within this specification, the terms SHALL, SHALL NOT, SHOULD, SHOULD NOT, MAY, NEED NOT, CAN, and CANNOT are to be interpreted as specified in Annex G of the ISO/IEC Directives, Part 2, 2001, 4th edition [ISODir2]. When used in this way, these terms will always be shown in ALL CAPS; when these words appear in ordinary typeface they are intended to have their ordinary English meaning.

All sections of this document, with the exception of Section 1 are normative, except where explicitly noted as non-normative.

The following typographical conventions are used throughout the document:

- ALL CAPS type is used for the special terms from [ISODir2] enumerated above.
- `Monospace` type is used to denote programming language, UML, and XML identifiers, as well as for the text of XML documents.

➢ Placeholders for open issues and/or changes that need to be made to this document prior to its reaching the final stage of approved Proposed Standard are prefixed by a rightward-facing arrowhead, as this paragraph is.

The specifications of data types and elements include a column for "cardinality" (usually abbreviated 'CARD' ) that describes if the element is either mandatory or optional, and whether the number of occurrences of the element can be more than one, or just one.

## 3 Abstract Data Content

This section specifies the abstract data content of a data file conforming to this standard.

A data file conforming to this standard SHALL consist of:

- An optional EFII File Header, as specified in Section 3.1.
- An EFII Root Record as specified in Section 3.2.
- Zero or more EFII Data Records, as specified in Section 3.3.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 3.1 EFII File Header

➢ TBD Define elements to go into a "header" area for an EFII file. The purpose of the header is to provide context about the file and its contained EFII data records. However, the header will not contain anything necessary to interpret the meaning of any EFII record; e.g., things like default values for EFII data records will not be in the header (such things would mean that the meaning of an EFII record could change if taken away from the header). Examples of things that might be useful to include in the header:

| Element Name | Type | Card | Description |
|---|---|---|---|
| ContentDate | DateTime | 1,1 | The date and time of generation of the data |
| Originator | | 0,1 | The identifier of the creator of the content of this file |
| FileContent | | 1,1 | A code describing the content of this data file. |
| ProcessStage | | 0,1 | A code indicating the stage of this file in the EURHISFIRM workflow |
| RecordCount | | 1,1 | The number of data records in the file. Can be a positive whole (integer) number, or zero (0). |

## 3.2 EFII Root Record

The EFII Root Record anchors a single EFII. This root record is needed in order to provide a unique EFII anchor to support multiple (historical) EFII Data Records should there be changes in aspects of the security or its reference data (e.g., name change, stock split, etc) that require keeping a record. Each EFII Root record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| EFII | EFII | 1 | The 12-character EFII of the financial instrument described by this EFII Data Record. |
| EFIICreationDate | DateTime | 1 | Date/time the EFII root record was initially created in the system |
| FinancialInstrumentStatus | FinancialInstrumentStatusEnum | 1 | The status of the financial instrument. This is not to be confused with the status of the registration; see RegistrationStatus. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| FinancialInstrumentIssueDate | DateTime | 0,1 | Date/time the Financial Instrument was issued, if known |
| FinancialInstrumentExpirationDate | DateTime | 0,1 | The date that the financial instrument ceased to exist, whether due to dissolution, merger or acquisition of the issuing firm, or the maturity, redemption or other action in the terms of the security. Omitted if the financial instrument has not ceased to exist, or if this EFII record contains a non-empty SuccessorEFII field. |
| FinancialInstrumentExpirationReason | FinancialInstrumentExpirationReasonEnum | 0,1 | The reason that a financial instrument ceased to exist. This element SHALL be present if FinancialInstrumentExpirationDate is present, and omitted otherwise. |
| EarliestRegistrationDate | DateTime | 1 | Date/time of the Registration record with the earliest historical date |
| LatestRegistrationDate | DateTime | 1 | Date/time of the Registration record with the most recent historical date |

## 3.3 EFII Data Record

An EFII Data Record describes a single EFII. Each EFII Data record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| EFII | EFII | 1 | The 12-character EFII of the financial instrument described by this EFII Data Record. |
| FinancialInstrument | FinancialInstrument (Section 3.3.1) | 1 | Attributes describing the financial instrument itself |
| Registration | Registration (Section 0) | 1 | Attributes describing the registration of this EFII. |
| Extension | Extension (Section 0) | 0,1 | An optional element for including data beyond the standard data elements in an EFII data file. This may include data specific to an DSU, data specific to a publisher of EFII data, and so on. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.3.1 Financial Instrument Section of EFII Data Record

The `Instrument` section of an EFII Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| | | | |
| `InstrumentName` | `Name` | 1,1 | The Registered Name of the Instrument, exactly as it appears on the official list.<br>If an Instrument is in a jurisdiction with more than one Registered Name (e.g., in different languages), this is the Primary Registered Name (see `OtherInstrumentName` for other names). |
| `OtherInstrumentName` | `Name` | 0,n | Other registered names of the financial instrument, whether in different languages in the jurisdiction of the legal issuer, or in different languages used in exchanges in foreign jurisdictions. `OtherInstrumentNameType` enumerated attributes will qualify the type of each `OtherInstrumentName` |
| `Ticker` | `Name` | | Ticker is a specific identifier for a financial instrument that reflects common usage. Tickers are not, however, unique to specific exchanges or specific pricing sources. Tickers may change in conjunction with Corporate Actions. |
| `ExchangeCode` | `ExchangeCodeEnum` | | Code for the trading venue or environment on which the instrument trades. If an exchange is specified, the code will be for the specified exchange. When not specified, the code will be according to the user default exchange, which can be the composite or primary exchange. |
| `SecurityType` | `SecurityTypeEnum` | | Classification (enumerated type) of the instrument type within its market sector. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| MarketSector | MarketSectorEnum | | Market Sector refers to the classification of the asset type of the instrument.<br><br>( For example: Commodity, Equity, Municipals ,Preferred, Money Market, Government, Corporate, Index , Currency, Mortgage) |
| LocalEFII | EFII | | Twelve character, alphanumeric identifier. The first 2 characters are upper-case<br><br>consonants (including "Y"), the third character is the upper-case "G", characters 4 -11 are any upper-case consonant (including "Y") or integer between 0 and 9, and the last character is a check-digit. An identifier is assigned to instruments of all asset classes, is unique to an individual instrument and once issued will not change for an instrument. For equity instruments an identifier is issued per instrument per trading venue (e.g., stock exchange or other marketplace with listed securities). |
| CompositeEFII | | | Twelve character, alphanumeric identifier. The first 2 characters are upper-case<br><br>consonants (including "Y"), the third character is the upper-case "G", characters 4 -11 are any upper-case consonant (including "Y") or integer between 0 and 9, and the last character is a check-digit. The Composite level of assignment is provided in cases where there are multiple trading venues for the instrument within a single country or market. The Composite Eurhisfirm Financial Instrument Identifier (EFII) enables users to link multiple EFIIs at the trading venue-level within the same country or market in order to obtain an aggregated view for that instrument within that country or market. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| ShareClassEFII | | | Twelve character, alpha-numeric identifier. The first 2 characters are upper-case consonants (including "Y"), the third character is the upper-case "G", characters 4 -11 are any upper-case consonant (including "Y") or integer between 0 and 9, and the last character is a check-digit. A Share Class level Eurhisfirm Financial Instrument Identifier is assigned to an instrument that is traded in more than one country. This enables users to link multiple Composite EFIIs for the same instrument in order to obtain an aggregated view for that instrument across all countries globally. |
| SecurityDescription | | | A description of the security |
| SecurityShortDescription | | | Alternate Short Description for a given security comprised of the ticker, coupon and maturity year (YY). For strips/scripts it returns the ticker, coupon, and maturity (M/YY). |
| OtherInstrumentNames | OtherInstrumentName | 0..n | An optional list of other Name instances for the Instrument. |
| LegalJurisdiction | ERegionCode | 0,1 | The jurisdiction of legal issuance and registration of the financial instrument (and on which the legalForm data element is also dependent). |
| InstrumentCategory | InstrumentCategoryTypeEnum | 0,1 | Indicates the general category of the type of financial instrument identified by this EFII data record |
| FinancialInstrumentStatus | FinancialInstrumentStatusEnum | 1 | The status of the financial instrument. This is not to be confused with the status of the registration; see RegistrationStatus. |
| FinancialInstrumentEvents | FinancialInstrumentEvent | 0,n | Corporate events that occurred during this historical period |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.3.2 Registration Section of EFII Data Record

The `Registration` section of an EFII Data Record in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `RegistrationCreationDate` | `DateTime` | 1 | Date/time this EFII record was initially created in the system |
| `RegistrationUpdateDate` | `DateTime` | 1 | Date/time that this historical EFII record was most recently updated in the system. |
| `RegistrationSource` | `RegistrationSourceType` | 0,n | A locator of the source of the financial instrument reference data in this registration record (mechanism TBD, typically company yearbooks, official price lists, and exchange registrations) |
| `RegistrationStatus` | `RegistrationStatusEnum` | 1 | Status of the EFII registration. This is not to be confused with the status of the financial instrument itself; see `FinancialInstrumentStatus`. |
| `RegistrationUpToDate` | `DateTime` | 0,1 | Should it exist and be known, the historical date up until which time the EFII identifying reference data for this financial instrument is valid. (Not the same thing as when the data was changed in the system). If present, a subsequent historical record for this EFII may exist with the revised information. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| RegistrationAsOfDate | DateTime | 0,1 | The date that this version of the reference data is known to be valid. This allows a record of historical changes to the financial instrument identifying data to be recorded. This date would typically be derived from either:<br>1. The publication date of the source reference, or<br>2. Information contained in the source reference |
| ResponsibleDSU | DSUID | 1 | The Identifier of the Data Submission Unit (DSU) that produced and manages this EFII registration. |
| DataMaturityStage | DataMaturityStageEnum | 0,1 | The current data maturity stage of this EFII record, or omitted if the data maturity stage is not known. |

### 3.3.3 Extension Section of EFII Data record

The Extension section of an EFII record may be used to include additional data not defined in this standard. For example, a DSU may use Extension to publish additional data elements it collects as part of registration.

- TBD: include the details of how this works. Basically, the idea is to use an XSD schema wildcard with namespace ##other, permitting the inclusion of XML elements from other XML namespaces.

## 3.4 Data Types

This section specifies the data types referenced by the tables in Section 3.2, in alphabetical order.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.1 DateTime Data Type

A value of type `DateTime` in a conforming to this standard SHALL be a point in time expressed as a string conforming to ISO 8601 having the following format:

- TBD: Provision for different calendars ? (i.e., Gregorian, Julian, etc. )

*YYYY*`-`*MM*`-`*DD*`T`*hh*`:`*mm*`:`*ss.sss*`TZ`

where the components of the above string are as follows:

- *YYYY* is the year
- *MM* is the month (01 = January, …, 12 = December)
- *DD* is the day of the month (01 = first day of the month)
- `T` is the single character 'T'
- *hh* is the hour (00 – 23)
- *mm* is the minute
- *ss.sss* is the second and milliseconds. From one to three digits may be used for milliseconds, or omitted entirely along with the decimal point.
- *TZ* is the time zone specifier, which can be either:
  - `Z` the single character 'Z', denoting Coordinated Universal Time (UTC); or
  - `+`*hh*`:`*mm* denoting a positive offset from UTC; or
  - `-`*hh*`:`*mm* denoting a negative offset from UTC

In the XML representation specified in Section **Error! Reference source not found.**, the XSD type `xs:dateTime` is used; however, whereas `xs:dateTime` permits the time zone specifier to be omitted, `DateTime` values in files conforming to this standard SHALL always include a time zone specifier.

*Explanation (non-normative): milliseconds are hardly necessary for EFII reference data, and likewise it might seem simpler to allow only "Z" as a time zone specifier; however, XML processing tools support the full syntax given above and it is not always possible to restrict such tools to avoid milliseconds or force the use of "Z" as the time zone specifier. The restriction that the time zone specifier must be present is equivalent to using XSD type `xs:timestamp`; however this was introduced in XSD 1.1 and not supported by the majority of XML processing tools which still only implement XSD 1.0.*

### 3.4.2 DCUID Data Type

A value of type `DCUID` in a file conforming to this standard SHALL be a [TBD]-character Data Collection Unit Identifier conforming to [TBD].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.3 DSUID Data Type

A value of type `DSUID` in a file conforming to this standard SHALL be a [TBD]-character Data Submission Unit Identifier conforming to [TBD].

### 3.4.4 ECountryCode Data Type

A value of type `CountryCode` in a file conforming to this standard SHALL be a 2-character country code conforming to ISO 3166-1 alpha-2 [ISO3166]. Note that ISO 3166-1 alpha-2 codes are all uppercase.

- TBD: CountryCode external code list must extend ISO country codes with start/stop dates and historical changes

### 3.4.5 ERegionCode Data Type

A value of type `RegionCode` in a file conforming to this standard SHALL be a code conforming to ISO 3166-2. Note that ISO 3166-2 codes are all uppercase.

- TBD: Like CountryCode code list, RegionCode external code list needs historical extensions and additions as well.

### 3.4.6 ERegistrationAuthorityType Data Type

A value of type `ERegistrationAuthority` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `ERegistrationAuthorityID` | `ERegistrationAuthorityEnum` | 0,1 | An identifier for the financial instrument **registry** of the financial instrument in the jurisdiction of legal registration, or in the appropriate registration authority. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EOtherRegistrationAuthority | String | 0,1 | A legacy / historical reference code of a registration authority which is not yet entered in the ERegistration Authorities List (RAL), or the designation of an interim register until such time as an entry from RAL can be delivered<br><br>. |

### 3.4.7 EFII Data Type

A value of type EFII in a file conforming to this standard SHALL be a 12-character Financial Instrument Identifier conforming to [OMG FIGI].

### 3.4.8 FinancialInstrumentCategory Data Type

A value of type FinancialInstrumentCategory in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| FinancialInstrumentCategory | FinancialInstrumentCategoryTypeEnum | 0,1 | Indicates the general category of the type of financial instrument identified by this EFII data record |

### 3.4.9 LanguageCode Data Type

A value of type LanguageCode in a file conforming to this standard SHALL be a 2-character language code conforming to [ISO639-1]. Note that ISO 639-1 language codes are all lowercase.

- TBD: Alternatively, we could use IETF language tags (RFC 4646) instead. IETF language tags can distinguish between variations of the same language in different countries. For example, the IETF language tags fr-CA and fr-FR denote Canadian French and French as spoken in France, respectively; in ISO 639-1 these would both be simply fr.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.10 Name Data Type

A `Name` is a string expressed in a natural language, including a code indicating which natural language is used.

A value of type `Name` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `lang` | `LanguageCode` | 0,1 | The language of `name` |
| `Name` | `String` | 1 | The name itself. |

### 3.4.11 OtherFinancialInstrumentName Data Type

A value of type `OtherFinancialInstrumentName` in a file conforming to this standard SHALL include data elements as specified below. Each `Name` element includes an optional language code, permitting `OtherFinancialInstrumentName` to be repeated as many times as necessary to express the same name type in multiple languages. When `type` is `PREFERRED_ROMANIZED_LEGAL` or `AUTO_ROMANIZED_LEGAL`, the language code specifies the language of the name prior to Romanization.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `type` | `FinancialInstrument NameTypeEnum` | 1 | The type of name represented by this `OtherFinancialInstrumentName` instance. The FinancialInstrumentNameType observes language, since 'Name' type has a language attribute. |
| `Name` | `Name` | 1 | The name. If `type` is `PREFERRED_ROMANIZED_LEGAL` or `AUTO_ROMANIZED_LEGAL`, then this value SHALL only include characters from the character set specified in Section 7. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.12 RegistrationSource Data Type

| | | | |
|---|---|---|---|
| RegistrationSourceCitation | RegistrationSourceCitationType | 0,1 | A locator of the source of the financial instrument reference data in this registration record (mechanism TBD, typically company yearbooks, official price lists, or exchange registrations) |
| RegistrationSourceLocatorKey | RegistrationSourceLocatorKeyType | 0,1 | A locator of the source of the financial instrument reference data in this registration record (mechanism TBD, typically company yearbooks) |
| RegistrationSourceLocatorAnchor | RegistrationSourceLocatorAnchorType | 0,1 | Additional information to allow pinpointing where in the RegistrationSource of the financial instrument data can be found |
| RegistrationSourceDCU | DCUID | 0,1 | The identifier of the Data Collection Unit that sourced the registration data |
| RegistrationSourceInstrumentID | String | 0,1 | An identifier of the financial instrument that may have been locally used or assigned by the DCU at collection time |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.13 RegistrationSourceCitation Data Type

An associative locator (link or key) that resolves to a DDI metadata block regarding the source of the financial instrument reference data in this registration record (mechanism TBD, typically company yearbooks -)

DDI 3.2 uses the `CitationType` Element for referencing publications (e.g. yearbooks). It contains 11 elements which are described in detail in the *DDI Lifecycle XML Schema*. The table below lists the suggested cardinality for each element and some notes on special uses. The name of each element is a link to its description in the DDI XML Schema.

| Element Name | Type | Card | Description |
|---|---|---|---|
| Title | DDI Title | 1,1 | |
| SubTitle | DDI SubTitle | 0,1 | |
| AlternateTitle | DDI AlternateTitle | 0,n | |
| Creator | DDI Creator | 0,n | |
| Publisher | DDI Publisher | 0,1 | |
| Contributor | DDI Contributor | 0,n | |
| PublicationDate | DDI PublicationDate | 0,1 | For monographs: SimpleDate<br>For serials (e.g. yearbooks): StartDate and EndDate (the date/year when the first and the last volume of a serial were published – the specific volume and page numbers of the year from which information about the financial instrument was taken can then be recorded in the RegistrationSourceLocatorAnchor element) |
| Language | DDI Language | 0,n | |
| InternationalIdentifier | DDI InternationalIdentifier | 0,n | |
| Copyright | DDI Copyright | 0,1 | |
| dc:isPartof | DDI dc:isPartof | 0,1 | Use for instance to record the title of the newspaper in which a stock exchange price list is published |

The Citation as we would want to use it in the Financial Instrument Reference Data would be considered as a DataSource in DDI 3.2 (Origin element).



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.4.14 RegistrationSourceLocatorAnchorType Data Type

The `RegistrationSourceLocatorAnchorType` provides additional information to allow pinpointing where in the RegistrationSource referenced by the `RegistrationSourceLocatorKey` can the financial instrument data be found

### 3.4.15 RegistrationSourceLocatorKeyType Data Type

An associative locator (link or key) that resolves to a DDI metadata block regarding the source of the financial instrument reference data in this registration record (mechanism TBD, typically company yearbooks -)

## 3.5 Enumerated Code Lists

This section specifies the enumerated code list data types (all having the suffix `Enum`) referenced by the tables in Sections 3.2 and 0, in alphabetical order.

### 3.5.1 FinancialInstrumentCategoryTypeEnum

The `FinancialInstrumentCategoryTypeEnum` value in an EFII record specifies the general classification category of the financial instrument.

- TBD: the table below is just an example with some possible categories, and is incomplete, and needs to be elaborated.

A value of type `FinancialInstrumentCategoryTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| | To Be Determined |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.5.2 FinancialInstrumentExpirationReasonEnum Code List

The `FinancialInstrumentExpirationReasonEnum` value in an EFII record specifies the reason that the financial instrument expired.

A value of type `FinancialInstrumentExpirationReasonEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| DISSOLVED | The financial instrument ceased to exist because the issuing entity was dissolved, and was legally shuttered. |
| CORPORATE_ACTION | The financial instrument was merged with another financial instrument due to an acquisition or merger of the issuing entity. |
| MATURITY | The terms of the financial instrument provide for its expiration and settlement |
| OTHER | The reason for expiry is not one of the above |

### 3.5.3 FinancialInstrumentNameTypeEnum Code List

The `FinancialInstrumentNameTypeEnum` value in an `FinancialInstrumentName` specifies how the name relates to the financial instrument.

A value of type `FinancialInstrumentNameTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| REGISTERED_NAME | Primary Registered Name of the security as designated by the issuer |
| OTHER_REGISTERED_NAME | Registered name of the financial instrument in an alternate language in the legal jurisdiction in which the financial instrument is registered |
| FOREIGN_REGISTERED_NAME | Primary Registered Name of the financial instrument on a trading venue in a sovereign jurisdiction that is not the jurisdiction where the financial instrument is registered |
| OTHER_FOREIGN_REGISTERED_NAME | Alternate Registered Name of the financial instrument on a trading venue in a sovereign jurisdiction that is not the jurisdiction where the financial instrument is registered |
| | |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| AUTO_ROMANIZED_LEGAL | Romanized form, auto-transliterated |

### 3.5.4 FinancialInstrumentStatusEnum Code List

The FinancialInstrumentStatusEnum value in an EFII record indicates the status of the financial instrument itself. This is not to be confused with the status of the EFII registration, which is specified by RegistrationStatusEnum (Section 3.4.8). See also Section **Error! Reference source not found.**, which illustrates how the FinancialInstrumentStatusEnum value changes over the lifecycle of an EFII registration.

A value of type FinancialInstrumentStatusEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| ACTIVE | As of the last report or update, the financial instrument reported is legally registered and operating. |
| INACTIVE | It has been determined that the financial instrument that was assigned the EFII is no longer legally registered and/or operating, whether as a result of:<br>1. Business closure<br>2. Acquisition by or merger of the issuing institution with another (or new) financial instrument<br>3. Contractual term or maturity reached (e.g., options or bonds )<br>4. Determination of illegitimacy [perhaps not required for historical entities] |
| NOT_SPECIFIED | The EFII record is in a state in which it does not provide information about whether the financial instrument is legally registered and operating. |

### 3.5.5 MarketSectorEnum Code List

Code signifying the market sector of the security.

| Code | Definition |
|---|---|
| COMMODITY | |
| EQUITY | |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| MUNICIPALS | |
| PREFERRED | |
| MONEY_MARKET | |
| GOVERNMENT | |
| CORPORATE | |
| INDEX | |
| CURRENCY | |
| MORTGAGE | |

### 3.5.6 RegistrationStatusEnum Code List

The `RegistrationStatusEnum` value in an EFII record indicates the status of the registration of the financial instrument with an DSU. This is not to be confused with the status of the EFII itself, which is specified by `FinancialInstrumentStatusEnum` (Section 3.5.4). See also Section **Error! Reference source not found.**, which illustrates how the `FinancialInstrumentStatusEnum` value changes over the lifecycle of an EFII registration.

A value of type `RegistrationStatusEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| PENDING | An application for an EFII that has been submitted and which is being processed and validated.<br>*NOTE: EFII registrations in the `PENDING` state are not intended for public release, but could be used internally between DSUs.* |
| ISSUED | An EFII Registration that has been validated and issued, and which identifies the financial instrument as of the last update. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| DUPLICATE | An EFII Registration that has been determined to be a duplicate registration of the same financial instrument as another EFII Registration; the DUPLICATE status is assigned to the non-surviving registration (i.e., the EFII that should no longer be used). Only one of the potential multiple identifiers will survive; for all other duplicate registrations:<br>1. The RegistrationStatus is set to DUPLICATE,<br>2. The EFII of the surviving EFII Registration is set in the SuccessorEFII data element of (each) duplicate EFII registration;<br>3. The LastUpdateDate is set to reflect the date of this update, and<br>4. No further updates of the DUPLICATE registration record will occur. |
| MERGED | An EFII registration for a financial instrument that has been merged into another financial instrument, such that this financial instrument no longer exists as an active financial instrument.<br>If<br>• After being issued an EFII, the financial instrument is merged with the financial instrument of another entity;<br>• Per agreements among the parties to the transaction, the EFII of the acquired or merged financial instrument will not be used to identify the surviving financial instrument (or if a new financial instrument is created that is issued a new EFII)<br>Then<br>1. The EFIIRegistrationStatus is set to "MERGED",<br>2. The EFIIRecordLastUpdate is set to reflect the date of this update, and<br>3. No further updates of the MERGED registration record(s) will occur. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| RETIRED | An EFII registration for a financial instrument that has ceased operation, without having been merged into another financial instrument.<br>If<br>• The responsible DSU determines by public sources that the financial instrument has ceased to exist (and the DSU seeks to confirm this status through all available channels)<br>Then<br>1. The EFIIRegistrationStatus is set to "RETIRED";<br>2. The EFIIRecordLastUpdate is set to reflect the date of this update;<br>3. The FinancialInstrumentExpirationDate is also set to the date of this update;<br>4. The FinancialInstrumentExpirationReason is set;<br>5. No further updates of the RETIRED registration record will occur. |

### 3.5.7 DataMaturityStageEnum Code List

A value of type DataMaturityStageEnum in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| UNVERIFIED | (e.g., "Raw") Data has been acquired without undergoing any EURHISFIRM data quality measures |
| VERIFIED | (e.g., "Collected") Source-level metadata attribution and data collection measures have been performed. This maturity stage is what the output of a Data Collection Unit (DCU) would achieve. |
| COMPLIANT | (e.g., "Harmonized") Source-level data elements have been harmonized to EURHISFIRM field-level Common Data Model conventions. In DDI terms: source variables have been mapped to conceptual (i.e., semantic) standards. This maturity stage is typically associated with the output of a Data Submission Unit (DSU) |
| CONSOLIDATED | (e.g., "Reconciled")Data previously collected and harmonized is compared with other EURHISFIRM data (whether previously published or submitted by other DSUs) in order to identify multiple references to the same unique object (legal entity, issued security, etc) and resolve duplicate references (identifiers) to the same object. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| CONSOLIDATED | (e.g., "Promoted"). After reconciliation (deduplication / unique identification ), data is promoted to the maturity stage of published Common Data Model EURHISFIRM data |
| EU REVISED | After being promoted, the data was revised or edited |

.

# 4 Change Management

- TBD. Use the <vnext> strategy for forward/backward compatibility, and/or include a version number in the XML. Needs to be explained

# 5 Examples (non-normative)

- TBD.

# 6 References

[ISO646] ISO, "Information technology -- ISO 7-bit coded character set for information interchange," ISO/IEC 646:1991.

[ISO3166-1] ISO, "Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes," ISO 3166-1:2013.

[ISO3166-2] ISO, "Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code," ISO 3166-2:2013.

[ISO8601] ISO, "Data elements and interchange formats – Information interchange – Representation of dates and times," ISO 8601:2004.

[ISO10646] ISO, "Information technology -- Universal Coded Character Set (UCS)," ISO 10646:2012.

[ISO17422] ISO, "Financial Services - Legal Entity Identifier (LEI)," ISO/DIS 17442:2012.

[ISODir2] ISO, "Rules for the structure and drafting of International Standards (ISO/IEC Directives, Part 2, 2001, 4th edition)," July 2002.

[XSD1] H. Thompson, D. Beech, M. Maloney, N. Mendelsohn, "XML Schema Part 1: Structures," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-1/.

[XSD2] P. Biron, A. Malhotra, "XML Schema Part 2: Datatypes," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-2/.

# 7 Appendix: Character Codes Allowed in Romanized Names

When a `Name` instance is of type `PREFERRED_ROMANIZED` or `AUTO_ROMANIZED`, the value of the `name` field SHALL consist only of non-control characters drawn from the "invariant



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

subset" of ISO 646. These characters are enumerated below. The "Hex Value" column indicates the code point value (expressed in hexadecimal) for each character in both ISO 646 and ISO 10646.

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| ! | Exclamation Mark | 21 | M | Capital Letter M | 4D |
| " | Quotation Mark | 22 | N | Capital Letter N | 4E |
| % | Percent Sign | 25 | O | Capital Letter O | 4F |
| & | Ampersand | 26 | P | Capital Letter P | 50 |
| ' | Apostrophe | 27 | Q | Capital Letter Q | 51 |
| ( | Left Parenthesis | 28 | R | Capital Letter R | 52 |
| ) | Right Parenthesis | 29 | S | Capital Letter S | 53 |
| * | Asterisk | 2A | T | Capital Letter T | 54 |
| + | Plus sign | 2B | U | Capital Letter U | 55 |
| , | Comma | 2C | V | Capital Letter V | 56 |
| - | Hyphen/ Minus | 2D | W | Capital Letter W | 57 |
| . | Full Stop | 2E | X | Capital Letter X | 58 |
| / | Solidus | 2F | Y | Capital Letter Y | 59 |
| 0 | Digit Zero | 30 | Z | Capital Letter Z | 5A |
| 1 | Digit One | 31 | _ | Low Line | 5F |
| 2 | Digit Two | 32 | a | Small Letter a | 61 |
| 3 | Digit Three | 33 | b | Small Letter b | 62 |
| 4 | Digit Four | 34 | c | Small Letter c | 63 |
| 5 | Digit Five | 35 | d | Small Letter d | 64 |
| 6 | Digit Six | 36 | e | Small Letter e | 65 |
| 7 | Digit Seven | 37 | f | Small Letter f | 66 |
| 8 | Digit Eight | 38 | g | Small Letter g | 67 |
| 9 | Digit Nine | 39 | h | Small Letter h | 68 |
| : | Colon | 3A | i | Small Letter i | 69 |
| ; | Semicolon | 3B | j | Small Letter j | 6A |
| < | Less-than Sign | 3C | k | Small Letter k | 6B |
| = | Equals Sign | 3D | l | Small Letter l | 6C |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| > | Greater-than Sign | 3E | m | Small Letter m | 6D |
| ? | Question Mark | 3F | n | Small Letter n | 6E |
| A | Capital Letter A | 41 | o | Small Letter o | 6F |
| B | Capital Letter B | 42 | p | Small Letter p | 70 |
| C | Capital Letter C | 43 | q | Small Letter q | 71 |
| D | Capital Letter D | 44 | r | Small Letter r | 72 |
| E | Capital Letter E | 45 | s | Small Letter s | 73 |
| F | Capital Letter F | 46 | t | Small Letter t | 74 |
| G | Capital Letter G | 47 | u | Small Letter u | 75 |
| H | Capital Letter H | 48 | v | Small Letter v | 76 |
| I | Capital Letter I | 49 | w | Small Letter w | 77 |
| J | Capital Letter J | 4A | x | Small Letter x | 78 |
| K | Capital Letter K | 4B | y | Small Letter y | 79 |
| L | Capital Letter L | 4C | z | Small Letter z | 7A |
| | Space | 20 | | | |

**8**



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 9.5 Legal Entity Data Artifact 1.05

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# EURHISFIRM Common Data Model Standard 1.0

## Legal Entity Data Artifact (LEDA)

Version of 20 November 2020

| Date | Revision | Description |
|---|---|---|
| 2020-07-23 | ver 1.01 | First Draft to be circulated |
| 2020-09-23 | ver 1.02 | Revisions to Header and Financial Data Elements |
| 2020-10-29 | ver 1.03 | Major revisions |
| 2020-11-01 | ver 1.04 | Additional formatting, revisions to EntityDataElement |
| 2020-11-18 | ver 1.05 | Numerous revisions |


## Abstract

First, the semantic content of these attributes must be fully specified. Second, some additional elements, such an indication of the status of the information, are necessary for effective use of the data. Third, the form the information takes at any given local point of source data capture must be such that it can be made to conform to a common standard, which must also be specified. This document proposes the standards necessary in these areas to support the EURHISFIRM Common Data Model.


## Status of this document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this document series will be maintained on SeaFile.

This draft is a **Working Draft** which can be circulated to any interested parties for review and comment. It is a draft document and may be updated, replaced or made obsolete by other documents at any time. It is inappropriate to use Working Drafts as reference material or to cite them as other than "work in progress." This is work in progress and does not imply endorsement by the EURHISFIRM ExCo.

Comments on this document should be sent to [TBD: insert mailing list or URL].



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## Table of Contents


*1 Introduction ..... 3*

*2 Terminology and Typographical Conventions ..... 4*

*3 Abstract Data Content ..... 4*

**3.1 LEDA EntityArtifact Record ..... 5**

3.1.1 EntityArtifactItems of the LEDA EntityArtifact ..... 6

3.1.2 Extension Section of LEDA EntityArtifact ..... 7

**3.2 Data Types ..... 8**

3.2.1 DateTime Data Type ..... 8

3.2.2 DCUID Data Type ..... 9

3.2.3 DSUID Data Type ..... 9

3.2.4 EntityArtifactCategory Data Type ..... 9

3.2.5 EntityArtifactItem Data Type ..... 9

3.2.6 EntityArtifactItemCategory Data Type ..... 10

3.2.7 EntityArtifactItemStatus Data Type ..... 10

3.2.8 EntityArtifactItemType Data Type ..... 10

3.2.9 ItemName Data Type ..... 11

3.2.10 LanguageCode Data Type ..... 11

3.2.11 Name Data Type ..... 11

3.2.12 OtherEntityArtifactItemName Data Type ..... 12

3.2.13 EntityArtifactSource Data Type ..... 13

3.2.14 EntityArtifactSourceCitation Data Type ..... 14

3.2.15 EntityArtifactSourceLocatorAnchorType Data Type ..... 15

3.2.16 EntityArtifactSourceLocatorKeyType Data Type ..... 15

**3.3 Enumerated Code Lists ..... 16**

3.3.1 EntityArtifactCategoryTypeEnum ..... 16

3.3.2 DataMaturityStageEnum Code List ..... 16

*4 Change Management ..... 17*

*5 Examples (non-normative) ..... 17*

*6 References ..... 17*

*7 Appendix: Character Codes Allowed in Romanized Names ..... 18*




This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

# 1 Introduction

The economic activity of businesses and firms is, and has been, expressed in financial statements that periodically report on the state and status of the business operations and financial "health" of organizations. Such financial statements and reports (e.g., balance sheets and income statements) have, for centuries, used classifications and categories that conform to the accepted accounting standards in the jurisdiction and historical era in which the reports have been prepared.

Although there has definitely been an evolution and ongoing revision of a number of primary accounting standards in the modern era (e.g., GAAP, IFRS, and "Local GAAP"), the fundamentals of balance sheet and income statement reporting have, for the most part, and at a high level, been reasonably stable and consistent over the historical time frame that is of interest to the scope of EURHISFIRM.

However, although balance sheets and income statements (and cash flow statements) have generally well-accepted concepts at the highest level of summary, or consolidation, of financial data, this high-level understanding of such basic concepts as balance sheet assets, liabilities and equity -- or income statement concepts of income, expenses, earnings (profit) -- is too general to be considered any kind of standard definition of the structure of financial statements. Even so, the structure of most financial statements reflects, in varying degrees and levels of detail, the basic underlying principles of financial accounting that are common to most enterprises.

At the highest level of summary, there is little if any information about the type of economic activities, lines of business, or the products and services that form the basis and rationale of a particular firm, especially when a firm has multiple lines of business or products that are offered to the marketplace. Such information is almost always contained in lower levels of the financial statements in which taxonomies and categories of classification and categorization of types of business activities (business lines, business units, products and services) are themselves summarized and itemized.

The standardization of detailed financial reporting taxonomies (such as those established in recent times by the eXtensible Business Reporting Language, or XBRL ) for the wide variety of companies in the industries and sectors of the economies of the different countries in Europe is outside the initial scope of the high-level standards for financial statements in the EURHISFIRM Common Data Model.

However, the economic data of businesses and firms contained in their historical financial records constitute a collection of the core elements of data in a historical financial Research Infrastructure such as EURHISFIRM, and as such is essential to the justification of the rationale to create the EURHISFIRM research infrastructure. Financial statement information (i.e., the labels associated with historical financial facts) should be semantically defined in EURHISFIRM, but – owing to the incomplete nature of many historical records -- it is not (initially) possible, nor necessarily advantageous, to standardize a normative structure of all line items and financial information that will or could be collected and recorded in the EURHISFIRM research infrastructure.

Hence, the further elaboration of the standardization of financial statement taxonomies will be an ongoing task that must follow the establishment of an initial framework and means to capture data contained in financial statements by first defining the types of data elements and facts that



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

are present in historical records that contain financial information that may or may not be found in the form of complete financial statements.

In addition to information contained in financial statements or other published financial data pertaining to legal entities, many other types of data (e.g., market prices, manufacturing output, macro economic data, etc. can also be captured in this format.

- .

## 2 Terminology and Typographical Conventions

Within this specification, the terms SHALL, SHALL NOT, SHOULD, SHOULD NOT, MAY, NEED NOT, CAN, and CANNOT are to be interpreted as specified in Annex G of the ISO/IEC Directives, Part 2, 2001, 4th edition [ISODir2]. When used in this way, these terms will always be shown in ALL CAPS; when these words appear in ordinary typeface they are intended to have their ordinary English meaning.

All sections of this document, with the exception of Section 1 are normative, except where explicitly noted as non-normative.

The following typographical conventions are used throughout the document:

- ALL CAPS type is used for the special terms from [ISODir2] enumerated above.
- `Monospace` type is used to denote programming language, UML, and XML identifiers, as well as for the text of XML documents.
- ➢ Placeholders for open issues and/or changes that need to be made to this document prior to its reaching the final stage of approved Proposed Standard are prefixed by a rightward-facing arrowhead, as this paragraph is.

The specifications of data types and elements include a column for "cardinality" (usually abbreviated 'CARD' ) that describes if the element is either mandatory or optional, and whether the number of occurrences of the element can be more than one, or just one.

## 3 Abstract Data Content

This section specifies the abstract data content of a data file conforming to this standard.

A data file conforming to this standard SHALL consist of:

- An optional LEDA File Header, as specified in Section 3.1.
- An LEDA Root Record as specified in Section **Error! Reference source not found.**.
- Zero or more LEDA EntityArtifacts, as specified in Section 3.3.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.1 LEDA EntityArtifact Record

The LEDA EntityArtifact Record establishes a context and container for the collection of a number of Artifact Items associated with a given legal entity identified by the ELEI, as of a particular historical date, and potentially covering a particular interval of time.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ELEI | ELEI | 1 | The ELEI of the organization |
| PLEI | PLEI | 0,1 | "Previous Identifier" of the organization, generally in the form of IdentifierDomain.Identifier, e.g. "SCOB.94342" |
| EntityArtifactItems | EntityArtifactItems | 1 | Artifact Items collected in this EntityArtifact |
| Extension | Extension | 0,1 | An optional element for including data beyond the standard artifact Items in a LEDA data file. This may include data specific to an DSU, data specific to a publisher of LEDA data, and so on. |
| EntityArtifactCreationDate | DateTime | 1 | Date/time this LEDA record was initially created in the system |
| EntityArtifactUpdateDate | DateTime | 1 | Date/time that this historical LEDA record was most recently updated in the system. |
| EntityArtifactSource | EntityArtifactSourceType | 1,n | A locator of the source of the data Artifact reference data in this EntityArtifact record (mechanism TBD, typically company yearbooks and exchange registrations) |
| EntityArtifactStatus | EntityArtifactStatusEnum | 1 | Status of the LEDA EntityArtifact. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityArtifactUpToDate | DateTime | 0,1 | Should it exist and be known, the historical date up until which time the LEDA data for this data Artifact is valid. (Not the same thing as when the data was changed in the system). If present, a subsequent historical record for this LEDA may exist with the revised information. |
| EntityArtifactAsOfDate | DateTime | 0,1 | The date that this version of the data is known to be valid. This date would typically be derived from either:<br>1. The publication date of the source reference, or<br>2. Information contained in the source reference |
| OriginalDCU | DCUID | 1 | The Identifier of the Data Collection Unit (DCU) that originally gathered or created the data in this EntityArtifact. |
| ResponsibleDSU | DSUID | 1 | The Identifier of the Data Submission Unit (DSU) that produced and manages this LEDA EntityArtifact. |
| DataMaturityStage | DataMaturityStageEnum | 0,1 | The current data maturity stage of this LEDA record, or omitted if the data maturity stage is not known. |

### 3.1.1 EntityArtifactItems of the LEDA EntityArtifact

The EntityArtifactItems section of an LEDA EntityArtifact in a file conforming to this standard SHALL include one or more entity Artifact Items as specified below.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityArtifactItem | EntityArtifactItemType | 1,n | One or more Artifact Items that are collected in this EntityArtifact |

### 3.1.2 Extension Section of LEDA EntityArtifact

The Extension section of an LEDA record may be used to include additional data not defined in this standard. For example, a DSU may use Extension to publish additional Artifact Items it collects as part of EntityArtifact.

- TBD: include the details of how this works. Basically, the idea is to use an XSD schema wildcard with namespace ##other, permitting the inclusion of XML elements from other XML namespaces.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 3.2 Data Types

This section specifies the data types referenced by the tables in Section **Error! Reference source not found.**, in alphabetical order.

### 3.2.1 DateTime Data Type

A value of type `DateTime` in a conforming to this standard SHALL be a point in time expressed as a string conforming to ISO 8601 having the following format:

- TBD: Provision for different calendars ? (i.e., Gregorian, Julian, etc. )

*YYYY*-*MM*-*DD*T*hh*:*mm*:*ss.sss*TZ

where the components of the above string are as follows:

- *YYYY* is the year
- *MM* is the month (01 = January, …, 12 = December)
- *DD* is the day of the month (01 = first day of the month)
- T is the single character 'T'
- *hh* is the hour (00 – 23)
- *mm* is the minute
- *ss.sss* is the second and milliseconds. From one to three digits may be used for milliseconds, or omitted entirely along with the decimal point.
- *TZ* is the time zone specifier, which can be either:
  - Z the single character 'Z', denoting Coordinated Universal Time (UTC); or
  - +*hh*:*mm* denoting a positive offset from UTC; or
  - -*hh*:*mm* denoting a negative offset from UTC

In the XML representation specified in Section **Error! Reference source not found.**, the XSD type `xs:dateTime` is used; however, whereas `xs:dateTime` permits the time zone specifier to be omitted, `DateTime` values in files conforming to this standard SHALL always include a time zone specifier.

*Explanation (non-normative): milliseconds are hardly necessary for LEDA reference data, and likewise it might seem simpler to allow only "Z" as a time zone specifier; however, XML processing tools support the full syntax given above and it is not always possible to restrict such tools to avoid milliseconds or force the use of "Z" as the time zone specifier. The restriction that*



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

*the time zone specifier must be present is equivalent to using XSD type* `xs:timestamp`*; however this was introduced in XSD 1.1 and not supported by the majority of XML processing tools which still only implement XSD 1.0.*

### 3.2.2 DCUID Data Type

A value of type `DCUID` in a file conforming to this standard SHALL be a [TBD]-character Data Collection Unit Identifier conforming to [TBD].

### 3.2.3 DSUID Data Type

A value of type `DSUID` in a file conforming to this standard SHALL be a [TBD]-character Data Submission Unit Identifier conforming to [TBD].

### 3.2.4 EntityArtifactCategory Data Type

A value of type `EntityArtifactCategory` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| `EntityArtifactCategory` | `EntityArtifactCategoryTypeEnum` | 0,1 | Indicates the general category of the type of data Artifact identified by this LEDA EntityArtifact |

### 3.2.5 EntityArtifactItem Data Type

An `EntityArtifactItem` in an LEDA EntityArtifact in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| | | | |
| `ArtifactItemConceptName` | `ItemName` | 1,1 | The Concept Name of the Artifact Item. |
| `ArtifactItemVariableName` | `ItemName` | 1,1 | The Variable Name of the Artifact Item. |
| `OtherArtifactItemNames` | `String` | 0..n | An optional list of other `Name` instances for the EntityArtifactItem, whether in different languages in the jurisdiction of the legal entity, or in different languages used in foreign jurisdictions. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| EntityArtifactItemCategory | EntityArtifactItemCategoryType | 0,1 | Indicates the domain category of the type of this EntityArtifactItem |
| EntityArtifactItemAsOfDate | Date | 0,1 | AsOf date of the EntityArtifactItem fact (assuming it is date-dependent) |
| EntityArtifactItemValue | String | 0,1 | Artifact Item value |
| EntityArtifactItemType | ElementDataType | 0,1 | Artifact Item type |
| EntityArtifactItemStatus | EntityArtifactItemStatusEnum | 1 | The status of the EntityArtifactItem. This is not to be confused with the status of the EntityArtifact; see EntityArtifactStatus. |

### 3.2.6 EntityArtifactItemCategory Data Type

An EntityArtifactItem in an LEDA EntityArtifact in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| | | | |
| ArtifactItemDomainName | Name | 1,1 | The Domain Name of the Data Concept. |
| ArtifactItemDefinition | URL | 1,1 | Link to semantic web definition of this element |

### 3.2.7 EntityArtifactItemStatus Data Type

An EntityArtifactItemValue in an LEDA EntityArtifact in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| ArtifactItemDataStatus | ElementDataStatus | 1,1 | Status of the ArtifactItemValue (TBD) |

### 3.2.8 EntityArtifactItemType Data Type

An EntityArtifactItemValue in an LEDA EntityArtifact in a file conforming to this standard SHALL include data elements as specified below.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| | | | |
| ArtifactItemDataType | ElementDataType | 1,1 | DataType of the ArtifactItemValue |
| ArtifactItemDefinition | URL | 1,1 | Link to semantic web definition of this element |

### 3.2.9 ItemName Data Type

A `ItemName` is a string expressed in a natural language, including a code indicating which natural language is used.

A value of type `ItemName` in a file conforming to this standard SHALL include data elements as specified below.

| Element Name | Type | Card | Description |
|---|---|---|---|
| lang | LanguageCode | 0,1 | The language of `name` |
| NameDomain | String | 1,1 | The Domain Name of the `ItemName` |
| Name | String | 1 | The name itself. |

### 3.2.10 LanguageCode Data Type

A value of type `LanguageCode` in a file conforming to this standard SHALL be a 2-character language code conforming to [ISO639-1]. Note that ISO 639-1 language codes are all lowercase.

- TBD: Alternatively, we could use IETF language tags (RFC 4646) instead. IETF language tags can distinguish between variations of the same language in different countries. For example, the IETF language tags `fr-CA` and `fr-FR` denote Canadian French and French as spoken in France, respectively; in ISO 639-1 these would both be simply `fr`.

### 3.2.11 Name Data Type

A `Name` is a string expressed in a natural language, including a code indicating which natural language is used.

A value of type `Name` in a file conforming to this standard SHALL include data elements as specified below.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Element Name | Type | Card | Description |
|---|---|---|---|
| lang | LanguageCode | 0,1 | The language of name |
| Name | String | 1 | The name itself. |

### 3.2.12 OtherEntityArtifactItemName Data Type

A value of type OtherEntityArtifactItemName in a file conforming to this standard SHALL include data elements as specified below. Each Name element includes an optional language code, permitting OtherEntityArtifactItemName to be repeated as many times as necessary to express the same name type in multiple languages. When type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, the language code specifies the language of the name prior to Romanization.

| Element Name | Type | Card | Description |
|---|---|---|---|
| type | EntityArtifactItemNameTypeEnum | 1 | The type of name represented by this OtherEntityArtifactItemName instance. The EntityArtifactItemNameType observes language, since 'Name' type has a language attribute. |
| Name | Name | 1 | The name. If type is PREFERRED_ROMANIZED_LEGAL or AUTO_ROMANIZED_LEGAL, then this value SHALL only include characters from the character set specified in Section 7. |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.2.13 EntityArtifactSource Data Type

| | | | |
|---|---|---|---|
| EntityArtifactSourceCitation | EntityArtifactSourceCitationType | 0,1 | A locator of the source of the EntityArtifactItemreference data in this EntityArtifact record (mechanism TBD, typically company yearbooks or exchange EntityArtifacts) |
| EntityArtifactSourceLocatorKey | EntityArtifactSourceLocatorKeyType | 0,1 | A locator of the source of the EntityArtifactItemreference data in this EntityArtifact record (mechanism TBD, typically company yearbooks) |
| EntityArtifactSourceLocatorAnchor | EntityArtifactSourceLocatorAnchorType | 0,1 | Additional information to allow pinpointing where in the EntityArtifactSource of the EntityArtifactItemdata can be found |
| EntityArtifactSourceDCU | DCUID | 0,1 | The identifier of the Data Collection Unit that sourced the EntityArtifact data |
| EntityArtifactSourceArtifactID | String | 0,1 | An identifier of the EntityArtifactItemthat may have been locally used or assigned by the DCU at collection time |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.2.14 EntityArtifactSourceCitation Data Type

An associative locator (link or key) that resolves to a DDI metadata block regarding the source of the EntityArtifactItemreference data in this EntityArtifact record (mechanism TBD, typically company yearbooks -)

DDI 3.2 uses the `CitationType` Element for referencing publications (e.g. yearbooks). It contains 11 elements which are described in detail in the *DDI Lifecycle XML Schema*. The table below lists the suggested cardinality for each element and some notes on special uses. The name of each element is a link to its description in the DDI XML Schema.

| Element Name | Type | Card | Description |
|---|---|---|---|
| Title | DDI Title | 1,1 | |
| SubTitle | DDI SubTitle | 0,1 | |
| AlternateTitle | DDI AlternateTitle | 0,n | |
| Creator | DDI Creator | 0,n | |
| Publisher | DDI Publisher | 0,1 | |
| Contributor | DDI Contributor | 0,n | |
| PublicationDate | DDI PublicationDate | 0,1 | For monographs: SimpleDate<br>For serials (e.g. yearbooks): StartDate and EndDate (the date/year when the first and the last volume of a serial were published – the specific volume and page numbers of the year from which information about the EntityArtifactItemwas taken can then be recorded in the EntityArtifactSourceLocatorAnchor element) |
| Language | DDI Language | 0,n | |
| InternationalIdentifier | DDI InternationalIdentifier | 0,n | |
| Copyright | DDI Copyright | 0,1 | |
| dc:isPartof | DDI dc:isPartof | 0,1 | Use for instance to record the title of the newspaper in which a stock exchange price list is published |

The Citation as we would want to use it in the EntityArtifactReference Data would be considered as a DataSource in DDI 3.2 (Origin element).



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

### 3.2.15 EntityArtifactSourceLocatorAnchorType Data Type

The `EntityArtifactSourceLocatorAnchorType` provides additional information to allow pinpointing where in the EntityArtifactSource referenced by the `EntityArtifactSourceLocatorKey` can the EntityArtifactdata be found

### 3.2.16 EntityArtifactSourceLocatorKeyType Data Type

An associative locator (link or key) that resolves to a DDI metadata block regarding the source of the EntityArtifact reference data in this EntityArtifact record (mechanism TBD, typically company yearbooks -)



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

## 3.3 Enumerated Code Lists

This section specifies the enumerated code list data types (all having the suffix `Enum`) referenced by the tables in Sections **Error! Reference source not found.** and 3.1, in alphabetical order.

### 3.3.1 EntityArtifactCategoryTypeEnum

The `EntityArtifactCategoryTypeEnum` value in an LEDA record specifies the general classification category of the EntityArtifact.

- TBD: the table below is just an example with some possible categories, and is incomplete, and needs to be elaborated.

A value of type `EntityArtifactCategoryTypeEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| | To Be Determined |

### 3.3.2 DataMaturityStageEnum Code List

A value of type `DataMaturityStageEnum` in a file conforming to this standard SHALL be one of the code strings specified in the following table:

| Code | Definition |
|---|---|
| `UNVERIFIED` | (e.g., "Raw") Data has been acquired without undergoing any EURHISFIRM data quality measures |
| `VERIFIED` | (e.g., "Collected") Source-level metadata attribution and data collection measures have been performed. This maturity stage is what the output of a Data Collection Unit (DCU) would achieve. |
| `COMPLIANT` | (e.g., "Harmonized") Source-level data elements have been harmonized to EURHISFIRM field-level Common Data Model conventions. In DDI terms: source variables have been mapped to conceptual (i.e., semantic) standards. This maturity stage is typically associated with the output of a Data Submission Unit (DSU) |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Code | Definition |
|---|---|
| CONSOLIDATED | (e.g., "Reconciled")Data previously collected and harmonized is compared with other EURHISFIRM data (whether previously published or submitted by other DSUs) in order to identify multiple references to the same unique object (legal entity, issued security, etc) and resolve duplicate references (identifiers) to the same object. |
| CONSOLIDATED | (e.g., "Promoted"). After reconciliation (deduplication / unique identification ), data is promoted to the maturity stage of published Common Data Model EURHISFIRM data |
| EU REVISED | After being promoted, the data was revised or edited |

.

## 4 Change Management

- TBD. Use the <vnext> strategy for forward/backward compatibility, and/or include a version number in the XML. Needs to be explained

## 5 Examples (non-normative)

- TBD.

## 6 References

[ISO646] ISO, "Information technology -- ISO 7-bit coded character set for information interchange," ISO/IEC 646:1991.

[ISO3166-1] ISO, "Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes," ISO 3166-1:2013.

[ISO3166-2] ISO, "Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code," ISO 3166-2:2013.

[ISO8601] ISO, "Data elements and interchange formats – Information interchange – Representation of dates and times," ISO 8601:2004.

[ISO10646] ISO, "Information technology -- Universal Coded Character Set (UCS)," ISO 10646:2012.

[ISO17422] ISO, "Financial Services - Legal Entity Identifier (LEI)," ISO/DIS 17442:2012.

[ISODir2] ISO, "Rules for the structure and drafting of International Standards (ISO/IEC Directives, Part 2, 2001, 4th edition)," July 2002.

[XSD1] H. Thompson, D. Beech, M. Maloney, N. Mendelsohn, "XML Schema Part 1: Structures," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-1/.



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

[XSD2] P. Biron, A. Malhotra, "XML Schema Part 2: Datatypes," W3C Recommendation, May 2001, http://www.w3.org/TR/xmlschema-2/.

# 7 Appendix: Character Codes Allowed in Romanized Names

When a `Name` instance is of type `PREFERRED_ROMANIZED` or `AUTO_ROMANIZED`, the value of the `name` field SHALL consist only of non-control characters drawn from the "invariant subset" of ISO 646. These characters are enumerated below. The "Hex Value" column indicates the code point value (expressed in hexadecimal) for each character in both ISO 646 and ISO 10646.

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| ! | Exclamation Mark | 21 | M | Capital Letter M | 4D |
| " | Quotation Mark | 22 | N | Capital Letter N | 4E |
| % | Percent Sign | 25 | O | Capital Letter O | 4F |
| & | Ampersand | 26 | P | Capital Letter P | 50 |
| ' | Apostrophe | 27 | Q | Capital Letter Q | 51 |
| ( | Left Parenthesis | 28 | R | Capital Letter R | 52 |
| ) | Right Parenthesis | 29 | S | Capital Letter S | 53 |
| * | Asterisk | 2A | T | Capital Letter T | 54 |
| + | Plus sign | 2B | U | Capital Letter U | 55 |
| , | Comma | 2C | V | Capital Letter V | 56 |
| - | Hyphen/ Minus | 2D | W | Capital Letter W | 57 |
| . | Full Stop | 2E | X | Capital Letter X | 58 |
| / | Solidus | 2F | Y | Capital Letter Y | 59 |
| 0 | Digit Zero | 30 | Z | Capital Letter Z | 5A |
| 1 | Digit One | 31 | _ | Low Line | 5F |
| 2 | Digit Two | 32 | a | Small Letter a | 61 |
| 3 | Digit Three | 33 | b | Small Letter b | 62 |
| 4 | Digit Four | 34 | c | Small Letter c | 63 |
| 5 | Digit Five | 35 | d | Small Letter d | 64 |
| 6 | Digit Six | 36 | e | Small Letter e | 65 |
| 7 | Digit Seven | 37 | f | Small Letter f | 66 |
| 8 | Digit Eight | 38 | g | Small Letter g | 67 |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

| Graphic Symbol | Name | Hex Value | Graphic Symbol | Name | Hex Value |
|---|---|---|---|---|---|
| 9 | Digit Nine | 39 | h | Small Letter h | 68 |
| : | Colon | 3A | i | Small Letter i | 69 |
| ; | Semicolon | 3B | j | Small Letter j | 6A |
| < | Less-than Sign | 3C | k | Small Letter k | 6B |
| = | Equals Sign | 3D | l | Small Letter l | 6C |
| > | Greater-than Sign | 3E | m | Small Letter m | 6D |
| ? | Question Mark | 3F | n | Small Letter n | 6E |
| A | Capital Letter A | 41 | o | Small Letter o | 6F |
| B | Capital Letter B | 42 | p | Small Letter p | 70 |
| C | Capital Letter C | 43 | q | Small Letter q | 71 |
| D | Capital Letter D | 44 | r | Small Letter r | 72 |
| E | Capital Letter E | 45 | s | Small Letter s | 73 |
| F | Capital Letter F | 46 | t | Small Letter t | 74 |
| G | Capital Letter G | 47 | u | Small Letter u | 75 |
| H | Capital Letter H | 48 | v | Small Letter v | 76 |
| I | Capital Letter I | 49 | w | Small Letter w | 77 |
| J | Capital Letter J | 4A | x | Small Letter x | 78 |
| K | Capital Letter K | 4B | y | Small Letter y | 79 |
| L | Capital Letter L | 4C | z | Small Letter z | 7A |
| | Space | 20 | | | |



This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489