Design Principles supporting the Creation of Tools for the Interoperability of European Digital Data Spaces and Marketplaces

i3-MARKET Consortium © 2022

Contents:

i3-MARKET

www.i3-market.eu

Supported by:

In collaboration with:

INTERNATIONAL DATA SPACES ASSOCIATION

BDV BIG DATA VALUE ASSOCIATION

GAIA-X data-infrastructure.eu

EUROPEAN COMMISSION

Achille Zappa and Martin Serrano

NUI Galway, Data Science Institute - Insight SFI Research Centre for Data Analytics

Galway, Ireland

achille.zappa@insight-centre.org   martin.serrano@insight-centre.org

1 Introduction

In today's Data Driven Society, there is a growing demand for a single European Data Market Economy [1] capable to preserve data security and privacy, enable data sharing and data transactions across data spaces/marketplaces, allowing the cross-domain exchange and support business innovation. There is a need for providing technologies for trustworthy (secure and reliable), data-driven collaboration and federation of existing and new future data spaces and data marketplace platforms. In this data-driven view special attention is given to industrial data, particularly sensitive commercial data assets from both SMEs and large industrial corporations, this attention is mainly derivate by the amount of data that is possible to be generated.

It is well known that despite various research and innovation attempts working on Big Data management and sharing, there is not a broadly accepted trusted and secure solution for federation of data marketplaces. The i3-MARKET consortium addressed this gap by developing the i3-MARKET backplane as a set of software technologies and backend solutions for a trusted (secure, self-governing, consensus-based, and auditable), interoperable (semantic-driven) and decentralized (scalable) infrastructure. The i3-MARKET Backplane is delivered as a set of software components that are aimed at facilitating access and competitiveness for SMEs to European data marketplaces, by allowing them to view, buy and sell data from/to multiple marketplaces, through one interface.

2 Practical Considerations

2.1 Vocabularies and Terminologies

Considering the need for data interoperability [2] the first important aspects to work on are:

a) The definition, creation and collection of data models that allow to share a common description of the data assets (as per the case of dynamic data offering descriptions, operations, services, data details, credentials, contracts, pricing, actors; etc.,

b) The development and implementation of data management systems and storage to manage data and information and meta data descriptions.

c) share the data models with the community to make use of those and work with people in improving and maintain the models for the present and future. d) The creation of data offering descriptions, management of controlled registries, mapping of information, interfaces among other components, and e) provide the links of data and actors for discovering and retrieve of necessary information, compiling of smart contracts and other operations.

2.2 Taxonomies

The use of standardized vocabularies is important to avoid redundancies or unnecessary data modeling work. Schemas and created data models and the interaction with templates and schemas is important mainly for enabling interoperability between nodes, user-friendly services, exchangeability of Data Assets, representation of actors (Marketplaces, Providers, Consumers, Owners) and data exchange between different instances in the Infrastructure Ecosystem. A variety of standard vocabularies already exist for main topics and sub-specific topics and domains, the most suitable ones to set up a common information model are selected and integrated into a high-level collection of vocabularies taxonomies and organized information like graphs.

2.3 Data Modeling

A high level approach called logic model that contains and represent all the data relationships needs to be created, in this line i3-MARKET model has included the necessary domains according to vocabularies and taxonomies and defined the missing parts for the main operational interactions & links among entities. The i3-MARKET Backplane uses a basic principle where the Information is part of an Ecosystem and thus the Infrastructure have to be able to combine and seamless exchange information and operations in a distributed data exchange architecture. The use of a meta-modelling perspective is always helpful to add a level of abstraction and to ensure the data. the i3-MARKET has raised certain requirements

that go beyond the simple main description of data sets, adding information models to define other entities, operations actors, sharing agreements, data details and data-related descriptions about offerings.

2.4 i3-Market Semantic Model

A semantic model is introduced to facilitate the formal semantic descriptions of concepts and proper-ties used in the context of the i3-MARKET Backplane and API. The existing data models covers only partially the requirements for the data management aspects but in i3-MARKET the backplane scopes linking and extend common vocabularies created the i3-MARKET Semantic Core Model, including pricing model, contractual model for data sharing agreements and service agreements alike for using smart contracts and to compile a collection of semantic information models.

i3-MARKET introduced the concept of "data offering" that allows to describe the capabilities and interfaces of an i3-MARKET data asset from a provider perspective and how this information can be used for discovery and access purpose but at the same time provides information towards the commercial use of the data asset. The specific description is called the Data Offering description and it is used in particular by the implemented component called the Semantic Engine and Framework to manipulate and manage the metadata that is used by other components.

To simplify both semantic model development and reuse, a modular design is beneficial. Based on the project specification and the requirements the i3-MARKET semantic model practical implementation addresses the collection, access, storage, use and selling of the data. The semantic model is called OCASUS and can be modularized according to specific scope which follows i3-MAKET data model.

2.5 OCASUS Data Lifecycle Principle

The i3-MARKET project envisages semantic services with enhanced discovery tasks, to satisfy this need a particular component called SEED is implemented to provide the basic query-based mechanisms on top of the metadata offerings, including complex discovery and retrieve checks that make sure, e.g., that the necessary information is retrieved by the actors and services. The first prototype of the Semantic Engine and Framework solution SEED is available and integrated into the i3-MARKET backplane. An additional important aspect is also the functionalities related to the creation and registration of the data offering descriptions and the management of local and federated Registries. Another implemented functionality is the metadata semantic Registry, with this feature, the i3-MARKET backplane can rely on scaled up metadata registry storage capacity to collect the semantic information that can be queried. One of the key aspects when designing a semantic model is reusing of knowledge. Once a semantic model is created for a domain, it should be (at least to some degree) reusable for other applications in the same domain. in i3-MARKET the creation and use of the O-CASUS model to cover the modeling and semantic needs lead to concept of O-CASUS library, which is an idea based on the Data Lifecycle Process:

• Compile Vocabularies and taxonomies in relation to Marketplaces Metadata, Operation and Management.

• Formalize the state of current marketplaces by using best practices and standards;

• Compile a collection of Vocabularies and Ontology for Collecting, Accessing, Store, Utilize and Sell Data.

OCASUS is the unique Link to a pricing module and the other standard vocabularies and taxonomies to cover the various parts of the data lifecycle as it is described and contained in i3-MARKET data models. The main design contributions in this position paper are based on the Semantic Data Modelling process as it is the result of practical implementations that serves as consolidation of the i3-MARKET data lifecycle principles and also the use of Semantic Models. The integration and extensions of other common Sematic Models to enable the mapping of the metadata and description of the data assets, contracts and operations, provided from i3-MARKET stakeholders, also plays an important role in the design process. The i3-MARKET semantic model captures the structural and semantic characteristics (metadata) of the various entities in each data asset offering in the Markets as data offering. More specifically, the core model uses this models for:

1) data Registration of metadata descriptions which corresponds to the da-ta harmonization process. In this way, each provided data asset will be registered in our Registry with concepts from the i3-MARKET data offering model in a semi-automatic way;

2) metadata linking where any provided data asset metadata will be linked with other relevant sources (or data assets) that exist in the Backplane;

3) data Discovery (for local or federated Registries) which involves the development of algorithms and software for supporting the selection of the most appropriate metadata that best match user preferences.

4) Management of information related to Smart Contract, Data access and transfer, pricing models, identity and credentials identifications, notifications. The i3-MARKET Models are used for capturing the structural and semantic metadata characteristics of the various entities involved in the i3-MARKET back-plane domain, whereas the underlying conceptual models facilitate the use of lightweight reasoning during the discovery and operational process e.g., for contracts and service/agreements, data exchange, data access/transfer and data selling operations.

The O-CASUS Semantic Model comprise a collection of ontologies and vocabularies to cover the concepts used in the i3-MARKET Backplane and this is the first demonstration the semantic models works. The main part is the semantic vocabulary and model in the project as i3-MARKET Semantic Core Model. The core used to collect metadata terminologies related to e.g.: system actors (e.g. links and identifiers for Providers, Markets-nodes, Consumers…) In addition, the model, imported and added definitions from other recognized vocabularies and ontologies, [3] [4] as follow:

- W3C Data Catalog (DCAT and DCAT-AP)
- W3C Verifiable Credentials and DID
- SKOS Knowledge Organization System
- SKOS Data Offering Categories
- IT Service Ontology
- ADMS license type vocabulary
- IANA Media Types
- ISO-4217 Currency Terminology
- The Ontology of units of Measure (OM) 2.0
- W3C Data Quality Vocabulary
- Terminologies for time intervals frequencies
- Domain Annotations

3 Best Practices

3.1 Data Marketplace/Spaces Actors

To simplify both semantic model development and reuse, an interoperable design the semantic models can be modularized according to their scope. Based on the data spaces and marketplaces analysis and the i3-MARKET project experience and produced specification the proposed domains are as follow:
- Organization/Management module
- Market Module
- Provider module
- Consumer module
- Owner Model
- Query module
- Data Offering module

- Contractual Parameters Module

The i3-MARKET enables the different stakeholders to participate in the data lifecycle, providers can offer or trade access to datasets via the backplane. Data-Offerings are used by data processors, which describes via metadata a set of Resources offered via the i3-MARKET marketplaces in order to enable data consumers to access the metadata descriptions and eventually access to the data assets. It typically encompasses a set of related Information.

3.2 Data Offerings

A central role plays the concept of "Data Offering", which represents a set of data assets resources contained in by a data marketplace or data space. The defined concepts and properties are utilized to uniquely identify and semantically annotate meaningful relationships and contexts of data assets as well as to specify semantically defined values for relevant operations. The needed vocabulary management, that is operated via Semantic Engine components, includes features such as the registration, discovery, retrieval, update, and removal of semantic descriptions, as well as the alignment between other semantic descriptions.

A Data-Offering provides a semantic description of the data assets provided to a Consumer once the Data-Offering is registered. The description also entails context and meta information about the Distribution, including information to, the Pricing for accessing the Resource(s), the License of the Information provided, Contractual Parameters and service description as URL for data access).

A provider registers its offerings on the marketplace by providing an offering description. It contains the information about the Data Assets, data service, categories of data assets, sub classes components of catalogues and resources, data services, categories of the offering (core:category). All relevant communication metadata are provided on how the offering can be accessed through the data service and service extension descriptions. As illustrated in Figure 1, the Data Offering module represents the initial conceptualization which is built around the DataOffering and its metadata.

3.3 Data Assets

The W3C Data Catalog Vocabulary (DCAT) -Version 3 (https://www.w3.org/TR/vocab-dcat-3/) related to parts as: Dataset, Distribution, DataService used in data offering description enables a publisher to
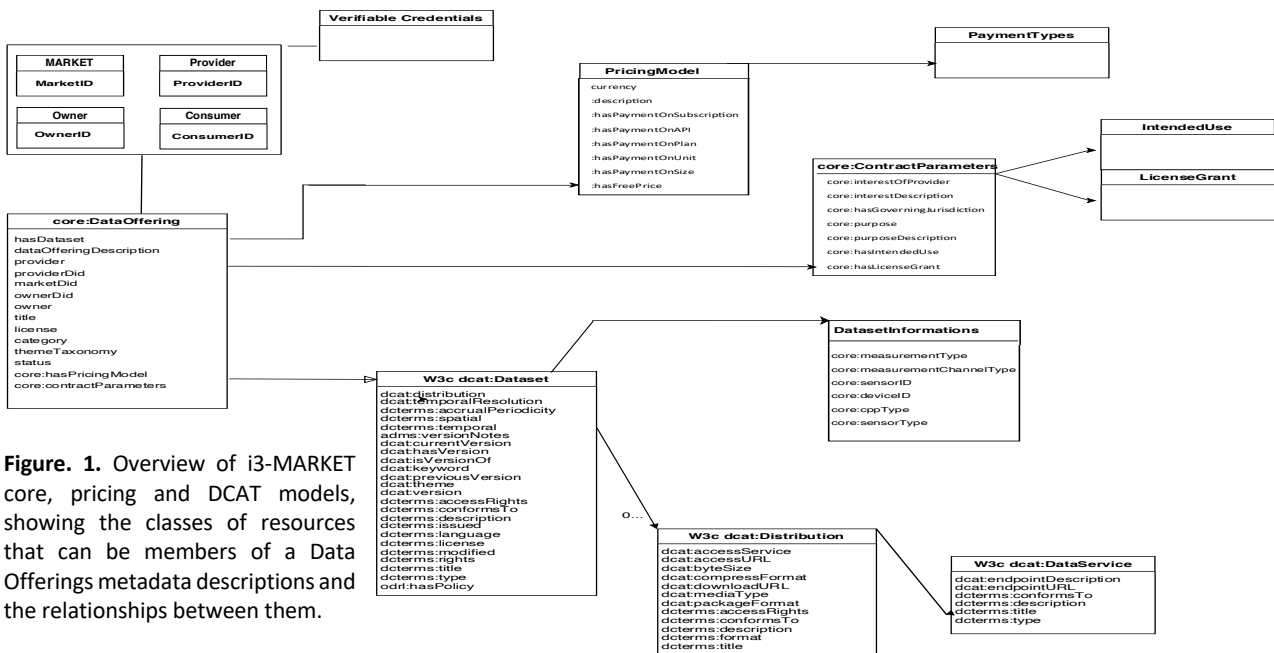
Figure. 1. Overview of i3-MARKET core, pricing and DCAT models, showing the classes of resources that can be members of a Data Offerings metadata descriptions and the relationships between them.

describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs. The data asset definition used in i3-MARKET is extended from the specific annotations to add extra information related to details that describe data contained in data sets. This information is used to give the possibility to providers to describe with more granularity the source and types of data that are shared/traded and annotations related to specific domains.

3.4 Smart Contracts

The smart contract parameters define the collection of the necessary information about the intended use and scope of the Data Offering/Dataset that will be used in the commercial transaction.

The Information derived from the i3-MARKET vocabularies Data Offering and from the details of the data sharing contract agreement that the Smart Contract Manager uses to compile the agreement between the parties (e.g., providers, consumers) is the binding legal relationship that need to be established in the smart contract.

The terminologies are used to indicate in the contract agreements information like:

• Data Offering descriptions
• Contract Parameters and intended use and grants
• Parties involved
• Validity and Duration period
• Duties Obligations
• Restrictions
• Pricing and Costs

4 Sharing Experiences

4.1 Project Repository – Open Access

The i3-MARKET project results including the different components implementation are shared not only with Consortium Partners but also with stakeholders and developers' communities. As part of open-source assets, the data models, documentations and files used in the i3-MARKET project are made available in both platforms, i.e. GitHub and GitLab:

• The i3-MARKET data Pack is the set of files, schemas and metadata model diagrams that represent the way the i3-MARKET semantics is organized and structured.
• The i3-MARKET Semantic Model info is the documentation that describes in detail all the taxonomies and vocabularies from needed domains used in i3-MARKET metadata descriptions.
• The Support and management process is the mechanism for how the data model is maintained following the interoperability requirements in i3-MARKET, if people want to contribute or have any suggestion for improving the semantic models.
• The Models Files shared in repository with releases versions where each section contains the online machine-readable files and other format for online accessibility, the files are maintained and updated regularly to keep the latest version of the models files up to date.

4.2 Data Space Compatibility

The i3-MARKET project is describing a Data Space in a formal way using the current standards from DCAT. The dcat:theme is used to give annotation and provide information about the domain categorization of the datasets. The i3-MARKET uses the themes as sub-categories to give more granularity in defining the domain annotations and in this way to create the classification and at the same time be able to allocate differences between multiple data spaces. In DCAT the domain of dcat:Theme was dcat:Dataset, which limited use of this property in other contexts. The domain has been relaxed in later revisions. In i3-MARKET was also added an upper level property for a Data Offering to annotate directly the High Level type of Category the data offering belongs to as core :category.

4.3 Data Protection and DID's

Credentials are a part of our daily lives; e.g. driver's licenses are used to assert that we are capable of operating a motor vehicle, university certificates can be used to assert our level of education, and government-issued passports enable us to travel between countries. These credentials provide benefits to us when used in the physical world, but their use on the Web continues to be elusive.

The verifiable credentials described in W3C Verifiable Credentials Data Model 1.0 is used in i3-MARKET. The difficulty of expressing digital credentials on the Web makes it challenging to receive the same benefits through the Web that physical credentials provide us in the physical world. This specification provides a standard way to express credentials on the Web in a way that is cryptographically secure, privacy respecting, and machine-verifiable.

The i3-MARKET project implemented the SSI& IAM Subsystems use DIDs that follow the W3C Decentralized Identifiers (DIDs) v1.0 specifications. Decentralized identifiers (DIDs) are a new type of identifier that enables verifiable, decentralized digital identity. A DID refers to any subject (e.g., a person, organization, thing, data model, abstract entity, etc.) as determined by the controller of the DID. In contrast to typical, federated identifiers, DIDs have been designed so that they may be decoupled from centralized registries, identity providers, and certificate authorities. Specifically, while other parties might be used to help enable the discovery of information related to a DID, the design enables the controller of a DID to prove control over it without requiring permission from any other party. DIDs are URIs that associate a DID subject with a DID document allowing trustable interactions associated with that subject.

5 Conclusions/Future Work

This position paper refers to best design practices and the presented results of the work related to the specification of data lifecycle and the use of Semantic Models to enable Data Interoperability and describe data assets within the i3-MARKET project. The i3-MARKET project describes data assets and resources semantically to enable metadata management in a i3-MARKET ecosystem. It defines the core models and specifications of the i3-MARKET Backplane environment. In particular, it is described the latest version of the i3-MARKET Semantic Models, revolving around the concept of Data Offering and the process in how data offerings are related to data assets. It captures most of the semantics associated to data offerings defined during specification of i3-MARKET's high-level architecture.

The implemented i3-MARKET solution allows that Data Offerings can be generated and processed by the i3-MARKET backplane implemented tools and particularly emphasis is set in the Semantic Engine with Enhanced Discovery Capabilities (SEED). The i3-MARKET uses specific sub-vocabularies for completing data spaces- and marketplaces-rich semantic models, opening the opportunity of defining synergies and compatibilities with other data spaces and marketplaces following the same semantic modeling path and by providing the experiences and the design principles.

6 References

1. Hans van der Veer and Anthony Wiles, "Achieving Technical Interoperability - the ETSI Approach," in ETSI, 2008.
2. "IoT Semantic Interoperability: Research Challenges, Best Practices, Recommendations and Next Steps," 2015.
3. Internet Assigned Numbers Authority (IANA). Media Types http://www.iana.org/assignments/media-types/
4. European Commission. Joinup. Asset Description Metadata Schema (ADMS). https://joinup.ec.europa.eu/solution/asset-description-metadata-schema-adms