Meeting Minutes: Adoption requirements workshop for the registry for semantic assets (SEMIC - A07.01)

Project:	SEMIC	**Date and Time:**	22/03/2024 13:00 - 15:00
Meeting Type:	Webinar	**Location:**	Virtual
Coordinators:	Claudio Baldassarre Nathan Ghesquière	**Issue Date:**	04/04/2024

Agenda of the webinar		
13:00 - 13:10	Introduction	Slides 1 - 3
13:10 - 13:25	Recap kick-off meeting	Slides 4 - 9
13:25 - 13:35	User categories	Slides 10 - 14
13:35 - 14:45	Use cases	Slides 15 - 22
14:45 - 15:00	Summary and next steps	Slides 23 - 28

Meeting Slides
LINK

Participants		
Name	**Initials**	**Organisation**
Ana Rosa Guzman	ARG	Ministerio de Política Territorial y Función Pública

Claudia Cristina Pollina	CCP	Dipartimento per la Trasformazione Digitale
Claudio Baldassarre	CB	DIGIT
Darius Amilevicius	DA	Informacinės visuomenės plėtros komitetas
Emidio Stani	ES	PwC EU Services
Felicitas Löffler	FL	Thüringer Finanzministerium
Florian Barthelemy	FB	PwC EU Services
Javier Andrés Arroyo	JAA	Ministerio de Política Territorial y Función Pública
Joren Verleyen	JV	PwC EU Services
Martynas Mockus	MM	Informacinės visuomenės plėtros komitetas
Matteo Fortini	MF	Dipartimento per la Trasformazione Digitale
Nathan Ghesquière	NG	PwC EU Services
Pascal Derycke	PD	Sciensano
Riitta Alkula	RA	Digi- ja väestötietovirasto
Sander Van Dooren	SVD	Flemish Government
Torbjörn Ull	TU	Bolagsverket

Full Meeting Minutes

Introduction Slides 1 - 3 **Speaker:** Claudio Baldassarre	Claudio opened the meeting and went over the agenda for the adoption requirements workshop: • Introduction • Recap kick-off meeting • User categories • Use cases • Summary & next steps
Recap kick-off meeting Slides 4 - 9 **Speaker:** Claudio Baldassarre	<u>Recap</u> Claudio started the meeting by going over a recap of the kick-off meeting and mentioned the positive feedback that came from that meeting. He also emphasized that this working group is the paveway for the future pilot that will be constructed in the next phase of the project. For this, it is implied that participants of this working group would be cooperating with the pilot as well. When a Member State is willing to cooperate with this pilot, it is expected from them to have

	the needed maturity level and/or have an overview of features that are needed for those Member States to be willing to adapt and use the registry. This way, these Member States could give valuable input with which the pilot could advance. <u>Registry vs. repositories</u> Next, Claudio went over the differences between the registry and national repositories. Here, he also mentioned what a registry and repository are exactly and apart from the differences, also what the similarities are between them. <u>Semantic operability at source</u> Thirdly, the semantic interoperability at source was explained. The registry is a spectrum to achieve interoperability. In order to do so, 3 possible ways are highlighted: • Downstream: each Member State maps their own models based on others found on the registry. • At source: This represents a process that could be done autonomously. Inspiration could be drawn from existing models. • Mediated: a mix between the previous 2. With this registry, all 3 possibilities would be covered. This way, the Member States have the opportunity to choose how they would interact with the registry. <u>Outcome and benefits</u> Regarding the outcome and benefits of the registry, Claudio continued by going over how this registry would be advantageous for the Member States when it would be adopted. The 3 main benefits of this registry would be: • Discoverability and findability • Co-designing • Harmonization <u>Questions and remarks</u> After this, Claudio asked the workgroup a question: ‘In the workflow of designing your models, where do you see the registry fitting in and for which output?’. Nathan added that if something was unclear in the previous topics, participants should not hesitate to ask questions about them. Ana Rosa: It is clear that the registry and the models it would have has its use. How could we handle the diversity with which every semantic model is documented? As Europeans, language barriers will always be a thing. Added to this, when connected to a national repository, every Member State has their own way of modelling and interacting with the models. It is not only about discovery, but also about accessibility.

- Claudio answered that the registry would not just simply be a redirection where the models would be exposed. The registry would have an experience that could be compared with the LOV repository. The registry would have its own interface and set of features. This will be discussed later. The registry would reprocess your input and the result would be given in the registry where you could see in which repository it is stored based on the metadata. What first needs to be established is whether each Member State is prepared to make a minimal effort to perform on their repository to be able for it to interact with the registry.
- Ana Rosa said that the current dimensional loss provides not only metadata on the models but also on each information element. Many models include properties/classes, and they sometimes pick up different aspects from other models. The question here should be more related to interacting with the national repositories because once you have found out in which repository the model you are searching for is stored, you will need to be able to interact with that repository.
- Riitta joined Ana Rosa saying that this kind of registry would not be helpful. The future is to source vocabularies when they are available and discoverable, so we could have genuine linked data. It is useful to find relevant sources on the registry, but it would only be dummy information from the repository where the model comes from. It would be beneficial to have a certain class linked to their source so different models that utilize that class could be seen as interoperable. Another beneficial aspect would be domain-specific information. It would be more relevant to see what has been done within a specific domain rather than what e.g. Italy has done. A common agreement on how data models would be described and the resources available would be needed.

Requirement
The registry pilot should enable domain specific filtering of models when looking for a semantic model.

- Felicitas mentioned it is important that the registry would be used to find models but also to look at others before creating a new model. Standards are also needed to consolidate and to be able to recommend what research already has been done. An example could be Life Sciences, which are organized with principles, guidelines and standards to work collaboratively on domain-specific semantic models. They use https://ontoportal.org/ as their registry software
- Torbjörn stated that apart from looking only on the business domain, he and Nordic colleagues also took an English definition that works all over the Nordic countries to provide, via API, an

	interconnection. This way, interoperability is achieved when a request is made. Doing so makes it available for external parties as well. Parallel to this registry, showing what the European Union has done and if something similar exists, could help. The Core Vocabularies from the European Union do not cover everything but should be used in adapting models. Torbjörn also agreed with Riitta stating that being able to filter on specific domains would be very beneficial, also for pushing domain-specific work under the umbrella of the European Union. Is this something that could be done?
	
	Requirement
	
	The semantic modelling done on the European Commission level should be findable on the registry.
	
	• Claudio answered that another group within the SEMIC-team is advancing on the Core Vocabularies. He added that this question and related information will be passed on to them. To recap this discussion, The current registry would have a 2-part problem:
	  • There should be easier access to the national repositories, as the models are stored there.
	  • It should be more domain specific.
	
	Another issue is that the core vocabularies are not publishable. Semantic interoperability would be achieved between Member States when their models would be aligned. Claudio also reiterated that Member States would have to take some actions themselves to be able to become interoperable with the registry.
	• Riitta commented that when core vocabularies are published as files, each repository has to implement them locally. Doing so means redundancy. In Riitta's opinion, national data models could be linked to the core models that the European Union has published as resolvable and negotiable data models. In addition to national models, domain specific (EU-wide) models like Torbjörn described (business domain modelling) would also be beneficial.
	• Claudio said he agrees, but still had some questions regarding this. How will someone know what models and what properties this would be? The registry would be made to promote discoverability of semantic models.
	• Sander agreed, but also said that it would be a pity because searching on the registry would generate a list of models. What would be maintained on the registry would go broader than only models/properties related to the core vocabularies. Even without the models from the European Commission, this would typically not be the kind of data models that would be found under the SEMIC umbrella. Another thing that could be helpful is the ability to compare a model with its past self and see the changes that

	were made. If this is a segmentation that would be made, from a design goal standpoint, it should be made as quickly as possible within a DCAT-AP space. **Requirement** The ability to compare a model with its previous versions and thus having an overview of the changes implemented with each iteration of the model. • Ana Rosa Stated that the models from the European Union are not only core vocabularies. These semantic assets do not cover everything. The Member States who are semantic leaders are very valuable to have as a link to the registry because they could help other Member States and build common semantic assets at the European level. • Claudio stated that we need those models at the very least. The semantic models in the European Union are exposed and discoverable on the registry as a reference pool. This goes beyond what the core vocabularies do, but they must expose different artifacts. The models exposed on the registry must be open data oriented and must have the ability to potentially have their model history with them. The ability to filter models (e.g. domain specific) goes hand in hand with the metadata of the model and should be standardized. Together with the Member States, a designing process for this standardization should be thought out.
User categories Slides 10 - 14 **Speaker:** Claudio Baldassarre	<u>Categories of users</u> After the questions and answers, Claudio continued by going over the user categories. The range of users is as follows: • Public administration worker: This person is the furthest from the technical side and is responsible for the searching and filtering on the registry. • Model engineer: This person develops/alters/… the model that was given to him by the public administration worker. • Software developer: The responsibility from the software developer is connecting their applications to the registry. • Casual browser: Someone who is just browsing the registry without a clear objective in mind.

Use cases Slides 15 - 22 **Speaker:** Claudio Baldassarre	<u>Exemplary use cases</u> Next, Claudio went over the context of some use cases that would be used to explain how the registry might be used. For these use cases, 3 different actors and the registry were introduced. *Use case 1: Preparation* Claudio went over the use case about publishing a model on the registry. A question for this use case was about the preparation and willingness of the Member States to publish metadata in the correct format. • Riitta mentioned they use an API interface. She continued saying that describing a model with an API call would be quite easy. Nothing needs to be done apart from the content regarding the ontologies. • Claudio asked if the API would be used to expose the content the registry needs. • Riitta answered that this was indeed the case. This way no extra harvesting would be needed. • Sander added that most models have already a specific domain mentioned in the metadata. The controlled vocabularies must be reworked a bit before they could be used within the registry. When you search for them, it could become difficult the more of them there are. • Matteo and Claudia both mentioned that in Italy, API connections from and to the national repository are in place. Therefore, direct API connection between Italy's repository and the EU wide Registry is feasible. • Claudio noticed that both approaches (API or metadata upload) are viable and what metadata would be used will be decided as soon as possible. It could be that there is more than one vocabulary, but this will be described early stage. **Requirement** Publishing a model to the registry should be possible through two methods: 1) Using an API to expose information from national repositories to the registry 2) Metadata harvesting from national repositories *Use case 2: Discovery* Claudio introduced the second use case, addressing the discoverability of models on the registry. Claudio emphasized that the features mentioned here could be beneficial, but emphasizes that these are hypothetical and could change based on the input received. Nathan added that the application domain and language, as mentioned earlier, could be seen as filter features. The question that

	accompanies this use case is thus also related to the features: Which features should be prioritized and which are missing?
	
	• Sander went over the ranking aspect that the registry could have. In order to be able to rank models and afterwards search faster, manual curation might be needed to rank datasets. Some ways to rank and rate models could be:
	  • Making the link with actual datasets
	  • Using the models
	  • The use of the linked data aspect to use a query on the data portal
	  A huge advantage would be using the DCAT datasets, but participants would need to put the link properly. This would be a big investment but could bring huge value to the entire data cycle.
	• Claudio reiterated that the features proposed are meant to inspire and not to have it as a fixed set of features yet.
	• Sander rebottled that filtering as a feature should be relevant. It would get more relevant when more data would be added to the registry but since the models themselves are the main point, the search feature would be the most important.
	
	Requirement
	
	Within the features the registry could offer, the most important ones that should be prioritized are searching and filtering.
	
	• Ana Rosa: An important notice would be to know that not everything is technically possible. The problem is that discoverability is based on the metadata for search. Different options should be considered, such as class name, property name, context, domain, … The latter is important because the same class could be used in different models within different sectors.
	• Claudio then asked if the search should be broadened with the aspects that were planned with the filter feature; if any features are missing.
	• Pascal mentioned that DCAT-AP3 introduced a new property, 'applicableLegislation'. With considering the context of the use cases proposed here, OpenEHR could be considered to be adopted as an official model to harmonize data. In a context like this, such a property would be nice to have.
	• Claudio answered that each source/repository would come with specific metadata. With the use of DCAT-AP, this problem would not exist. Filter capacities would become bigger.
	• Ana Rosa added that with metadata-based ordering, results could be ranked by topics that someone deems important.
	
	Use case 3: Accessibility

	Claudio began by mentioning that the content of this use case, about being connected to the registry, was already mentioned by Riitta. The question for this use case is about what ways someone would like to be able to retrieve models from the registry.
	
	• Sander suggested that regarding accessibility, a model should be first found and only then could matching properties be compared. Linked metadata could be behind these properties and could be referenced to get content from the original model. Regarding the user experience, it should not feel as linked data.
	• Claudio mentioned that the registry would be based on the FAIR-principles. If an import mechanism is used, every time a model is loaded, import would be written as 'go to this location' and the right version of the URI is chosen. Once a mechanism to access the models is in place, it could be rendered to your liking.
	• Sander believed this was more of an engineering question: how you get access to the data, how you would access the metadata and how you would fetch the model.
	• Riitta mentioned that there are many options to access the registry.
	  • Having a core vocabulary on the European level. That way nothing specific must be done to the specific data model and you could say your data model is derived from a specific subclass from a general class. This would work as long as the European core vocabularies have URI addresses.
	  • Other Member States could use those core vocabulary references to download (in JSON, RDF, … format). Afterwards, they could make some modifications but the reference to the original European data model would remain.
	  • The model could also be downloaded and the references to the original model could be taken away, making the new model entirely your own but sparing yourself the trouble of inventing anything from scratch.
	
	  With the last two points, the ability to download models should exist. For this, a serialization should be put in place. Models could also be stored locally, but they should have a high-level data model to refer to and could be referred to but only locally as well.
	
	• Claudio asked if Member States would be ready to do this and said this was something to be discussed during the next workshop.
	• Lastly, Felicitas mentioned being able to download models in RDF, JSON-LD, TTL would be nice, but also being able to browse through the ontology online would be quite helpful.
	
	This last requirement is not seen as in scope of the registry, as this option lies on the repository side (The Member State maintaining a national repository).

Use case 4: Collaboration

Claudio introduced this use case by asking how Member States expect to co-design models on the registry; what would help with the co-designing process.

- Sander mentioned that during the co-designing process, being able to discover what other Member States are doing is already a good first step. When the contact details are available on the registry, users have the possibility to contact them outside the registry, which is more than enough. It should not be made into a separate feature.

Requirement
The contact details of the creator(s) of a certain model should be retrievable from that models' metadata, which gives Member States the chance to contact the creator(s) directly.

- Riitta emphasized that co-designing is an essential element in creating common data models.
- Felicitas said that it could be a possibility to let everybody suggest terms and properties but only a small group is permitted to edit that specific model.
- Martynas proposed to use graphical representations for semantic assets. Something that could be used is e.g. WebVOWL.
- Sander added that a link to an issue tracker, such as done on GitHub is also an option. This way, you would be able to communicate and suggest a change in a place where actual collaboration happens.
- Claudia suggested that GitHub could be a good starting point and afterwards, other people/entities could be added. The reuse of semantic assets should be kept in mind. If another entity would want to reuse a certain model, the model should come with instructions on how to use/apply this model.
- Claudio said models could be aligned, but entities should search other models they could use (a part of) beforehand.

Use case 5: Integration

Claudio opened the last use case by saying a balanced use of deciding what metadata to use should be made. This way, a decision process would be made that includes everybody, making it possible to influence the decisions and to see the results. Other than with metadata, through API is the other solution. 2 dataflow styles were proposed and are:

- Upstream: the exposed information would go from the Public Administration to the registry.

- Downstream: the exposed information would go from the registry to the Public Administration.

The question regarding the last use case was which information stream would be preferred. If anyone has some experience with either one of those streams, why would it (not) be suitable for the registry Claudio continued. Nathan added whether everybody understands the difference between the 2 streams.

- Ana Rosa said that an open modelling tool would help to reach the quality criteria (e.g. ArchiMate).
- Claudio asked how Member States allow their content to be discovered.
- Claudia answered saying that they have a harvesting process in place.
- Ana Rosa said that both streams require a mapping transformation for an automated integration.
- Riitta added to that, saying that they already have 2 ways to download information
  - Browse the interface and download the serialization (RDF, Turtle).
  - Make a search request to the API interface (JSON).

  Riitta further asked what would be done when some models would be private, such as restricted access to (a segment of) users.
- Sander wondered whether this was something that should be kept in scope already or should be kept for the pilot. According to him, when you want to foster collaboration, models should be open. Sander also mentioned that dataspace building blocks (e.g. IDSA protocol) for restricted access could be used in the future pilot.

Requirement
Not every model would be accessible to anyone, so access rights is considered important. Open access is preferred but restricting access should be explored.

- Claudio said he did not have an answer to this yet, but this would be further investigated. This however already paves the way for the future pilot.
- Emidio explained a bit further the difference between the 2 streams. He said the downstream workflow means the registry will harvest the models from the repository from the Member State. The upstream workflow however does the opposite, and the Member State will access the registry via an API. Which workflow would be chosen depends on the use of the registry and the national repositories. With the downstream workflow for example, the registry would need to know when it could get

	access to the repository environment regarding timing, security, …
Summary & next steps Slides 23 - 28 **Moderator:** Claudio Baldassarre	<u>Roadmap</u> After going over the use cases, Nathan went over the timeline for this working group. He highlighted that the second workshop, about the technical requirements, will be held in May. A lot of content for that workshop has already been touched during this workshop and will be looked at. Even further, the pilot that would commence after the summer would be looked at how the realization could be done. <u>GitHub</u> Nathan continued by presenting the dedicated GitHub for the registry. Here, the result of this workshop will be posted and feedback could be logged via the 'issues' function. <u>Closing</u> To close this workshop, Claudio invited all participants that they do not need to wait to address us. A collaboration outside and/or before the next workshop is possible. Claudio also thanked everybody for their collaboration during this workshop and wished everybody a good weekend.

Conclusion

In the context of preparing for an EU-wide Registry of semantic models' pilot, a first workshop around the adoption requirements for this future pilot was organized. This was the first of 2 workshops, where the latter will be organized around the technical requirements such as metadata and APIs. This workshop emphasized the need for participating Member States to be adequately prepared to adapt and use the registry, as well as the willingness to cooperate with the pilot with the continuation of this project.

The discussions highlighted the differences between the registry and national repositories, and the concept of semantic interoperability. The registry was seen as beneficial for its discoverability and findability, co-designing, and harmonization. Concerns were raised about the registry's integration into the workflow of designing models, considering the diversity of semantic models due to language barriers and different modelling methods.

The session also addressed the need for the registry to enable domain-specific filtering of models and to be more than just a redirection to the models. The importance of a common

agreement on how data models should be described was emphasized, along with the need for standards and the ability to compare models across different domains.

The registry was identified as potentially useful for various user categories, and use cases were presented to illustrate how the registry might be used, emphasizing the need for Member States to be prepared and willing to publish metadata in the correct format.

Overall, the meeting underscored the importance of the registry for promoting interoperability among Member States, the need for clear guidelines on its use, and the necessity for a common agreement on how data models should be described. The participants agreed that the registry should provide an interface and set of features that enhance user experience, rather than just being a redirection to the models.

Recognizing the diversity in the working methods of different Member States, it is clear that the development of a tool like the registry for semantic assets requires a collaborative effort. Each state has its unique approach to semantic modelling and data management, and these differences must be taken into account to ensure the registry is universally adaptable and beneficial.

Moving forward, a set of requirements will be considered for the future pilot program. These requirements will be based on the discussions and feedback from the workshop, taking into account the diverse needs and working methods of the Member States.

Requirements

To have a clear overview of the captured requirements, below is a summary of all requirements that were listed during this workshop.

- Having the ability to search for models through domain specific filtering.
- The registry should provide Member States the possibility to get an overview about what has been done on the European Commission level regarding semantic modelling and creating vocabularies.
- Publishing a national model to the registry should be possible through two methods:
    - Using an API to expose information from national repositories to the registry
    - Metadata harvesting from national repositories
- Having the ability to compare a model with its former self. This way, changes within that model would become clear and the reason why these changes were made could be clarified.
- Regarding the features the registry should offer, the focus should be on foundational searching and filtering options.
- The creator(s) of a certain model should be listed within the metadata. Having the contact details available on the registry would give other Member States the opportunity to contact the creator(s) of certain models for collaboration for new models or for information for example.
- The access rights that other Member States would have is also important. To foster collaboration, having access to all models would be preferred, but restricted access could be granted if the need is there.