# Handbook and Style Guides

Different 'general comments' have led to the conclusion that distinguishing clearly between the goals, targeted audience and content of the '**eGovernment Core Vocabularies Handbook**' and the '**eGovernment Core Vocabularies Style Guides**' is urgently needed.

Subsections below reflect our opinion on this. In any case, the general comments are emerging as critical inputs for the scoping of the handbook content and the disambiguation between what should go in the handbook and what in the style guides.

## Handbook

### Audience

Policy makers, CIO, business owners and project leaders. Rationale for these targets is provided in farther sections below.

### Goals and content (TOP-DOWN approach)

1. Provide General Context:

* Alignment with the EU general principles and policy;
* Alignment with the European Commission principles and strategy;
* Alignement with the COM Interoperability strategy and principles, more concretely:
  + - References to the legal framework related to interoperability, standardisation and reusability (cfr. EIF Decision, Regulation 2015/2012, eur-lex texts on reusability, etc.);
    - Summary of ISA2 goals, actions and initiatives;
    - Summary of the SEMIC goals, works and road-map.

1. Provide Public Administation Interoperability-specific Context
2. The saliency of interoperability in the EU Digital Public Service Value Chain;
3. The role of the semantic interoperability;
4. The importance of sharing and reuse.
5. The need of a EU Catalogue of Reusable eGovernment Core Vocabularies:

* Explain and justify the need;
* Set the principles and recommendations for the identification and selection of reusable resources (cfr. the pros-and-cons comment, below);
* Illustrate with examples.

1. eGovernment Core Vocabularies:

* Definition
* What’s in ‘core’?
* Who should initiate them?
* When and how to develop a new eGovernment Core Vocabulary?
* Recap: ‘the story-line’ (illustrate the end-to-end process, including the identification of the need, roles, actions, groups and flows of tasks, etc.)
* The need of sharing new eGovernment Core Vocabularies (Sharing-by-design[[1]](#footnote-1))
* Principles and recommendations for the governance of eGovernment Core Vocabularies
* The Style Guides as an enabler of the semantic interoperability:
  + What are the Style Guides: introduction
  + How the Style Guides enable semantic interoperability
  + Why they should be systematically applied, by whom, under what circumstances (what for, etc.)

1. Application Profiles:

* Definition
* Who should initiate them?
* When and how to develop Application Profiles: story-line?
* Use Cases, examples
* Are APs shareable and reusable? When, how, by whom?
* Principles and recommendations for the governance of APs

1. The Style Guides as an enabler of the semantic interoperability:

* What are the Style Guides;
* How the Style Guides enable semantic interoperability;
* Why they should be systematically applied, by whom, under what circumstances (what for, etc.)

## Style Guides

### Audience

Project leaders and technical implementers. Rationale for these targets is provided in farther sections below.

### Goals and content

The revision of the current content of the eGovernment Core Vocabularies Style Guides.

# COMMENT 1: Pros and Cons about the reuse of vocabularies

## General considerations:

Sharing and reuse do not apply only vocabularies, but any potentially reusable 'semantic interoperability solution' (i.e. ontological and non-ontological resources, including core vocabularies, controlled vocabularies, validations artefacts, documentation, etc.). This argument should be clearly stated and illustrated both in the handbook and the guides, since it is especially important for the ‘sharing-by-design principle’.

## Pros

Considerations on what could be included in the handbook follow. They are based on abundant literature on this topic, namely very recent works developed for COM[[2]](#footnote-2).

* **Semantic interoperability**: semantic interoperability is enabled. This is well explained in the latest version of EIRA, see the Key Interoperability Enablers View: Ontologies are Key IOP Enablers in the semantic layer. For the Enablement to be activated the ABB (ontology) is to be ‘Shared and Reused’ (Sharing and Reuse are the two critical events resulting into components that can be considered Key IOP Enablers -> The Structural Interoperability Saliency is defined in terms of ‘Sharing and Reuse’);
* **Modularity**: modularity is promoted in front of monolithic solutions. Modularity is the architectural approach making possible the Structural Interoperability. The more modular the more reusable;
* **Standardisation**: openly sharing modular components promotes the standardisation of semantic artefacts. The (re)use of open standards is one important recommendation in the EIF and a critical factor for the enablement of interoperability. One major benefit of open standards is the proliferation of compatible solutions that are Vendor-independent;
* **Data is the asset**: the reuse of standard vocabularies and ontologies provide the opportunity to **consume** existing and up-to-the-date datasets. These data are invaluable assets for the implementation of the Public Service Policy and the Digital Public Service Delivery.
* **Focus and time-to-market**: the reuse of open standard domain-agnostic semantic artefacts, e.g. ontological and linguistic resources, allows policy, business and technical implementers to focus on the policy area and business domain problems to be solved. This also reduces the cost of maintenance and evolution and boosts the time-to-market;
* **Stability and sustainabililty**: 'Catalogues’ of reusable vocabularies are (to be) founded on sound governance principles. One of such ‘sound’ governance principles is that the vocabularies to reuse need to be developed and maintained by a stable SSO, SDO or Community ensuring the open access to their governance processes and developments. Also, these organisations have special interest in 1) releasing stable developments (i.e. provenly usable), and 2) ensuring backwards compatibility. This benefits directly the business and digital continuity, which are critical for the the long-term sustainability of the Public Services and the justification of the investment in their development and maintenance.

## Cons

* **Compatibility and *etiquette[[3]](#footnote-3)***: When SDO and Communities develope vocabularies they sometimes couple their conceptual models to specific design and implementation paradigms. Thus, many vocabularies that are of the interest for Public Administrations are only available as RDF models (some of them are not even represented graphically, e.g. FOAF, DCT, other). Some SDO Technical Committees and vocabulary publishers (e.g. OASIS UBL[[4]](#footnote-4), the Publications Office of the European Union footnote:[link:https://docs.oasis-open.org/ubl/UBL-2.3.html[UBL] produces W3C XSD and JSON Schemas. OP's https://op.europa.eu/en/web/eu-vocabularies[EU Vocabularies] publishes all the authority tables in different formats, XML, HTML, SKOS, Genericode.]) are moving towards the production of 'syntax-binding-agnostic' models that would decouple the models and ontologies from 1) the syntax used to express the core model, 2) the expression of additional richer semantics (to add on-top-of the core model), and 3) core reusable rules and restrictions applicable to the core model and the added-on semantics. SEMIC is promoting this methodology for the development and maintenance of vocabularies and ontologies. However, for those vocabularies and resources (e.g. validation methods and artefacts) that are only available under one design and implementation approach and technology, 'mappings' and 'adjustments' between the original specification and the desired design/implementation need to be done. One example of this has been the need to 1) map the DCAT RDF T-Box (and the vocabularies reused by DCAT) to XML Schemas, and 2) to adjust certain datatypes used in RDF to other preferred in XML (e.g. rdf:resource to either xsd:anyURI or xsd:anyType). These actions should to be reported back to the SDO maintaining the specification, as well as the need of decoupling the model from the technological aspects (and producing alternative syntax-bindings). In the meantime, the situation described imposes:
* *The need of constantly watch the evolution of the SDO's works;*
* *The need of maintaining the inner developments that depend on the SDO's works evolution;*
* *The need of re-adapting tools and solutions;*
* *The need of well communicating and disseminating the evolution of the inner works originated by this evolution imposed from the outside;*
* **Adaptibility**: One of the problems of reusing externally governed resources is that requirements that should fall within the compass of one resource, e.g. one vocabulary, are not covered sufficiently or at all. When this happens, the alternatives are not always viable and need to be dealt with internally. At least three situations can be identified related to this problem: pace, customisation and extensibility. The identification of these situations is a relevant task when building the catalogue of reusable ontological and linguistic resources:
* ***Pace***: new legal and business requirements need to be introduced timely in the vocabularies. If the SDO responsible for the maintenance of the vocabulary is not reactive[[5]](#footnote-5) and adapts the pace of its developments to the [re]user needs, the user will need to either look for 1) adapting the vocabulary to its needs, 2) to come up with its own *\_ad hoc\_* solution, or 3) look for combinations of alternative vocabularies that partially meet the user requirements. The optimal option would be the first one, that the vocabulary is adaptable. One indicator that the vocabulary is adaptable is that it can be customised and extended.
* ***Customisation*** *and* ***extensibility***: vocabularies are not truly reusable if they set very strict conditions or constraints[[6]](#footnote-6). As a matter of fact, constraints and particular business entities cannot be fully identified until the vocabulary is 'applied' to a context and, more specifically, to a business domain. Hence, a relevant factor (i.e., a *\_sematic interoperability enabler\_*, cfr. EIRA Key Interoperability Enablers) when identifying and selecting reusable candidate vocabularies is whether they are flexible enough so they can be, later on, be 'profiled'. The ISA2 Core Standard and Specification Vocabulary (https://joinup.ec.europa.eu/solution/core-standards-and-specifications-vocabulary-cssv[CSSV]) and the W3C https://www.w3.org/TR/dx-prof/[The Profiles Vocabulary] use the following definition of 'Application Profile': A [data/application] specification that constrains, extends, combines, or provides guidance or explanation about the usage of other [data/application] specifications".
* **Accessibility\***: a relevant criterion for the selection of reusable resources is how unrestricted is the access to the resources, related documentation, reference implementations and to the participation in their evolution. These are also key factors that moderate the sharing and reuse of the semantic interoperability solutions, and therefore are Key Interoperability Enablers that contribute to the saliency of the interoperability in the European Digital Public Service Value Chain (EDPSVC). They should be taken into account when drafting the principles and recommendations for the cataloguing of reusable semantic solutions.

# COMMENT 2: Target audience

The target audience would vary depending on whether we are referring to the Handbook or to the Style Guides.

In our opinion, Cecile's suggestion of documenting the process of developing a Core Vocabulary is convenient and **should go in the handbook**, since it is of the interest of 'everyone'. The handbook should be the first reference document that ends up leading to the Style Guides.

Implementers, therefore, should also be recommended to have a look into the handbook to grasp the context and the 'story' of how a Core Vocabulary is expected to be developed, maintained, evolved, etc.

Some rationale to support the target audience of both documents, the handbook and the style guides, follow:

**== COMMENT 3: target audience, what's an eGCV (definition and justification), just Core Vocabularies? and what about APs?**

\* About the target audience see comments in sections above.

\* About the definition of what is an eGovernment Vocabulary and the justification of their existence:

. Both should go in the handbook;

. The Style Guides should summarise them and refer to the handbook for more details.

This would be consistent with the audience, goals and content of both documents.

\* About the scope of the Handbook and Guides: only eGCV? Our opinion is that the scope should encompass:

\*\* *\_Handbook\_*: principles, recommendations and how-tos for the starting and inception, development and governance of eGovernment Core Vocabularies and Application Profiles;

\*\* *\_Style Guide\_*: architectural and technical guidelines and examples for the implementation, testing, validation and use of eGovernment Vocabularies and Application Profiles.

**== COMMENT 4: Reuse needs further clarification**

**=== What is normative and what non-normative**

Yes, an effort should be done to clarify what is normative and what is just recommended, to whom it affects and how. However, this is question is related to another dimension which is:

**[.text-center]**

'*\_Who should make the decision of when to develop a new eGovernment Core Vocabulary?\_*

**[.text-left]**

And similarly for Application Profiles.

The answer to this question may be **\*\_politically problematic\_\***. A trade-off in the expression of these responsibilities could be necessary.

The right place to introduce recommendations on this aspect should be the Notebook, not the Style Guides.

General principles and recommendations could be issued though +++<u>in the Style Guides</u>+++. Some of them are already present in the Guides, but would need to be complemented with a kind-of-tag 'Normative' or 'Non-Normative', possibly accompanied by an explicit icon:

\* Import, do not re-define (NORMATIVE),

\* Do not change the datatype of the original entity (NORMATIVE),

\* Comment in-line (NON-NORMATIVE), etc.

In the +++<u>Notebook</u>+++, a general principle could be added in the sense that NORMATIVE guidelines should be applied when the semantic interoperability is threatend if the guideline is not thoroughly applied.

1. Reusing is not activated if the Sharing-by-design principle is not applied! [↑](#footnote-ref-1)
2. See http://www.heppnetz.de/files/dke2008.pdf and https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3315457, as well as ISA2 Actions such as Sharing and Reuse, CAMSS, SEMIC governance, etc. [↑](#footnote-ref-2)
3. The term 'ettiquette' is used in some scientific literature about interoperability. It has been applied related to the ICT communications fields to refer to 'protocols of protocols', to 'content-negotiation' and to 'syntax-bridging'. Cfr. XXXXX-REFERENCE]. [↑](#footnote-ref-3)
4. OASIS UBL-2.3: https://docs.oasis-open.org/ubl/UBL-2.3.html [↑](#footnote-ref-4)
5. There may be many reasons for the unresponsiveness of an SDO (or of a particular SDO TC), e.g. the SDO is not affected by the legislation, the SDO's roadmap is misaligned with the incoming requirement, the SDO policy does not allow for the inclusion of requirements that are not sufficiently global (i.e. international or of insufficient interest to a critical mass), etc. One example of highly responsive specification developers are OASIS business-oriented TC, e.g. UBL, LegalDocML, other.]. [↑](#footnote-ref-5)
6. The normalisation and standardisation of restrictive vocabularies become severely hampered. [↑](#footnote-ref-6)