



Publications Office
of the European Union

Directorate A - Information Management
Unit A.1 - Standardisation

CRDM-2

First set of best practices developed in phase 2

Date:	10 December 2020
Version:	V1.00
Authors:	Makx Dekkers
Unit(s)/sector:	Unit A.1 -- Standardisation
Revised by:	
Target:	
Reference number:	
URL:	
Associated procedure:	

Note: the best practices are available on-line at
<https://webgate.ec.europa.eu/fpfis/wikis/display/ReferenceData/T5+-+Best+practices>.



Contents

1.	Best practice: Availability	3
2.	Best practice: Communication and notification	5
3.	Best practice: Extensibility	7
4.	Best practice: Feedback and request handling.....	9
5.	Best practice: Multilingual descriptions	11
6.	Best practice: Multiple formats and access mechanisms.....	13
7.	Best practice: Persistent identifiers.....	15
8.	Best practice: Release, Versioning and Archiving	17
9.	Best practice: Reuse of existing internal reference data.....	19
10.	Best practice: Reuse of external reference data.....	20

1. Best practice: Availability

What

Corporate reference data assets must be available without interruption with sufficient measures to ensure performance and resilience through trusted services and platforms. The wider use of corporate reference data makes it necessary that the asset can be accessed at anytime from anywhere at the Commission.

How

In practice, the level of commitment for each reference data asset must be determined based on an analysis of how the asset is used. For example, if users download the whole asset at regular intervals, the asset needs to be available at the days and times that such downloads take place. If on the other hand, an asset is accessed continuously, it must be accessible at any time, although it could still be possible to schedule maintenance downtime, for example at night.

Measures to be taken to ensure the target availability may include using resilient hardware and power sources, off-site mirroring with automatic switchover in case of outage, and well-defined alerting tools and procedures. Management tools should allow multiple layers to be able to satisfy different sets of quality attributes in terms of flexibility, scalability and resilience.

Who

The maintainer of the reference data asset is responsible for advertising the target availability in terms of uptime and scheduled downtime and for taking measures that ensure the target availability. Operational responsibility can be taken up by internal staff, shared with other Commission services or outsourced externally.

Benefits

Providing clear commitment about the target availability and meeting those targets will ensure that applications that use the reference data can run smoothly, and create trust in the reference data asset and the maintaining service.

Challenges

Commitments concerning availability require resources, both in terms of technical environment and in terms of staff time. This needs to be carefully planned, always taking into consideration the target; higher commitments to availability requires more resources, so a reasonable balance need to be struck.

Examples

There are currently no examples of documented commitments for availability. However, many of the corporate reference data assets are continuously available, such as the [EU Vocabularies](#) at OP and the [INSPIRE registry](#) at JRC. It has been noted that explicit commitments for the corporate reference data assets could be formalised in Service Level Agreements.

References

Improving Availability, Data Protection and System Performance. SiriusEdge, 2015.
<https://edge.siriuscom.com/infrastructure-operations/improving-availability-data-protection-and-system-performance>

Four nines and beyond: A guide to high availability infrastructure. Atlassian, 2018.
<https://www.atlassian.com/blog/statuspage/high-availability>

Best Practices For IT Professionals To Improve Network Availability And Performance.
CleanTech, 2016. <http://cleantechservices.ca/information-technology/best-practices-for-it-professionals-to-improve-network-availability-and-performance/>

9 best practices for improving your website uptime. Monitis, 2017.
<https://www.monitis.com/blog/9-best-practices-for-improving-your-website-uptime/>

2. Best practice: Communication and notification

A communication and notification channel must be available for all users to be kept informed of any changes in the assets that they use, so that they can take necessary actions, for example to make change requests before a cut-off date, to participate in review of proposed changes or to plan changes to local systems to prepare for changes in the reference data.

How

If the reference asset is associated with a clearly delineated user community, the existing channels for that community should be used. If usage is more differentiated and the extent of the user community is not known, any channels available should be used, for example a notice on the website associated with the asset.

Stakeholder groups like the Local Data Correspondents, Digital Stakeholders Forum, the Information Management Team and others should be kept informed of changes in corporate reference data assets.

In case of changes that are not backward compatible, the announcement of impending changes needs to be made with sufficient advance notice so that users can plan changes in their systems and applications.

Possibly a subscription service for announcements related to the corporate reference data assets could be provided by the EC Data Catalogue.

Who

The maintainer of the asset is the first responsible to make sure that communication and notification is organised in such a way that the user community and other stakeholders are informed of changes and any other event relevant to the users.

Benefits

A good communication and notification approach will give users sufficient warning that changes are going to take place. This will allow them to assess and plan any actions necessary to prepare for the changes, for example to make changes to their systems and plan re-ingestion of the asset if they hold a local copy, thereby ensuring that systems that use the asset will continue to operate smoothly.

Challenges

The maintenance of a good communication and notification approach requires continuing attention and assigned resources. Good planning and explicit inclusion of the communication and notification actions in a project plan governing the maintenance activities will help in the efficient and effective use of resources.

Examples

At BUDG changes are communicated within the coordination network supported by regular “What’s new” notes.

MARE sends out weekly e-mails to the users in which changes are announced.

For EU Vocabularies, OP publishes the dates of the trimestral release schedule and sends out notifications ten days before a release.

References

7 Best Practices for Communicating During Change. Implementation Management Associates IMA, 2017. <https://www.imaworldwide.com/blog/7-best-practices-for-communicating-during-change>

The Foundations of Successful Reference Data Management. TopQuadrant, 2019. https://www.topquadrant.com/docs/whitepapers/TopBraid_ReferenceDataManagementWhitepaper-3-18-15.pdf

How to Communicate Product Changes To Your Users. Uservoice, 2020. <https://community.uservoice.com/blog/how-to-communicate-product-changes-to-your-users/>

3. Best practice: Extensibility

What

Reference data assets have to allow for local extensions in which concepts are added that are only relevant for particular applications. Local applications may have a need for an extended or enriched set of concepts, either for additional concepts or for a deeper level of concepts underneath existing concepts. Extensions need to be visible also on the corporate level to other users of the asset.

How

For every asset, specific extension rules need to be defined, e.g. no extensions allowed, only extensible with narrower values, extensible with values on any level.

For reference data assets for which extensions are allowed with new concepts, a proposal for a new concept needs to be analysed whether the proposed concept:

- a) is well defined, i.e. with a definition that is unambiguous, states the essential attributes and is not circular
- b) has the same type as the other concepts already in the asset, avoiding that the asset violates the rule that an asset should only contain concepts of one type
- c) does not conflict or overlaps with concepts already in the asset which could create confusion, i.e. the new concept needs to be clearly different from all the other concepts in the asset
- d) has semantics that fall strictly within the semantic of a broader term, in case the asset is structured as a hierarchy and the extension is proposed as a narrower term for one of the existing concepts

Corporate reference data assets may be enriched for particular purposes, for example by defining specific groupings of concepts like regions that consist of a set of countries, adding translations of descriptions, or addition of links to other information, including relations of a concept in one asset to a concept in another asset.

New concepts and enrichments must be incorporated in the corporate asset and maintained by the service that is responsible for the maintenance of the asset.

The maintainer should provide possibilities to select a particular application-specific view of the asset, i.e. a 'use context', for example suppressing some concepts, e.g. non-recognised countries for legal context, or using a subset of a larger asset, e.g. using only EU countries from the corporate asset for countries.

The maintainer should make all DGs aware of extensions that are included in the asset, to avoid the situation that multiple DG request a similar extension rather than use one that is already there.

Who

Extension may be proposed by users of the asset to meet their specific requirements. In doing so, they should document their use case as well as provide a clear definition of the proposed concept or enrichment

The maintainer of the asset should analyse the extension and provide feedback to the proposer and, if the proposal is adopted, should use the available communication and notification channels to inform the other users of the asset.

Benefits

Managing the extensions as part of the asset and thereby making such extensions visible on the corporate level will facilitate wider use, and avoids duplication of similar concepts.

Challenges

Integrating extensions into the asset, rather than keeping the extension local, implies that it will take longer to get the extension ready for use, due to the involvement of the maintainer of the asset and the process of evaluating whether the extension is appropriate as part of the asset.

In order to keep the delays to a minimum, the maintaining service needs to ensure that there are human resources available to react quickly to proposal and respond to proposals in a timely manner.

Whenever there is disagreement about the proposed extension between the proposer and maintainer, the Coordination Group may be involved to make a final decision.

Examples

At JRC, the concept of extension is present in all code lists in the INSPIRE Registry. There is an attribute for code lists that informs whether the values are extensible or not and in which manner.

References

The Reference Data Pattern: Extensible and Flexible. Vertabelo, 2016.

<https://www.vertabelo.com/blog/the-reference-data-pattern-extensible-and-flexible/>

Designing for Extensibility: An action research study of maximizing extensibility by means of design principles. Univ. Gothenburg, 2009.

https://gupea.ub.gu.se/bitstream/2077/20561/1/gupea_2077_20561_1.pdf

The Definitions Book: How to Write Definitions. Unified Compliance Framework.

<https://www.unifiedcompliance.com/education/how-to-write-definitions/>

4. Best practice: Feedback and request handling

What

Clear and well-documented procedures and processes must be established to allow stakeholders of a reference data asset to provide feedback and make change requests.

How

Information provided to the stakeholders must include a reference to the responsible person, unit, or committee, a functional mailbox for feedback and requests, and a link to an issue tracker if available. Submitted feedback and change requests must be handled in a reasonable timeframe, acknowledging receipt, and providing information on the process that will be followed to handle the submission including an indication when the submitter can expect to receive a response.

The procedure for request handling should be made public. This should include information about the general timeline for processing and who is involved in taking decisions.

A common approach across the Commission may be considered e.g. using a shared infrastructure with functional mailboxes, issue tracker, and possibly also automated workflow tools.

Who

The feedback and request mechanism must be provided by the maintainer of the asset. The maintainer must also commit to handle the feedback and follow up in a reasonable timeframe, if at all possible, in the order of days.

It is advisable that maintainer keep track of dependencies in order to understand which users and systems would be affected by a particular change.

Benefits

Visibility of the way that feedback is handled and how changes in the reference data are organised ensures a transparent process and creates trust. Users will be able to see who is responsible, how requests and feedback can be submitted and how requests are handled.

Challenges

The handling of feedback and change requests imply that the maintainer of the asset has resources available to follow up on submissions. An issue tracker and automated workflow tools may help in ensuring timely handling and resolution of issues.

Examples

Most DGs/services already have an approach with a functional mailbox, for example OP-EU-VOCABULARIES@publications.europa.eu for the EU Vocabularies at OP, MARE-DATA-MANAGEMENT@ec.europa.eu for the code lists related to fishing at MARE, jrc-inspire-support@ec.europa.eu for the INSPIRE code lists at JRC. In addition, JRC has an [issue tracker](#) where the status and progress of requests can be followed. Furthermore, the management of code lists at JRC is based on [ISO 19135](#) Geographic information — Procedures for item registration.

References

5 best practices for managing reference data. Lights On Data, 2018.

<https://www.lightsondata.com/5-best-practices-for-managing-reference-data/#>

Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Feedback.
<https://www.w3.org/TR/dwbp/#feedbacksection>

5. Best practice: Multilingual descriptions

What

All assets and concepts must follow the [language policy of the European Commission](#). The European Union is a multilingual environment where the European Commission should set an example of good practice. The stated aim of the language policy is to strike a reasonable balance between respect for speakers of the EU's many languages and practical considerations such as the cost of translation.

How

Taking into account the expected use and available resources, it needs to be decided for each asset in which languages descriptions need to be provided. In any case, all text should always be accompanied by an indication of the language in which it is written. Following the language policy, content should be published at least in English, and, if possible, also in French, German and Spanish. Assets that are governed by European legislation should be made available in all 23 official languages of the European Union. In specific cases, translations beyond the official EU languages (regional languages, languages of EEA and accession countries) could be provided. Some assets may not need translations, e.g. place names in the national language or biological nomenclature in Latin.

Who

The maintainer of an asset is responsible for the provision of translated content. When official translations are required, the Directorate-General for Translation (DGT) should be involved to provide support. In specific cases, national authorities may be involved in translation of text into their national language. However, DGT would be able to provide translation of labels and descriptions only in exceptional cases, including into German and French. Since such translations are very resource-intensive, there would need to be a demonstrable need for the language versions. When rough translations are needed, the machine translation service eTranslation, provided by DGT, can be used. It is recommended that maintainers use this service rather than an external service like Google Translate.

Benefits

The European Union is a multilingual environment where the European Commission should set an example of good practice. While it may be expected that information in English can reach a majority of users, the provision in multiple languages increases the usefulness of the information and fosters its wide reuse.

Challenges

Translation of information and maintenance of translated text is cost- and labour-intensive. Therefore, the language skills and expectations of the intended user community need to be analysed. Based on that analysis, a balance should be found between cost and benefit of translations.

Examples

Several reference data assets include descriptions in all official languages of the EU, and some also include additional languages, e.g. languages of EEA countries. Examples are [Eurovoc](#) at OP and the [INSPIRE Registry](#) at JRC. Other assets contain concepts described in English, French and German (Eurostat), or in English, French, German and Spanish (DEVCO).

References

Best Practices for Multilingual Linked Open Data. Jose Emilio Labra Gayo, Univ. Oviedo, 2012.
<https://www.slideshare.net/jelabra/best-practices-for-multilingual-linked-open-data>

Guidelines for Multilingual Thesauri. IFLA, 2009.
<https://archive.ifla.org/VII/s29/pubs/Profrep115.pdf>

Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Metadata.
<https://www.w3.org/TR/dwbp/#metadata>

Best Practices for registers and registries & Technical Guidelines for the INSPIRE register federation. JRC registry team, 2017. Best Practice 8: Provide registers in different languages.
<https://inspire.ec.europa.eu/file/1749/download?token=jaBumb59>

6. Best practice: Multiple formats and access mechanisms

What

A corporate reference data asset must be available for distributions in multiple, machine-readable formats to allow reuse by applications that use different technologies. These formats may include schema expressions using RDF/XML, JSON-LD and XSD.

How

The formats in which reference data assets are distributed have different machine-readable semantic expressiveness. From lower to higher expressiveness:

1. Human-readable information like HTML and PDF
2. Basic machine-readable expressions like Excel or CSV
3. Syntactic expressions like XML
4. Semantic expressions like RDF/SKOS

The choice of expression should be made based on the expectations of the primary users of the asset. However, it needs to be acknowledged that it is possible to provide data defined with higher expressiveness in a format on a lower level, but not the other way around. For example, data in RDF/SKOS can be converted to XML or XLS but an XML expression cannot easily be converted to RDF/SKOS. It is therefore recommended to strive for definition of reference data with the highest semantic expressiveness as much as possible. The use of semantic tools for the creation and management of reference data should be considered.

It is important to make clear to the users what the limitations of the distributed formats are, for example when the level of detail is different (e.g. simplified, complete) for different distributions.

Distribution of assets may be done both in a ‘push’-approach where the maintainer takes the initiative to deliver the asset to the users in a format that they can integrate in their systems, or in a ‘pull’-approach in which the users select and download a file that contains the asset, or access it through a data service like an API or SPARQL endpoint.

Who

The maintaining DG/service/unit is responsible for the definition of the asset and its concepts. The actual creation of the distribution formats may be outsourced to other services or suppliers but the responsibility remains with the maintainer.

Benefits

The provision of multiple formats allows a wide range of applications to reuse the corporate reference data asset.

Challenges

Current assets, for example code lists, may have been created in a format that serves a particular application. This may have been an application that uses data in XML and therefore the code list will have been created as an XML code list, for example the [UBL 2.1 default currency code list](#). Following this best practice may imply the redesign of the code list with a higher semantic expressiveness to allow for the distribution in additional formats.

Examples

Most DGs publish code lists both human-readable formats (HTML, PDF), spreadsheet format (Excel, CSV) and machine-readable formats (XML, RDF). In terms of machine-readable formats, MARE publishes in [UN-FLUX](#) XML and Excel); OP provides RDF/SKOS and XML for the EU Vocabularies; at JRC the central INSPIRE registry provides multiple formats, including (custom and ISO 19135-1) XML, RDF/XML, JSON, ATOM and CSV.

References

W3C Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Data Formats. <https://www.w3.org/TR/dwbp/#dataFormats>

W3C Data on the Web Best Practices. W3C Recommendation 31 January 2017. Best Practice 19: Use content negotiation for serving data available in multiple formats.

<https://www.w3.org/TR/dwbp/#Conneg>

W3C Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Data Access. <https://www.w3.org/TR/dwbp/#dataAccess>

Best Practices for registers and registries & Technical Guidelines for the INSPIRE register federation. JRC registry team, 2017. Best practice 6: Provide registers in different formats.

<https://inspire.ec.europa.eu/file/1749/download?token=jaBumb59>

7. Best practice: Persistent identifiers

What

Concepts in a reference data asset must be identified by persistent identifiers.

How

The concepts in internally defined assets may be identified with HTTP URIs, such as managed under the <http://data.europa.eu/> subdomain, but also references of the type ‘code/scheme/version/agency’ (e.g. UN/CEFACT), or ‘asset/code/validity period’ can be valid identifiers. In any case, an identifier must unambiguously identify a single concept. As the concepts of a reference data asset will be used as values in the description of data resources, it needs to be clear in all cases what the exact meaning of such a concept is, and therefore there needs to be an unambiguous way for the user to look up the meaning of the concept.

It is advisable to include the identifier as part of the documentation of the concept.

The URIs assigned to assets and concepts should be resolvable, i.e. they should lead to (information about) the concept being identified.

Wherever assets and concepts already have URIs, those existing URIs should continue to be used. It is not recommended to mint new URIs when the existing URIs are already known to and used by the users.

Who

If a creator or maintainer wishes to mint URIs, they may contact the Publications Office to request a collection identifier under data.europa.eu for the asset. In that case, the maintainer must also design a structure for the ‘local part’ of the URI so that a reference to a concept in the asset can resolve to the description of that concept.

The maintainer must commit to guaranteeing that the identified assets and concepts are indeed persistent. Identifiers, once created and published, must always identify the same object (asset or concept).

Benefits

A persistent identifier ensures that the meaning of an identified concept is fixed and that an identifier for a concept is not reassigned to a different concept. Persistence is essential for providing the necessary trust in the data and for allowing the development of data-based services.

Challenges

If a DG wishes to use identifiers under data.europa.eu, the Publications Office can support with more information and best practice information on how to design a local resolution mechanism.

Examples

MARE uses identifiers that include Alias, code, validity period which reflect a navigation path in the register.

At JRC, all INSPIRE code lists and all concepts in those code lists are identified by HTTP URIs like <http://inspire.ec.europa.eu/theme> (for the INSPIRE themes register) and <http://inspire.ec.europa.eu/theme/au> (for one of the codes).

At OP, the EU Vocabularies use by HTTP URIs like <http://publications.europa.eu/resource/authority/country> (for the list of countries) and <http://publications.europa.eu/resource/authority/country/AUT> (for Austria).

References

Persistent Identifiers Best Practices. CEOS Data Stewardship Interest Group, 2016. <https://earth.esa.int/documents/1656065/2265358/CEOS-Persistent-Identifier-Best-Practices>

Persistent Identifiers. Digital Preservation Coalition, 2017. <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>

Persistent identifiers: Consolidated assertions. Group of European Data Experts GEDE, 2017. https://www.rd-alliance.org/system/files/PID-report_v6.1_2017-12-13_final.pdf

10 Rules for Persistent URIs. SEMIC, 2012. <https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>

Towards a common approach for the management of persistent HTTP URIs by EU Institutions, SEMIC, 2014. https://joinup.ec.europa.eu/sites/default/files/inline-files/Towards-a-common-approach-for-the-management-of-persistent-URIs-by-EU-institutions_v.0.53.pdf

Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Data Identifiers. <https://www.w3.org/TR/dwbp/#DataIdentifiers>

Best Practices for registers and registries & Technical Guidelines for the INSPIRE register federation. Best Practice 2: Use resolvable HTTP(S) URIs as identifiers for registers and register items. JRC registry team. <https://inspire.ec.europa.eu/file/1749/download?token=jaBumb59>

8. Best practice: Release, Versioning and Archiving

What

The current and all previous versions of a reference data asset must be kept for access. If possible, also all versions of the individual concepts in the asset should be kept. In most cases, reference data need to reflect changes in the environment. New countries or organisations are established, old ones are discontinued, current affairs and evolving knowledge require new concepts to be described. The release schedule must be documented and published, even if releases are done ad-hoc. The current and all previous versions of a reference data asset must be kept for access.

How

A versioning mechanism must be applied, either based on version numbers or time stamps. This could be applied to the whole asset, to the concepts within the asset, or to both.

The history of previous releases including change logs or audit trails must be kept. It is important that older versions are preserved and that no reference data is deleted. It is also good practice never to remove concepts but to mark them as deprecated or withdrawn.

The release schedule should be published in line with user expectations with a balance between stability and flexibility, and should also allow users to plan ahead for the integration of a new release in their systems, especially when a new version is not backward compatible with previous versions. It should also foresee the possibility to process urgent changes outside of the established schedule but only as well-justified exceptions.

Long-term archiving should be foreseen in compliance with the Digital Preservation Strategy. In addition, snapshots of assets could be stored in a long-term archival repository at regular intervals.

Who

The maintainer of the asset is responsible for the establishment of the versioning approach and the release schedule. Archiving issues should be considered, ideally at an early stage in the design and creation of an asset, in consultation with OIB/HAS.

Benefits

Users of reference data will in most cases need to plan ahead for changes to their systems whenever reference data change, for example to enable systems and data to take into account new concepts. If the reference data is properly versioned, not all systems need to upgrade to a newer version at the same time, and even obsolete assets may still be in use in legacy systems. Preserving and archiving reference data will allow researchers into the history of data resources to understand the meaning of historical data.

Challenges

The preservation of reference data assets implies the availability of technical resources and sufficient storage facilities.

Examples

MARE has an approach with scheduled releases with a cut-off date for requests in September, which gives the Member States time to schedule IT resources for the next year to implement changes. MARE also provides two dates for a release, the 'allow to use' date and the 'compulsory to use' date.

At Eurostat, the regulation that governs NUTS specifies that the classification should be stable for at least three years.

OP uses and publishes a trimestral release schedule for the EU Vocabularies, but special releases are published more frequently.

References

Change and Release Management Policy for DCAT-AP. SEMIC, 2017.

<https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/document/change-and-release-management-policy-dcat-ap>

Change management methodology and process for the SRF.

<https://joinup.ec.europa.eu/collection/sharing-and-reuse-it-solutions/change-management-methodology-and-process-srf>

Klump, J., Wyborn, L., Downs, R., Asmi, A., Wu, M., Ryder, G., & Martin, J. (2020). Principles and best practices in data versioning for all data sets big and small. Version 1.1. Research Data Alliance. DOI: <https://doi.org/10.15497/RDA00042>

Best Practices for Storing, Archiving and Preserving Data. Smithsonian Libraries.

<https://library.si.edu/research/best-practices-storing-archiving-and-preserving-data>

Data archiving best practices: The difference between backups and archives. TechTarget, 2010.

<https://searchdatabackup.techtarget.com/tip/Data-archiving-best-practices-The-difference-between-backups-and-archives>

Best Practices for Digital Archiving; An Information Life Cycle Approach. Gail Hodge, 2000.

<https://www.dlib.org/dlib/january00/01hodge.html>

Digital preservation strategy. Secretariat-General SEC(2020) 850.

Best Practices for registers and registries & Technical Guidelines for the INSPIRE register federation. Best Practice 4: Use well-defined statuses. JRC registry team, 2017.

<https://inspire.ec.europa.eu/file/1749/download?token=jaBumb59>

Best Practices for registers and registries & Technical Guidelines for the INSPIRE register federation. Best Practice 5: Do not delete items. JRC registry team, 2017.

<https://inspire.ec.europa.eu/file/1749/download?token=jaBumb59>

Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Data Versioning. <https://www.w3.org/TR/dwbp/#dataVersioning>

Data on the Web Best Practices. W3C Recommendation 31 January 2017. Section Data Preservation <https://www.w3.org/TR/dwbp/#dataPreservation>

9. Best practice: Reuse of existing internal reference data

What

As a default, new IT systems developed by and for the Commission must reuse reference data that already exist at the Commission, and in particular, the assets that are managed under this reference data management policy.

How

This reuse is governed by the comply-or-explain policy that is designed to encourage the reuse of such existing reference data unless such reuse would not be possible or feasible. This comply-or-explain policy has the following aspects:

1. New IT systems developed by or for a Commission service that require the use of reference data in one of the categories in the list of corporate reference data assets, must use the recommended reference data asset
2. Existing systems that are undergoing maintenance should consider using the recommended reference data assets if possible
3. Deviation of the rules (1) and (2) may occur in cases where the use of the recommended reference data asset would not be possible or feasible
4. In case of deviation as stated in rule (3), a thorough explanation must be provided

The evaluation of conformance and the assessment of the explanation mentioned under rule (4) above will be done for new systems by the IT Investment Team.

Who

All Commission services that are planning or engaged in developing new IT systems and applications need to verify that any reference data relevant for the system or application is already available at the Commission as part of the collection of corporate reference data assets.

The maintainer should provide information about how the asset can be used, e.g. if it can be copied or cached and how synchronisation can be done.

Benefits

The reuse of internally available reference data will increase the interoperability across Commission services and IT systems, at the same time avoiding unnecessary double work in defining and maintaining reference data.

Challenges

Shared reuse of reference data implies that a development project has less autonomy in defining its own reference data based on its specific requirements and processes for changes and additions may introduce delays. The comply-or-explain policy includes the option not to use the corporate asset if there are justifiable reasons that this is not possible or feasible.

10. Best practice: Reuse of external reference data

What

As far as possible, externally defined reference data have to be reused, provided that an external asset is managed by a trustworthy source which satisfies criteria of good design and good maintenance.

How

Such trustworthy sources include decentralised EU agencies like EEA, EIGE and ECDC, international organisations like the UN, FAO, WHO or OECD, or standardisation bodies like ISO or CEN.

For the inclusion of external resources in the list of corporate reference data assets, apart from an evaluation of the design and maintenance approach, also aspects such as cost and reuse conditions need to be considered.

For internal use, it may be necessary to enrich the external source, for example by adding concepts that are missing in the source but are necessary for the systems and service at the Commission, or by adding local usage notes, or translated descriptions.

The external source can either be used by reference, in which case the concepts are used directly from the source, or copied to a location under the control of a Commission service, in which case the concepts are used from the copy at the Commission.

Who

For externally defined reference, it is important that a maintainer within the Commission is assigned. This DG/service is responsible for monitoring any changes that are made to the external source, as well as managing the local enrichments. The maintainer needs to establish good communication with the external provider to have early warnings of changes.

Benefits

The reuse of widely used external reference data increases the potential for wider interoperability of resources that are described with it. If the source asset is copied to a Commission service and enriched, the wider interoperability can be combined with the meeting of local requirements.

Challenges

Reuse of external reference data brings with it the dependency on the external organisation that maintains it. It is therefore important that the external maintainer has well-defined processes and policies. At the same time, such processes and policies may result in long delays in making changes or additions.

Examples

Several EU Vocabularies at OP reuse codes defined in ISO standards, e.g. countries (ISO 3166), languages (ISO 639) and currencies (ISO 4217), but with some extensions and modifications.

DEVCO reuses the DAC (Development Assistance Codes) code list from OECD and the SGD (Sustainable Development Goals) terminology from the UN without modifications.

The fishing-related code lists at MARE are taken from FAO, some with and some without modifications.

References

Data on the Web Best Practices. W3C, 2017. <https://www.w3.org/TR/dwbp/#ReuseVocabularies>