

SC8DI07171

**D04.03 Report on the tailoring of the pilot to DG
MARE needs**

Tailoring the Reference Data Deployment Adaptor (REDDA)
development to the needs of DG MARE



Document Metadata

Property	Value
Release date	2014-12-08
Status	Accepted
Version	1.00
Authors	Stijn Goedertier – PwC EU Services Gerben Hoogeboom– PwC EU Services Nikolaos Loutas – PwC EU Services Stéphane Roulier – PwC EU Services Brecht Wyns – PwC EU Services
Reviewed by	Pieter Breyne – PwC EU Services Jesper Abrahamsen - European Commission, DG COMP Roberto Atienza - European Commission, DG COMP Julian-Daniel Jimenez-Krause - European Commission, DG COMP Manuel Perez-Espin - European Commission, DG COMP Carsten Schott - European Commission, DG COMP (external consultant) David Berger - European Commission, Joint Research Centre Jesus Hermida - European Commission, Joint Research Centre Xavier Prud'hon - European Commission, DG MARE
Approved by	Susanne Wigard – European Commission, ISA Programme Vassilios Peristeras – European Commission, ISA Programme Athanasios Karalopoulos – European Commission, ISA Programme

This study was prepared for the ISA Programme by:

PwC EU Services

Disclaimer:

The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

Document History

Version	Date	Description	Action
0.01	2014-05-12	Creation	Creation
0.02	2014-06-02	Internal review	Review
0.03	2014-06-11	Updated	Updated
0.04	2014-06-17	Internal review	Review
0.05	2014-06-30	Update	Update
0.06	2014-07-17	Update	Update
0.07	2014-08-01	Update	Update
0.08	2014-08-04	Internal review	Review
0.09	2014-08-05	Address comments	Update
0.10	2014-08-05	Internal review	Review
0.11	2014-08-06	Address comments	Update
0.12	2014-08-14	Internal review	Review
0.13 - 0.14	2014-08-19 - 2014-08-22	Updates throughout the document <ul style="list-style-type: none"> - Addressing feedback from review process - Rewritten "context" - Summarized description of management processes (D04.02) - Summarized overview of generic and proposed tool chains (D04.02) - Elaboration of change requests 	Update
0.15	2014-08-26	Internal review	Review
0.16	2014-08-27	Address comments - Specification of change requests	Update
0.17	2014-08-27	Internal review	Review
0.18	2014-08-28	Minor updates addressing comments	Update
0.19	2014-09-10	Specification of Re3Gistry functional coverage	Update
0.20	2014-09-16	Address comments from DG COMP	Update
0.21- 0.23	2014-09-30	Internal review	Review
0.24	2014-10-09	Address comments	Update
0.25	2014-10-22	Restructure the deliverable	Update
0.26	2014-10-22	Internal review	Review
0.27	2014-10-24	Address comments	Update

D04.03 Report on the tailoring of the pilot to DG MARE needs

0.28	2014-10-27	Internal review and update	Update
0.29	2014-10-28	Address update	For review
0.30	2014-10-29	Internal review and delivered for acceptance	For acceptance
0.31	2014-11-26	Address some comments from DIGIT	For acceptance
0.32	2014-12-02	Finalisation and delivery for acceptance	For acceptance
0.33	2014-12-04	Finalisation and delivery for acceptance	For acceptance
1.00	2014-12-08	Accepted	Acceptance

Table of Contents

1.	INTRODUCTION	6
1.1.	CONTEXT	6
1.2.	OBJECTIVE AND SCOPE.....	8
1.3.	APPROACH.....	8
1.4.	GLOSSARY	9
2.	ADAPTATION OF THE METHODOLOGY FOR METADATA MANAGEMENT AND GOVERNANCE	11
3.	COVERAGE OF STAKEHOLDER REQUIREMENTS BY RDC/REDDA.....	13
3.1.	STAKEHOLDERS REQUESTS AND NEEDS	13
3.2.	GAP ANALYSIS OF DG MARE’S REQUIREMENTS AND REDDA.....	15
4.	CONCLUSION AND NEXT STEPS	24
5.	WORKS CITED	27
ANNEX I.	ANALYSIS OF RDC.....	28
I.1.	TESTING AND DEPLOYING RDC	30
ANNEX II.	REFERENCE DATA MANAGEMENT AND GOVERNANCE REQUIREMENTS – DG MARE.....	31
ANNEX III.	TOOLS FOR METADATA MANAGEMENT	33
III.1.	POOLPARTY SEMANTIC SUITE	33
III.2.	FAO VOCBENCH	33
III.3.	POOLPARTY & FAO VOCBENCH: AN ALTERNATIVE TO REDDA	34
III.4.	R3GISTRY	36
ANNEX IV.	PROPOSED CHANGE REQUESTS FOR THE ADAPTATION OF REDDA	42
IV.1.	MAKE REDDA REUSABLE BY THIRD PARTIES AS OPEN-SOURCE SOFTWARE	42
IV.2.	MAKE REDDA AVAILABLE AS A SERVICE TO COMMISSION SERVICES	43
IV.3.	SUPPORT STANDARD FORMATS FOR IMPORT AND EXPORT	44
IV.4.	REFERENCE DATA PUBLICATION ON THE RE3GISTRY	45
IV.5.	REFERENCE DATA PUBLICATION ON WEBDAV / SVN / GIT	45
IV.6.	USER AND ACCESS RIGHTS	46

List of Tables

TABLE 1: GLOSSARY	9
TABLE 2: MOSCOW	13
TABLE 3: DG MARE’S REQUIREMENTS FOR REFERENCE DATA MANAGEMENT TOOLS.....	14
TABLE 4: DIMENSIONS FOR COMPARING REQUIREMENTS.....	16
TABLE 5: COVERAGE OF REQUIREMENTS BY REDDA	17
TABLE 6: ASSESSMENT OF ALTERNATIVES FOR REFERENCE DATA MANAGEMENT TOOLS BY DG MARE	25
TABLE 7: REFERENCE DATA MANAGEMENT REQUIREMENTS	31
TABLE 8 - COVERAGE OF POOLPARTY, FAO VOCBENCH AND REDDA.....	34
TABLE 9: DIMENSIONS FOR COMPARING REQUIREMENTS.....	36
TABLE 10: COVERAGE OF REQUIREMENTS BY RE3GISTRY	37

List of Figures

FIGURE 1: DATA FLOW BETWEEN AUTHENTIC SOURCE, REDDA AND AN OPERATIONAL SYSTEM	7
FIGURE 2: METADATA MANAGEMENT LIFECYCLE [ISA_D42].....	11
FIGURE 3: GENERIC TOOL CHAIN FOR METADATA MANAGEMENT	11
FIGURE 4: PROPOSED TOOL CHAIN FOR METADATA MANAGEMENT	12
FIGURE 5: EXISTING AND PLANNED FEATURES OF RDC	28

1. INTRODUCTION

This report is commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission, in the context of its Action 1.1 on semantic interoperability¹. It reports on the work performed to coordinate the adaptation to the needs of the Directorate-General for Maritime Affairs and Fisheries (DG MARE) of the generic **Reference Data Component (RDC)**, which will evolve to **REDDA**: the **Reference Data Deployment Adaptor**. REDDA is currently being developed by DG COMP with the intention to be also used by other EU Institutions. In order to ensure reusability, the tool solution should take into account the requirements of other stakeholders. In the scope of this report, we studied the reference data management requirements of DG MARE, for the marine and maritime domain, and assessed them against the functionalities to be provided by REDDA.

1.1. Context

In 2014, the ISA Programme commissioned a study to develop specifications for the governance and management of reference data based on the needs of DG COMP [Eur14_D4_3].

The Reference Data Component (RDC) was developed by DG COMP and the ISA Programme to support the management of reference data as a building block integrated into the Generic Interoperable Notification Services (GENIS) system. The purpose of GENIS is to allow Member States to notify the European Commission of planned State aid cases, e.g. when an entity receives advantages from a public authority².

Based on the outcomes of the previous study [Eur14_D4_3], the ISA Programme and DG COMP initiated an update process for RDC to improve its reusability. Up to now, the RDC has not reached the required levels for the management and governance of reference data. The RDC has been developed to address the specific requirements of State Aid control systems; therefore the building block is not easily interoperable and reusable.

In order to deal with the shortcomings of the RDC, a new development has been launched, called Reference Data Deployment Adaptor (REDDA). REDDA should evolve into a standardised, reliable and reusable solution to manage and govern reference data. Within REDDA, DG COMP will extend the features of the RDC in order to better comply with the methodology for metadata management and governance developed by the ISA Programme.

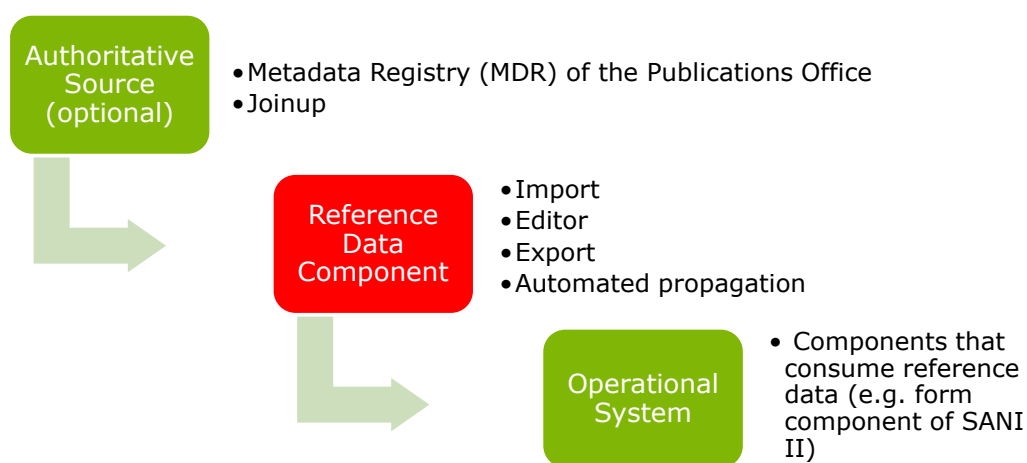
¹ Action 1.1 of the ISA Programme:
http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm

² State aid: http://ec.europa.eu/competition/state_aid/overview/index_en.html

As depicted in Figure 1, the REDDA will form the missing link between several authoritative sources of reference data and operational systems that consume that data. In essence, the REDDA will allow DG COMP to:

- import reference data from different sources, such as the Publications Office Metadata Registry and its Named Authority List for country codes;
- edit the imported reference data; and
- export it to a suitable format or automatically deploy the reference data into their operational systems.

Figure 1: Data flow between authentic source, REDDA and an Operational System



REDDA aims to be reused by several EU institutions for supporting the management and governance of reference data. The context in which REDDA could be implemented includes among others:

- **The marine and maritime domain** where DG MARE and the JRC need a cross-sector methodology and supporting tools for the management and governance of reference data in light of the Common Information Sharing Environment (CISE)³. CISE is currently being developed jointly by the European Commission and the EU/EEA Member States, including civilian and military authorities and agencies operating in the maritime field. CISE will integrate existing surveillance systems and networks and give all concerned authorities access to the information they need for their missions at sea;
- **The health and consumers domain** where DG SANCO (DG Health and Consumers) could use REDDA to maintain among others the taxonomies on Plant Protection Products Authorisations⁴.

³ The Common Information Sharing Environment – CISE:
http://ec.europa.eu/maritimeaffairs/policy/integrated_maritime_surveillance/index_en.htm

⁴ DG Health and Consumers, Plant Protection Products (PPP) Authorisations
http://ec.europa.eu/dgs/health_consumer/information_systems/index_en.htm#ppp
http://health.testproject.eu/PPP/ppp_taxonomies.html

- **The EU legislative decision making process** where the Publications Office of the EU could use REDDA for managing the Named Authority Lists and/or Eurovoc [Pub14].

1.2. Objective and scope

The objective of this pilot is to analyse to which extent REDDA could be tailored to support the needs of domains other than State Aid control and by other entities, such as in the marine and maritime domain managed by DG MARE.

The scope of this study includes the analysis of the extension of REDDA to meet the needs and requirements of DG MARE. We identify the missing functionalities and assess the feasibility of their implementation.

1.3. Approach

The following approach is followed in this pilot:

1. **Adaptation of the methodology**⁵ for reference data management and governance to the needs of DG MARE. The methodology shall include those **processes and tools relevant for DG MARE**, for instance those related to **information security management**. The adapted methodology shows that alternative technological solutions can support each phase of reference data management. The adapted methodology is described in Chapter 2.
2. **Gap analysis** between **DG MARE's tool requirements for reference data management** and the features to be implemented in **the generic REDDA solution**. As REDDA reuses and extends the core of RDC, the collection of functionalities of REDDA was derived after compiling input from the following sources: the documentation of the existing RDC solution, bilateral communications with DG COMP and the previous study referred to earlier [Eur14_D4_3]. The findings of our analysis were validated by DG MARE in the context of a validation workshop that took place on 1 September 2014. The coverage of the requirements is discussed in section 3.1.
3. **Coordination of the adaptation of REDDA and provision of recommendations** summarised in section 4.

⁵ ISA Program May 2014, D4.2 'Methodology and tools for Metadata Governance and Management for EU Institutions'

1.4. Glossary

Table 1 provides an overview of the terms and abbreviations which are commonly used throughout the report.

Table 1: Glossary

Term	Description
CISE	The Common Information Sharing Environment for the Marine and Maritime domain.
Code list	Complete set of data element values of a coded simple data element [ISO 9735-1:2002, 4.14].
Data model	A data model is a collection of entities, their properties and the relationships among them, which aims at formally representing a domain, a concept or a real-world thing.
DG COMP	Directorate-General for Competition
DG MARE	Directorate-General for Maritime Affairs and Fisheries
FAO	Food and Agricultural Organisation of the United Nations
GENIS	The Generic Interoperable Notification Services (GENIS) Information System is used to manage and support the exchange of information between Member States and the Commission within the State Aid Notification Process, where Member States notify the European Commission of planned State aid. GENIS is also known as the State Aid Notification Interactive (SANI-2), and is the successor to the existing SANI.
Interoperability	According to the ISA Decision, interoperability means the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the exchange of data between their respective ICT systems.
Metadata	Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information. (National Information Standards Organization, 2004)
Metadata alignment	Metadata alignment is the harmonisation of structural metadata either by forging a wide consensus on the use of a common specification for structural metadata or through the creation of mappings between terms of two or more specifications.
Metadata governance	Metadata governance comprises well-defined roles and responsibilities, cohesive policies and principles, and decision-making processes that define, govern and regulate metadata.
Metadata management	Metadata management is defined as the good practice of putting in place people, processes, and systems to plan, perform, evaluate, and improve the lifecycle of metadata.
Named Authority List	Controlled vocabulary for use in naming particular entities consistently [ISO/DIS 25964-2].
RDC	Reference Data Component is a component already integrated into the Generic Interoperable Notification Services (GENIS) platform.
REDDA	Reference Data Deployment Adaptor In August 2014, DG COMP introduced REDDA as the new development based on RDC.
Reference data	Reference data is small, discrete sets of values that are not updated as part of business transactions but are usually used to impose consistent classification. Reference data normally has a low update

Term	Description
	frequency. Reference data is relevant across more than one business systems belonging to different organisations and sectors.
RFC	Request For Change A form used to record details of a request for a change and is sent as an input to change management by the change requestor.
SKOS	Simple Knowledge Organization System – RDF Vocabulary for the representation of key reference data such as code lists, and taxonomies.
Structural metadata	Data model or reference data
Taxonomy	Scheme of categories and subcategories that can be used to sort and otherwise organize items of knowledge or information [ISO/DIS 25964-2].
Thesaurus	Controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms [ISO 25964-1:2011].

2. ADAPTATION OF THE METHODOLOGY FOR METADATA MANAGEMENT AND GOVERNANCE

In context of Action 1.1 of the ISA programme, the European Commission developed a **generic methodology for the management and governance of structural metadata** (see Figure 2) [ISA_D42] [ISA_D44]. This chapter adapts this methodology to the reference data management requirements of DG MARE, focusing on the tools to be used for managing reference data.

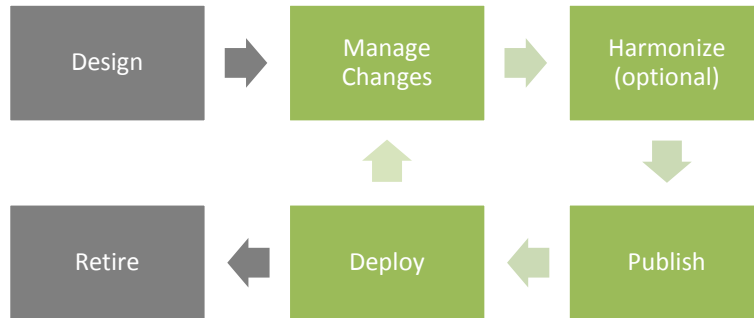
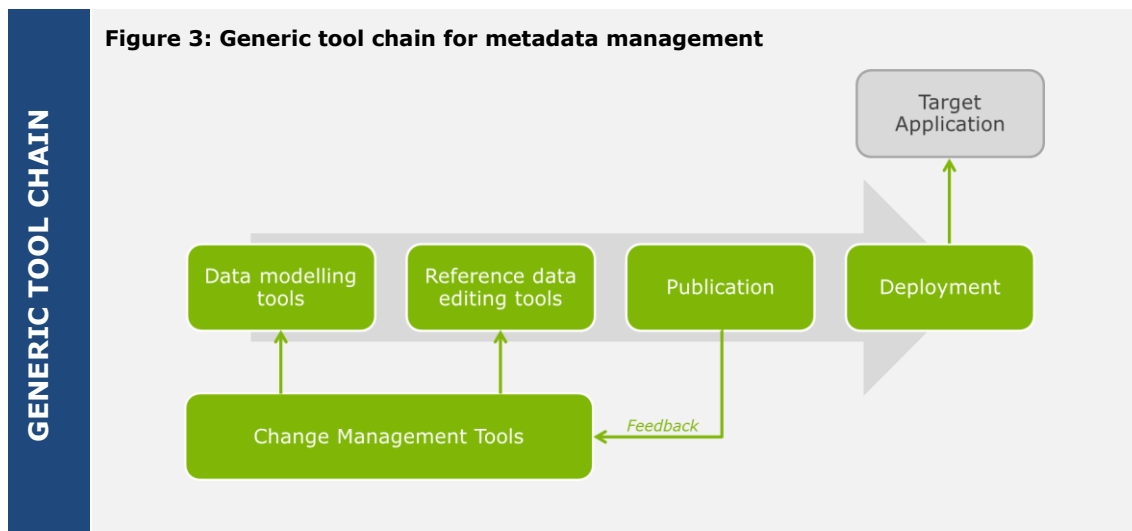
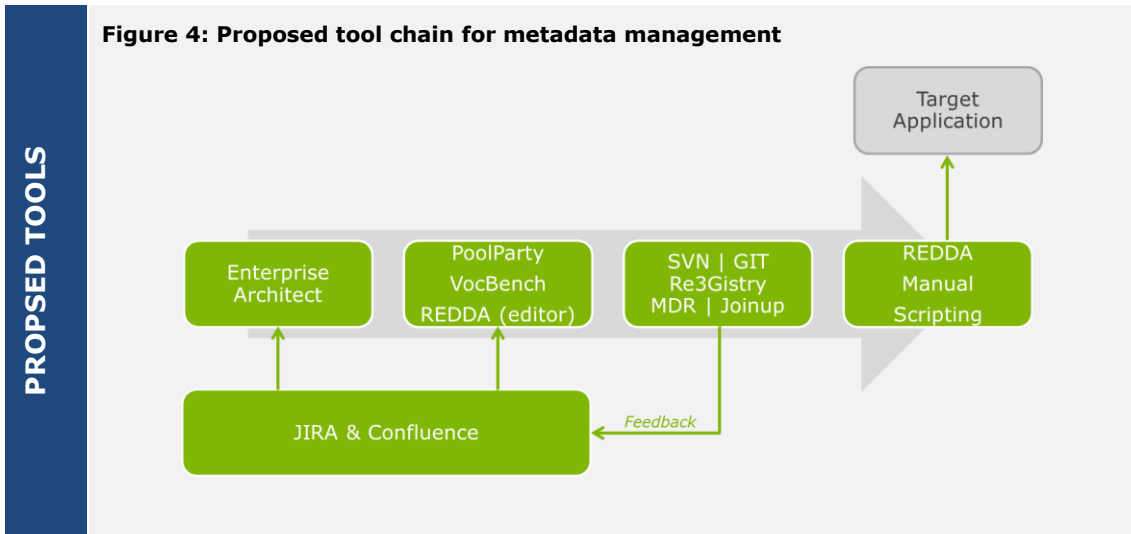


Figure 2: Metadata Management Lifecycle [ISA_D42]

The generic methodology for metadata management and governance described also a chain of tools which could be used to support the management processes [ISA_D42]. The generic tool chain depicted in Figure 3 indicates the types of tools which are needed to manage the lifecycle of structural metadata, i.e. covering both data models and reference data. A number of alternative tools can be used in each of the categories, as displayed in Figure 4. The different tools are described in Annex III.





3. COVERAGE OF STAKEHOLDER REQUIREMENTS BY RDC/REDDA

This chapter analyses the coverage and provides a gap analysis between the features of REDDA and the reference data management requirements of DG MARE. As REDDA reuses and extends the core of RDC, the collection of functionalities of REDDA was derived after compiling input from the following sources: the documentation of the existing RDC solution, bilateral communications with DG COMP and the previous study referred to earlier [Eur14_D4_3].

The functionalities of RDC as it is currently used by DG COMP in the State Aid Notification system are detailed in Annex I. Section 3.1 gives an overview of the stakeholder requests regarding the management, governance and tools to support reference data management. In section 3.2, the stakeholder requests are mapped to the features of REDDA, where possible and as foreseen at the time of the analysis. The identified gaps lead to the creation of proposed change requests for REDDA, which are documented in Annex III.4 below. These change requests will enable to close the gaps and to provide a new solution which will meet the needs of DG MARE.

The findings of our analysis were validated by DG MARE in the context of a validation workshop that took place on 1 September 2014.

3.1. Stakeholders requests and needs

This section describes the reference data management tool requirements for the marine and maritime domain expressed by DG MARE and the Joint Research Centre (JRC).

These requirements were elicited in light of a previous study on the management of structural metadata for the marine and maritime domain [ISA_D44]. In the context of reference data management, governance and tools, these requirements have been validated by the stakeholders, i.e. DG MARE and the JRC.

Given the scope of this pilot, we are focusing on the requirements for tools that can support the management of reference data (see Table 3).

We used the following dimensions for assessing those requirements:

Table 2: MoSCoW

MoSCoW	
Must	Describes a requirement that must be satisfied in the final solution for the solution to be considered a success
Should	Represents a high-priority item that should be included in the solution if it is possible.
Could	Describes a requirement which is considered desirable but not necessary.
Would	Represents a requirement that may be considered for the future

Table 3: DG MARE's requirements for reference data management tools

ID	Domain	Requirement	MoSCoW
TR1.	Workflow	Allow a user who is responsible of a workflow task to delegate it to another user.	Could
TR2.	Workflow	Ability to define workflows in the solution.	Could
TR3.	Workflow	Provide predefined circuits of validation to the application.	Could
TR4.	Workflow	Provide the necessary validation process (workflow) to ensure prevention of mistaken or unauthorised changes done by the team responsible to manage the Reference Data (create, update, and delete).	Must
TR5.	Tools integration	Allow integration with a ticketing system, e.g. JIRA.	Could
TR6.	Tools integration	Support integration with other tools used during the reference data design, maintenance and documentation phases.	Could
TR7.	Search	Enable users to search data by using several ways, e.g. free text, filter, tags, and metadata.	Must
TR8.	Search	Provide an option to search for a specific version of a reference dataset; by default search will be done on the last version.	Could
TR9.	Repository	Offer a publicly available repository of re-usable reference data – for those parts of reference data that are not subject to privacy/confidentiality constraints	Should
TR10.	Reference data import	Enable static import of reference data in standard format, especially supporting UN/CEFACT code lists.	Must
TR11.	Reference data format	Manage different reference data formats, both machine-readable and human-readable.	Must
TR12.	Reference data export	Enable export of reference data in a standard format,	Must
TR13.	Packaging	Provide a package allowing the easy deployment of the selected reference data management tool.	Must

ID	Domain	Requirement	MoSCoW
TR14.	Navigation	Allow users to navigate visually (tree, table, etc.) through data elements in a graphical way.	Could
TR15.	Multilingualism	Provide a mechanism to manage multilingualism for reference data.	Must
TR16.	Log	Provide a log of actions performed on reference data.	Should
TR17.	Log	Provide a log of actions performed by a user.	Must
TR18.	Licensing	Release the reference data component with an open source licence (or any other appropriate licence of choice).	Must
TR19.	Documentation	Provide complete documentation and guidelines for IT and business users.	Should
TR20.	Data Dictionary	Provide public HTTP URIs for the different reference data entities.	Should
TR21.	Collaboration	Allow the user to work in a collaborative way.	Could
TR22.	Classification	Provide to the user a way to classify the concept schema, concept or terms created in the reference data with a taxonomy mechanism.	Could
TR23.	Security	Provide a user interface for manage users, enabling the administrators to create, update or delete users and roles, as well as the possibility to assign roles to a user.	Should
TR24.	Security	Provide a mean to limit the access rights of the reference data. Manage access rights for different parts of the reference data (limited access).	Must
TR25.	Security	Limit the access to reference data according to its current status.	Must

3.2. Gap analysis of DG MARE's requirements and REDDA

This section contains a comparison of the needs and requirements of DG MARE against the functionalities of REDDA. Based on the findings, it will be decided whether it is feasible to tailor REDDA to the specific requirements of DG MARE.

Table 5 lists the requirements of DG MARE and compares them based on two dimensions:

- MoSCoW: the level of importance placed on the requirement
- Costs/benefits: an estimate of the efforts compared to the benefits

Table 4: Dimensions for comparing requirements

Dimension 1: MoSCoW	
Must	Describes a requirement that must be satisfied in the final solution for the solution to be considered a success
Should	Represents a high-priority item that should be included in the solution if it is possible.
Could	Describes a requirement which is considered desirable but not necessary.
Would	Represents a requirement that may be considered for the future
Cost benefits /	Implementation timeline

Table 5: Coverage of requirements by REDDA

Domain	Requirement	MoSCoW	Coverage REDDA	Observations REDDA
Workflow	Allow a user who is responsible of a workflow task to delegate it to another user.	Could	No	This requirement is currently not foreseen in REDDA.
Workflow	Ability to define workflows in the solution.	Could	No	This requirement is currently not foreseen in REDDA.
Workflow	Provide predefined circuits of validation to the application.	Could	No	This requirement is currently not foreseen in REDDA.
Workflow	Provide the necessary validation process (workflow) to ensure prevention of mistaken or unauthorised changes done by the team responsible to manage the Reference Data (create, update, and delete).	Could	No	This requirement is currently not foreseen in REDDA.
Tools integration	Allow integration with a ticketing system, e.g. JIRA.	Could	No	This requirement is currently not foreseen in REDDA. Should this option be developed, it would be possible to manage the workflow with JIRA for instance instead of developing a new functionality.

Domain	Requirement	MoSCoW	Coverage REDDA	Observations REDDA
Tools integration	Support integration with other tools used during the reference data design, maintenance and documentation phases.	Could	Part.	It is possible to import and export reference data in REDDA. Therefore a workaround could be to export reference data maintain in other tools and to import it in REDDA. Nevertheless an exchange format has to be chosen.
Search	Enable users to search data by using several ways, e.g. free text, filter, tags, and metadata.	Must	Yes	REDDA will provide a generic function to search reference data. Depending on the user experience, it could be necessary to plan further development.
Search	Provide an option to search for a specific version of a reference dataset; by default search will be done on the last version.	Could	No	A search function exists but could be improved in order to give a better end-user experience.
Repository	Offer a publicly available repository of re-usable reference data – for those parts of reference data that are not subject to privacy/confidentiality constraints.	Should	No	This requirement is currently not foreseen in REDDA.

Domain	Requirement	MoSCoW	Coverage REDDA	Observations REDDA
Reference data import	Enable static import of reference data in standard format, especially supporting UN/CEFACT code lists.	Must	Part.	REDDA enables import of reference data in XML with a format defined by DG COMP. Therefore, standard formats have to be implemented.
Reference data format	Manage different reference data formats, both machine-readable and human-readable.	Must	Part.	Only CSV and XML are currently supported. SKOS support is foreseen.
Reference data export	Enable export of reference data in a standard format.	Must	Part.	The RDC currently exports reference data in XML with a format defined by DG COMP. Therefore, standard formats will have to be supported by REDDA.
Packaging	Provide a package allowing the easy deployment of the selected reference data management tool.	Must	Part.	A package has been released on the Joinup platform. The re-use of the tool is not easy as there is not complete documentation explaining the main technical requirements.
Navigation	Allow users to navigate visually (tree, table, etc.) through data elements in a graphical way.	Should	No	This requirement is not foreseen in REDDA. In light of this pilot, this requirement should not be implemented.

Domain	Requirement	MoSCoW	Coverage REDDA	Observations REDDA
Multilingualism	Provide a mechanism to manage multilingualism for reference data.	Must	Yes	The requirement is fully implemented in REDDA and can be used as is.
Log	Provide a log of actions performed on reference data.	Should	Part.	A log database exists in REDDA. It is not necessary to extend the current functionalities in the context of this pilot.
Log	Provide a log of actions performed by a user.	Must	Part.	A log database exists in REDDA. It is not necessary to extend the current functionalities in the context of this pilot.
Licensing	Release the reference data component with an open source licence (or any other appropriate licence of choice).	Must	No	The licensing model has to be chosen.
Documentation	Provide complete documentation and guidelines for IT and business users.	Should	No	The package is provided as is without any documentation.
Data Dictionary	Provide public HTTP URIs for the different data entities.	Should	No	This requirement is not foreseen in REDDA.
Collaboration	Allow the user to work in a collaborative way.	Could	No	This requirement is currently not foreseen in REDDA. In the context of the pilot, this is not necessary to implement.

Domain	Requirement	MoSCoW	Coverage REDDA	Observations REDDA
				Nevertheless, this requirement should be considered in the next steps.
Data Dictionary	Provide public HTTP URIs for the different reference data entities.	Should	No	This requirement is currently not foreseen in REDDA.
Classification	Provide to the user a way to classify the concept schema, concept or terms created in the reference data with a taxonomy mechanism	Could	Yes	Tagging mechanism is already available in the system. This functionality enables flexibility but needs to be managed accordingly (governance).
Security	Provide a user interface for manage users, enabling the administrators to create, update or delete users and roles, as well as the possibility to assign roles to a user.	Should	Part.	A mechanism exists but it does not allow to dynamically managing user roles.
Security	Provide a mean to limit the access rights of the reference data. Manage access rights for different parts of the reference data (limited access).	Must	No	A workaround is possible now but it could be interesting to improve the function in the future REDDA solution.

D04.03 Report on the tailoring of the pilot to DG MARE needs

Domain	Requirement	MoSCoW	Coverage REDDA	Observations REDDA
Security	Limit the access to reference data according to its current status.	Must	No	A workaround is possible now but it could be interesting to improve the function in the future REDDA solution.

Based on the analysis of Table 5, the following gaps were identified (***Must or Should requirements which are not or partially implemented***):

- **Repository:** Offer a publicly available repository of re-usable reference data – for those parts of reference data that are not subject to privacy/confidentiality constraints;
- **Reference data import:** Enable static import of reference data in standard format, especially supporting UN/CEFACT code lists;
- **Reference data format:** Manage different reference data formats, both machine-readable and human-readable;
- **Reference data export:** Enable export of reference data in a standard format;
- **Packaging:** Provide a package allowing re-using easily the tool developed to fulfil these requirements;
- **Navigation:** Allow users to navigate visually (tree, table, etc.) through data elements in a graphical way;
- **Log:** Provide a log of actions performed on reference data;
- **Log:** Provide a log of actions performed by a user;
- **Licensing:** Release the reference data component with an open source licence (or any other appropriate licence of choice);
- **Documentation:** Provide complete documentation and guidelines for IT and business users;
- **Data Dictionary:** Provide public HTTP URIs for the different reference data entities;
- **Security:** Provide a user interface for manage users, enabling the administrators to create, update or delete users and roles, as well as the possibility to assign roles to a user;
- **Security:** Provide a user interface for manage users, enabling the administrators to create, update or delete users and roles, as well as the possibility to assign roles to a user; and
- **Security:** Limit the access to reference data according to its current status.

4. CONCLUSION AND NEXT STEPS

Based on the gap analysis of section 3.2, **REDDA only covers a part of the tool requirements of DG MARE.** Indeed, REDDA is based on the RDC which has originally been developed to fit the needs of DG COMP in the context of the State Aid Control. Extending those functionalities to cover the needs of an additional domain is not trivial.

The priorities of the different domains when it comes to reference data management also impact the further development of REDDA and its reusability in different contexts. For example, the **key use case of REDDA is the automatic deployment of reference data in production systems.** This is a very important feature for DG COMP in the context of State Aid Control and significant effort is invested there. However, **this feature is currently not required by DG MARE** for managing marine and maritime reference data.

For DG MARE, there are two aspects that are especially relevant (high-priority features):

- a. **Managing access to reference data:** It is foreseen that some parts of the CISE consolidated model will not be publicly available. Therefore, the methodology and the generic tool should consider users with different access rights to information of different sensitiveness level.
- b. **Domain users:** The software tool will be used by experts in several domains (maritime surveillance, authority lists, health, etc.). Therefore, the tool shall be generic enough to be used within any of those domains.

Additionally, to successfully extend the functionalities of REDDA to meet the requirements from DG MARE, the following issues should be tackled:

- REDDA needs a better user manual and documentation;
- REDDA needs to comply with standards for representing reference data; and
- REDDA needs to have a governance or business model guaranteeing its long term sustainability.

As a result, DG COMP has to take into account the previous issues in the development of REDDA to enable DG MARE, and possibly others, to **start using REDAA as soon as it will be released.**

In order to close the identified gaps, we have formed a number of proposed change requests to REDDA, which are documented in Annex IV. It is up to the Commission to perform a cost-benefit analysis to assess the feasibility of implementing any of those change requests. Alternatively, DG MARE could investigate the use and possible adaptation of other open-source tools available on the market, such as the FAO VocBench and the Re3gistry (see Annex III). The table below summarises the pros and cons for these two alternatives.

Table 6: Assessment of alternatives for reference data management tools by DG MARE

Alternative	Pros	Cons
Implement the changes requests to align REDDA with the requirements of DG MARE	<ul style="list-style-type: none"> • Capitalise on the effort produced by DG COMP to implement a reference data component; • Provide a generic component to other DGs. 	<ul style="list-style-type: none"> • Implement a large number of change requests to fulfil the requirements of DG MARE; • Develop the necessary change requests to meet the needs of the DG MARE will take a longer time.
Re-use open source tools available on the market	<ul style="list-style-type: none"> • Cover a large part of the requirements of DG MARE; • Capitalise on the experience of the Publications Office of the EU which implemented this solution. 	<ul style="list-style-type: none"> • Budget new investment to implement the solution; • Require specific development to meet some requirements of the DG MARE.

Regardless of the alternative chosen, the following actions should be undertaken by DG COMP to ensure the reusability and sustainability of REDDA:

- **Identify stakeholders:** As REDDA is a currently in development, DG COMP should identify stakeholders interested in reusing the solution;
- **Involve the stakeholders:** To allow REDDA to be reusable by third parties, it is necessary to involve the end users in the development process of the solution, e.g. by collecting their requirements and by allowing them to test early prototypes. This will increase the reusability of the solution and will facilitate the buy-in of the stakeholders;
- **Manage the communication with the community:** In order to enable reusability of one of the two solution, it is necessary to define a communication plan which enables third parties to be informed of the next steps which will done for the new solution.
- **Organise a Proof Of Concept with REDDA and get feedback from the end-users:** this task should enable to identify the pros and cons of the stakeholders who participate to the Proof of Concept;
- **Analyse the result of the feedbacks:** Based on the feedbacks of the stakeholders, this action should allow identifying the next major functionalities or improvements to implement; and
- **Review of the new functionalities/improvements and prioritisation:** Based on the previous analysis, the task would detail the functionalities/improvement implementation in terms of planning, resources, costs, etc., and based on that it should be possible to define priorities.

5. WORKS CITED

- [Eur14_D4_3] European Commission - ISA Programme, "Report on the implementation of a Metadata Management pilot for DG COMP," 2014.
- [Pub14] Publications Office. (2014) Publications Office. [Online].
http://publications.europa.eu/index_en.htm
- [ISA_D44] European Commission - ISA Programme, "D4.4 Report on implementation of a Metadata Management pilot for the Common Information Sharing Environment (CISE)," European Commission, 2014.
- [Nat04] National Information Standards Organization , "Understanding Metadata," 2004.
- [COMP14] European Commission - DG COMP, "Pilot for reference data management - Brochure," 2014.
- [ISA_D42] European Commission - ISA Programme, "Methodology and tools for Metadata Governance and Management for EU Institutions and Member States," 2014.
- [JRC14] European Commission - Joint Research Centre, "Re3gistry software v0.3," 2014.
- [ISA_D45] European Commission - ISA Programme, "D4.5. Metadata alignment pilot in the EU institutions and Member States," 2014.
- [JRC13] R. S. Smith, M. Lutz, A. Perego, and R. Sgaolin, "Building a missing item in INSPIRE: The Re3Gistry," 2013.

Annex I. ANALYSIS OF RDC

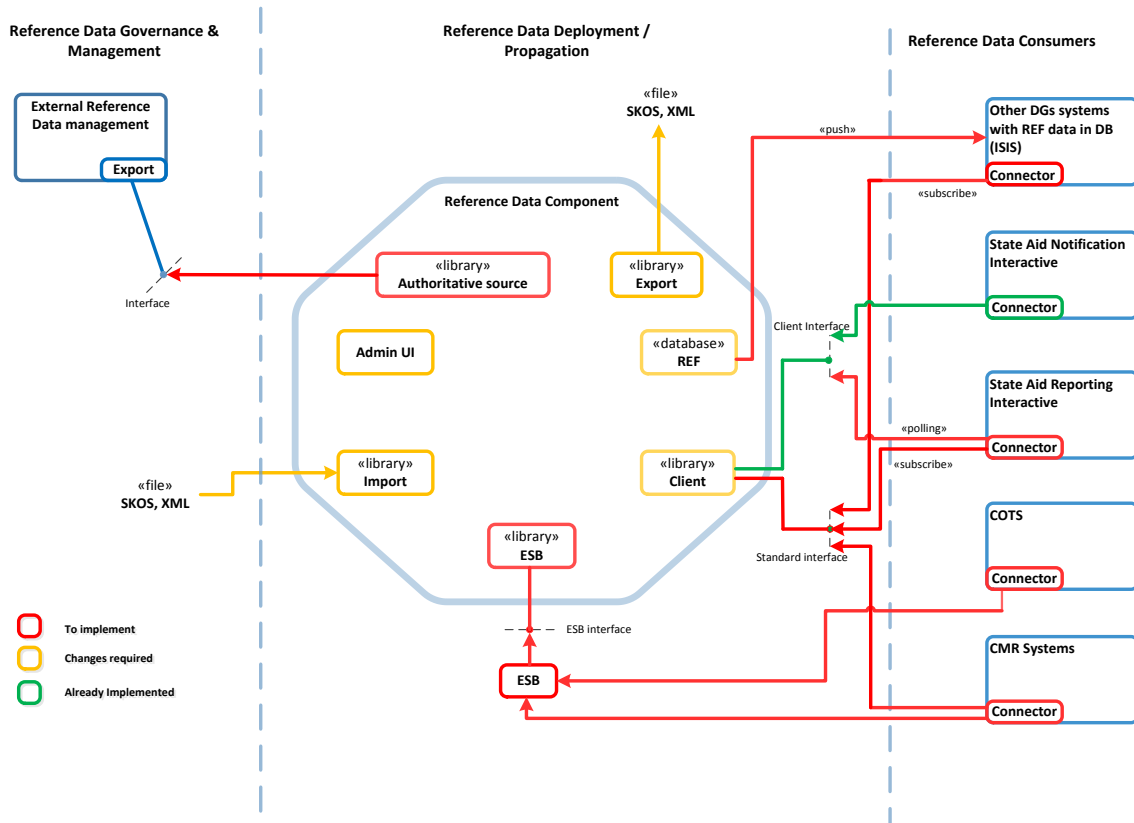
Based on the documentation of the existing RDC solution, bilateral communications with DG COMP and the previous study referred to earlier [Eur14_D4_3], we managed to analyse the RDC building block and by extension REDDA.

RDC currently supports the following features:

- **Create, read, update, delete** reference data using the Web-based graphical interface;
- **Import** reference data from an XML file; and
- **Export** reference data to an XML file.

Figure 5 provides a schematic overview of the existing and planned features of the tool. Emphasis is placed on the export of reference data in standardised formats, such as SKOS and XML.

Figure 5: Existing and planned features of RDC



The metadata governance process currently adopted by DG COMP for the RDC involves three roles:

- **Administrator:** Users with this role can create and delete projects, groups, project managers, normal user;
- **Project Manager:** The Project Manager can create Standard Users and give them access to projects; and

- **Standard user:** a Standard User will be assigned to one or more projects and, when logged in will have access to one or more projects.

According to DG COMP, RDC supports several main features for managing reference data, metadata and enterprise master data:

- **Architecture:** The software is designed in a way that allows it to run as a single instance on a server, while serving multiple client organisations. The Component categorises reference data in projects, to which users and managers are assigned;
- **Graph Data:** The domain model of the reference data in the tool includes Project, Group, Reference Data Item, Representation, Order and Tag entities. Ownership, lifecycle, protection and data segments are defined for each entity;
- **Versioning:** Versioning is carried out at the Group Entity and at the Reference Data Entity level. These entities can get project-specific start and end dates assigned;
- **Data staging:** Import and export capabilities in XML are supported. Import and export is script driven, so it can be adapted to the specifications of different systems.
- **Decoupling:** Clients of the Reference Data Management Component operate entirely on locally cached data. By decoupling the tool from its clients, an outage of the Component does not lead to an interruption in the client's system;
- **Multilingualism:** each reference data item can have labels in any language;
- **Deployment:** The Component can be deployed as a service, integrated into an application as a standalone application;
- **Historical data:** Clients can consume reference data according to a timestamp. By doing so, RDC allows to serve reference data as it was available at any point in time in the past.

RDC is intended for reference data and master data. RDC allows for a clear distinction between reference data or master data and the business logic which will be in the application layer. To this extent, RDC can be re-used by other systems as a plugin, via web-services, via API, or using a dedicated client. It is currently designed for use within DG COMP and would need further work before it can be made available as a generic solution for interoperability.

The RDC editor feature structures Reference Data as follows:

- **Project:** Reference data is categorised in projects. The SANI-2⁶ project for example contains the reference data which is linked to the State Aid Notification Infrastructure.
- **Group:** projects have zero, one or more groups. Groups represent concept schemes, for example a country code list.
- **Reference data entity:** each group consists of reference data entities. By defining different start and end dates to a reference data entity in different projects, each system will be able to access the version which is relevant. For example, the HR system might need Serbia as part of its reference data while DG COMP might not yet need it in its system.
- **Representation:** reference data entities can have one or more representations (e.g. alpha-3 and alpha-2 codes for countries).
- **Ordering:** groups can have one or more orderings for the reference data entities included in it.
- **Classification:** reference data can be described with tags which enable to found an item based on the tags describing it.

I.1. Testing and deploying RDC

DG COMP provided access to the testing environment of RDC. This enabled us to analyse the tool from a functional point of view.

As a package of RDC was released in August 2014 on the Joinup platform, this enabled us to try out the deployment of the solution in an environment which differs from the one of DG COMP.

The objectives of this deployment were to:

- Identify the difficulties which could be encountered by third parties; and
- Provide recommendations to DG COMP to improve the reusability of the RDC solution through its evolution to REDDA.

The package contains a war file, therefore we tried to deploy it on a Tomcat server, the package was deployed but the tool was not working because of library dependency. The logs generated by the application server were too generic to find a solution and to continue further and complete the deployment.

Here are the main issues which could be tackled:

- The package is provided as is without any explanation about the technical requirements (application server, database, framework, etc.); and
- There is no documentation to explain how to deploy the package and how to use the tools.

⁶ SANI-2: State Aid Notification Interactive 2, also referred to as GENIS.

Annex II. REFERENCE DATA MANAGEMENT AND GOVERNANCE REQUIREMENTS – DG MARE

The table below describes the reference data management requirements expressed by DG MARE. These requirements were collected in the context of [ISA_D44] and where validated in the context of this pilot. These requirements form the basis for eliciting the reference data management tool requirements listed in section 3.

Table 7: Reference data management requirements

ID	Domain	Requirement
R1.	Documentation, publishing and licensing	Marine and maritime reference data needs to be described following unambiguous guidelines using a vocabulary based on standard format, such as SKOS.
R2.	Documentation, publishing and licensing	The documentation of the reference data should provide all the necessary elements (e.g. guidelines, tutorials, tools) to public administrations to map their legacy reference dataset to standardised marine and maritime reference data.
R3.	Documentation, publishing and licensing	Every version of the reference data should be tracked enabling an access to all the version implemented and should be easily accessible for all stakeholders.
R4.	Documentation, publishing and licensing	Reference data should be available in both human- and machine-readable formats.
R5.	Documentation, publishing and licensing	Reference data should be persistent. Older versions should be deprecated but not deleted.
R6.	Documentation, publishing and licensing	Reference data management should be decoupled from the data model management. That means a separate lifecycle, a separate versioning, etc., as reference data changes frequently than data models.
R7.	Re-use of standards	The lifecycle of the reference data should be aligned with the standardisation process of different standardisation bodies, such as UN/CEFACT, especially in cases when reference data from these bodies are re-used.
R8.	Access	Some parts of marine and maritime reference data could describe sensitive, confidential and/or private information. Therefore access to this reference data should be restricted.

ID	Domain	Requirement
R9.	Timeliness	For the reference data to remain relevant, it should allow for continuous improvement. Adaptations to the needs of the users should be delivered timely.
R10.	Timeliness	The governance structure should foster the agility of the maintenance and evolution of the reference data, allowing requirements of the users to be implemented timely. Therefore, a flat and lean governance structure, involving only a few hierarchical levels, should be put in place.
R11.	Enforcement	Support the next generation of IT systems and regulation. Existing IT systems should not be asked to change their data models or modus operandi. In cases when change is required, it is recommended to pursue exports to the common data models. New IT systems as well as new regulation could support natively (or take into account) the common data models.

Annex III. TOOLS FOR METADATA MANAGEMENT

III.1. PoolParty Semantic Suite

PoolParty⁷ Semantic Suite is a web-based platform to manage enterprise metadata and linked data based on semantic knowledge models (taxonomies, thesauri, ontologies and knowledge graphs). Metadata management is based on W3C's Semantic Web standards RDF and SKOS. PoolParty's APIs (based on W3C's SPARQL standard) allow integration of semantic technologies with other systems like search engines, CMS, etc.

PoolParty has the following characteristics:

- It enables a set of functionalities that allow a knowledge team to work in an efficient manner:
 - Support for Open Web Standards like RDF and SKOS;
 - Support for collaborative editing;
 - Support for editorial workflow, with user roles;
 - Support for multilingual terminology.

III.2. FAO Vocbench

VocBench⁸ is a web-based, multilingual editing and workflow tool that manages thesauri, authority lists and glossaries using SKOS standard. The tool has been developed by the Food and Agriculture Organisation (FAO) of the United Nations. VocBench is designed to meet the needs of semantic web and linked data environments. VocBench enables collaborative editing thanks to administration and group management features.

VocBench has the following characteristics:

- It is an application with a set of functionalities which allows a knowledge team to work in an efficient manner:
 - Support the Open Web Standards like RDF, OWL and SKOS;
 - Support collaborative editing;
 - Support editorial workflow, with user roles;
 - Support of multilingual terminology.
- The community is growing and includes several institutions;
- The development was led by FAO and should meet requirements for a wide range of organisations or companies;
- The plugin mechanism allows add-in functionality without disturbing the core system;
- Development can be shared with the community;
- It is made available under a GPL licence.

⁷ PoolParty Semantic Suite: <http://www.poolparty.biz/>

⁸ VocBench 1.3.1: <http://aims.fao.org/tools/VocBench-2>

III.3. PoolParty & FAO Vocbench: an alternative to REDDA

PoolParty and FAO Vocbench provide the same level of functionalities and the choice of one of these two solutions require an analysis.

Nevertheless the two solutions enable to cover partially the needs of DG MARE as described in the table below:

Table 8 - Coverage of PoolParty, FAO Vocbench and REDDA

Domain	Requirement	Coverage REDDA	Coverage PoolParty	Coverage FAO Vocbench
Workflow	Allow a user who is responsible of a workflow task to delegate it to another user.	No	Yes	Yes
Workflow	Ability to define workflows in the solution.	No	No	No
Workflow	Provide predefined circuits of validation to the application.	No	Yes	Yes
Workflow	Provide the necessary validation process (workflow) to ensure prevention of mistaken or unauthorised changes done by the team responsible to manage the Reference Data (create, update, and delete).	No	Yes	Yes
Tools integration	Allow integration with a ticketing system, e.g. JIRA.	No	No	No
Tools integration	Support integration with other tools used during the reference data design, maintenance and documentation phases.	Part.	Part.	Part.
Search	Enable users to search data by using several ways, e.g. free text, filter, tags, and metadata.	Yes	Yes	Part.
Search	Provide an option to search for a specific version of a reference dataset; by default search will be done on the last version.	No	Yes	No
Repository	Offer a publicly available repository of re-usable reference data – for those parts of reference data that are not subject to privacy/confidentiality constraints	No	No	No

Domain	Requirement	Coverage REDDA	Coverage PoolParty	Coverage FAO Vocabench
Reference data import	Enable static import of reference data in standard format, especially supporting UN/CEFACT code lists.	Part .	Yes	Yes
Reference data format	Manage different reference data formats, both machine-readable and human-readable.	Part .	Yes	Yes
Reference data export	Enable export of reference data in a standard format,	Part .	Yes	Yes
Packaging	Provide a package allowing the easy deployment of the selected reference data management tool.	Yes	Yes	Yes
Navigation	Allow users to navigate visually (tree, table, etc.) through data elements in a graphical way.	No	Yes	Yes
Multilingualism	Provide a mechanism to manage multilingualism for reference data.	Yes	Yes	Yes
Log	Provide a log of actions performed on reference data.	Part .	Yes	Yes
Log	Provide a log of actions performed by a user.	Part .	Yes	Yes
Licensing	Release the reference data component with an open source licence (or any other appropriate licence of choice).	No	No	Yes
Documentation	Provide complete documentation and guidelines for IT and business users.	Part .	Yes	Part.
Data Dictionary	Provide public HTTP URIs for the different reference data entities.	No	No	No
Collaboration	Allow the user to work in a collaborative way.	No	Yes	Yes
Classification	Provide to the user a way to classify the concept schema, concept or terms created in the reference data with a taxonomy mechanism.	Yes	Yes	No
Security	Provide a user interface for manage users, enabling the administrators to create, update or delete users and roles, as well as the possibility to assign roles to a user.	Part .	Yes	Yes
Security	Provide a mean to limit the access rights of the reference data. Manage	No	Yes	Yes

Domain	Requirement	Coverage REDDA	Coverage PoolParty	Coverage FAO Vocabench
	access rights for different parts of the reference data (limited access).			
Security	Limit the access to reference data according to its current status.	No	Yes	Yes

III.4. R3gistry

This section contains a comparison of the needs and requirements of DG MARE with the functionalities currently implemented in Re3gistry.

Table 10 lists the requirements of DG MARE and compares them based on two dimensions:

- MoSCoW: the level of importance placed on the requirement

Table 9: Dimensions for comparing requirements

Dimension 1: MoSCow	
Must	Describes a requirement that must be satisfied in the final solution for the solution to be considered a success
Should	Represents a high-priority item that should be included in the solution if it is possible.
Could	Describes a requirement which is considered desirable but not necessary.
Would	Represents a requirement that may be considered for the future
Cost benefits /	Implementation timeline

Table 10: Coverage of requirements by Re3gistry

Domain	Requirement	MoSCoW	Coverage Re3Gistry	Observations Re3gistry
Workflow	Allow a user who is responsible of a workflow task to delegate it to another user.	Could	No	This requirement is not implemented in the Re3Gistry.
Workflow	Ability to define workflows in the solution.	Could	No	This requirement is not implemented in the Re3Gistry.
Workflow	Provide predefined circuits of validation to the application.	Could	No	This requirement is not implemented in the Re3Gistry.
Workflow	Provide the necessary validation process (workflow) to ensure prevention of mistaken or unauthorised changes done by the team responsible to manage the Reference Data (create, update, and delete).	Could	No	The workflow is implemented according to the ISO 19135 actions (addition/clarification/retirement/invalidations/supersession).
Tools integration	Allow integration with a ticketing system, e.g. JIRA.	Could	No	This requirement is not implemented in the Re3Gistry.
Tools integration	Support integration with other tools used during the reference data design, maintenance and documentation phases.	Could	Plan	It would be feasible to customise the import and export format in order to be used with external tools.
Search	Enable users to search data by using several ways, e.g. free text, filter, tags, and metadata.	Must	Yes	Free text, multilingual search is available. The filter, tags are planned for the next release.

Domain	Requirement	MoSCoW	Coverage Re3Gistry	Observations Re3Gistry
Search	Provide an option to search for a specific version of a reference dataset; by default search will be done on the last version.	Could	Part .	The search functionality works only on the last version.
Repository	Offer a publicly available repository of re-usable reference data – for those parts of reference data that are not subject to privacy/confidentiality constraints	Should	Yes	The INSPIRE Registry service and the registers are public and available.
Reference data import	Enable static import of reference data in standard format, especially supporting UN/CEFACT code lists.	Must	Yes	Import from CSV.
Reference data format	Manage different reference data formats, both machine-readable and human-readable.	Must	Yes	Several standards are supported, such as xml, html, json and rdf/skos
Reference data export	Enable export of reference data in a standard format,	Must	Yes	The Re3gistry can export the data in any standard/custom format. Examples: xml, html, json, atom, csv, rdf/skos.
Packaging	Provide a package allowing the easy deployment of the selected reference data management tool.	Must	Yes	The source code and a complete documentation are provided in the package.
Navigation	Allow users to navigate visually (tree, table, etc.) through data elements in a graphical way.	Should	Plan .	Using tables and links between

Domain	Requirement	MoSCoW	Coverage Re3Gistry	Observations Re3Gistry
				items in the web pages. No graphs.
Multilingualism	Provide a mechanism to manage multilingualism for reference data.	Must	Yes	The multilingual is supported by the Re3gistry both import and export.
Log	Provide a log of actions performed on reference data.	Should	Yes	All the changes are tracked in the database. Including also item versioning.
Log	Provide a log of actions performed by a user.	Must	Yes	The interface provides a log table. The user is notified by email about all the actions.
Licensing	Release the reference data component with an open source licence (or any other appropriate licence of choice).	Must	Plan	A direct reference to the data license in every item is planned. The Re3gistry software is open source.
Documentation	Provide complete documentation and guidelines for IT and business users.	Should	Yes	User interface manual and software manual are available in the software package. There is also a complete guide to extend the software (from v.0.4). The user interface also provides simple help functionalities.

Domain	Requirement	MoSCoW	Coverage Re3Gistry	Observations Re3gistry
Data Dictionary	Provide public HTTP URIs for the different reference data entities.	Should	Yes	Unique identifier for each item not dependent to the language/format representation. The representation is handled by the HTTP content negotiation.
Collaboration	Allow the user to work in a collaborative way.	Could	No	Collaboratively working on register content is out of scope
Classification	Provide to the user a way to classify the concept schema, concept or terms created in the reference data with a taxonomy mechanism.	Could	Yes	The Re3gistry support the hierarchical structure of items. The hierarchical structure of items can be described by parents, collections and extensibility.
Security	Provide a user interface for manage users, enabling the administrators to create, update or delete users and roles, as well as the possibility to assign roles to a user.	Should	Part	A mechanism exists but it does not allow to dynamically managing user roles.
Security	Provide a mean to limit the access rights of the reference data. Manage access rights for different parts of the reference data (limited access).	Must	No	The INSPIRE Registry content is open.

Domain	Requirement	MoSCoW	Coverage Re3Gistry	Observations Re3gistry
Security	Limit the access to reference data according to its current status.	Must	No	This requirement is not implemented in the Re3Gistry.

Annex IV. PROPOSED CHANGE REQUESTS FOR THE ADAPTATION OF REDDA

This section will describe the necessary adaptations to REDDA in order to fit the needs and requirements of **DG MARE** and the **ISA Programme**.

IV.1. Make REDDA reusable by third parties as open-source software

Requested by: DIGIT

Requirements: R2, TR13, TR18, TR19

Status: Completed

Current situation:

- Before August 2014: REDDA is embedded in the IT system of DG COMP.
- After August 2014: The component is published and promoted via Joinup⁹; as a package. Nevertheless technical documentation is not available which does not enable third parties to deploy it properly in their infrastructure. Even if the tool should be deployed, it would be not possible to the business to work with it efficiently as the functional documentation is not provided. Regarding this package, the licence model is still missing.

Desired situation: The software should be made re-usable by other DGs and Member States. To that end, a number of changes need to be implemented:

- The tool should be packaged as a software so that it can be installed on non-DG COMP systems;
- The documentation for the tool should be further elaborated and disseminated:
 - **Create a brochure:** The brochure should allow business users to understand all features of the tool as well as the benefits brought to their organisation by REDDA; and
 - **Technical documentation** should be created to enable the IT department of an organisation to install and configure the tool. An example of such documentation is the Re3Gistry technical documentation (European Commission - Joint Research Centre, 2014), which is included in the installation package on Joinup¹⁰.
- REDDA should be made available as an open source solution to the Member States of the European Union;
- REDDA should be made available under an open software licence, such as the EUPL.

⁹ A project for this was already created on Joinup: <https://joinup.ec.europa.eu/asset/rd/description>

¹⁰ The Re3Gistry installation package can be downloaded from <https://joinup.ec.europa.eu/software/re3gistry/description>

Impact or risks: If this change is not implemented, third parties will not be able to use REDDA.

IV.2. Make REDDA available as a Service to Commission services

Requested by: DIGIT

Requirements: R8, TR24, TR25

Status: Open

Current situation: RDC is an online tool that is currently only used by DG COMP for managing the GENIS reference data.

Desired situation: REDDA should be made available as a service to other DGs and Commission services. This means that for usage within the Commission, users would not be required to install the software on their own servers, but that the already deployed software should be made available to other DGs. In order to make REDDA available as a service, the tool should take into account the following technical considerations:

- Proper identity and access management (IAM):

As some of the CISE reference data is confidential, both the access to the application itself as the access to reference data projects within the application should be properly managed. Apart from keeping unauthorised users from accessing the application, REDDA should allow users to manage access rights on the level of the reference data projects, groups or even entities. For example, users from DG COMP should not have access to sensitive code lists ("reference data groups") in the CISE project.

- Continuity of service:

In order to successfully make REDDA available as a service to a number of organisations, the tool should assure a close-to-zero downtime. The architecture should be developed in a way that it can deal with server issues, upgrades, debugging and backups without interrupting the availability to the clients.

- Scalability:

For the software to serve a potential high amount of clients, its architecture should be scalable. Scalability should be interpreted in terms of

- Capacity management: can the servers and the application handle an increased load?
- Extension of the user-base: the application needs to provide APIs that enables users to integrate the REDDA data with their systems.

Besides technical considerations, making software available as a service entails a number of consequences for the business, such as:

- Training: in order to have the service used in a proper way, the owner of the application should organise trainings for potential users and administrators;

- Service Level Agreements: since the software will be made available as a service to other DGs and institutions, SLAs will need to be created in order to cover the legal aspects of the service delivery.
- Pricing models: the maintenance and development of REDDA as a service will entail significant costs for its owner. Therefore, a cost allocation model should be developed before launching the software as a service;
- IT support: providing a service to clients means that these clients should have access to relevant IT support on the tool on a permanent basis; and
- Evolutive maintenance: as the service may be extended to new users over time and as the needs of existing users could change, the tool will need to be continuously developed in order to stay relevant.

Impact or risks: If this change is not implemented, it will be impossible to make REDDA available as a service, which would be a significant impediment to the actual usability of the tool outside DG COMP. Making REDDA available as a service without taking into account the above mentioned considerations might lead to technical and legal issues.

IV.3. Support standard formats for import and export

Requested by: DG MARE

Requirements: R1, R4, TR10, TR11, TR12

Status: Open

Current situation: The RDC allows to import or to export reference data in an XML format which is specific to DG COMP. It is currently not possible to import or export in a standard like SKOS and its extension SKOS-XL.

Desired situation: In order to enable a better interoperability of the reference data, REDDA should support import and export to and from the following formats:

- RDF/SKOS
- JSON
- XML

The main benefits expected:

- Interoperability: reduce interoperability conflicts and allow integration with other building blocks ; and
- Re-usability: Foster the re-use of REDDA.
- Avoid data lock-in.

Impact or risks: Not supporting standardised export and import formats is a significant impediment to the interoperability with other tools and IT systems. Such interoperability is key in order to assure the re-usability of the software and its reference data.

IV.4. Reference data publication on the Re3gistry

Requested by: DG MARE

Requirements: R3, R4, R5, R11, TR6, TR9, TR20

Status: Open

Current situation: The RDC supports deployment of the reference data to some DG COMP-specific databases and IT systems.

Desired situation: In order to support the design, maintenance and publication of reference data, REDDA should extend the RDC so that it could be integrated with the Re3Gistry reference data repository.

Risk: Risks are limited since there is a workaround. Although an integration of REDDA with the Re3Gistry would bring additional features to both tools, the integration is not crucial in order to meet the reference data requirements from the stakeholders. As REDDA supports exporting reference data, the exported data can still be transformed in order to meet the import requirements of the Re3Gistry.

IV.5. Reference data publication on WebDAV / SVN / Git

Requested by: DIGIT, DG MARE

Requirements: R3, R4, R5, TR6, TR9, TR20

Status: Open

Current situation: The RDC does not provide a mechanism for the publication of reference data.

Desired situation: REDDA must extend the RDC so that it can provide this functionality in order to enable humans and/or machines to access and process reference data, through the use of common interfaces such as WebDAV, Git, SubVersion, or (read-only) HTTP(s).

The tool should support:

- Provisioning all versions (full versioning of temporal changes and language versions);
- Read-access over HTTP/s; and
- Write-access over WebDAV, Git or Subversion.

The main benefits expected:

- Re-usability: third parties will be able to easier re-use reference data.
- Avoid data lock-in.

Impact or risks: Limited risk as there is a workaround. The export feature would allow to manually creating a file, which can afterwards be manually published using WebDAV, Git, SubVersion or via other tools such as the Re3Gistry.

IV.6. User and access rights

Requested by: DG MARE,

Requirements: R8, TR23, TR24, TR25

Status: Open

Current situation: As REDDA is based on the development done for RDC; REDDA provides only user management limited to three profiles with a hardcoded matrix mechanism to manage the access. Therefore to define new roles and rights, it is necessary to request for new development.

Desired situation: REDDA must provide the option for users to define customised roles, responsibilities and access rights. The metadata management and governance methodology prescribes a clear validation workflow which assures segregation of duties. In order to support this workflow, the end-users of REDDA need to have the opportunity to define custom user roles and to link access rights and responsibilities to these roles. More specifically, the tool needs to:

- Foresee a role for members of the **operational team** and for members of the **governance committee**. Within the operational team, **editors** and **reviewers** need to be defined. However, as stated above, it is recommended not to encode these user roles into the code of REDDA, but rather foresee the option for users to define their own roles;
- Allow user roles to be defined on the level of the reference data project. A member of the governance committee of project A, for instance, might be a member of the operational team in project B;
- Specify the validation status of a reference data element, e.g.: draft, for review, validated. The statuses should be linked to user roles and access rights. For example, only the reviewer of a reference data element should be able to set the status to "validated". Once the element is validated, the editor should not be able to edit the element any longer;
- Safeguard "segregation of duties" in the create, update and retire processes, e.g. the user validating a change should be different from the user that drafted that change; and
- Keep a log of all actions performed in REDDA. For the purpose of maintaining an audit trail, the logs should be able to give an overview of activities performed per user, per reference data element and per reference data project.

The main benefits expected:

- Data security: managing the access to private and/or sensitive reference data;
- Risk management: management processes will be under control and roles and responsibilities will be clearly defined and enforced within a team;
- Cost reduction: as there is no need to go into development for specifying user roles into the system, the installation and maintenance costs for end users will be significantly lower;

- Easier customisation: because of the ability to define and modify roles and responsibilities on the appropriate level of granularity, customised to the needs of the organisation, access rights will be managed more effectively; and
- Re-usability: the tool will be accepted easier by the end users and will reinforce the re-use of the tools.

Impact or Risks: If this change is not implemented, the tool will not answer to the need of the stakeholders, which might result in non-acceptance. Users could limit the potential risks by setting up business procedures to put in place the necessary controls. In that case, the IT processes will not play a role in supporting the risk control processes. As IT controls tend to be more reliable than business controls, their absence might result in inefficiencies and erroneous reference data.