

Document Metadata

Property	Value
Date	18/05/2017
Version 1.00	
Authors	Alexandru Droscariu – PwC EU Services Ana Fernández de Soria Risco – PwC EU Services Emidio Stani – PwC EU Services Ioana Novacean – PwC EU Services
Reviewed by	Nikolaos Loutas – PwC EU Services Susanne Wigard – European Commission
Approved by	Susanne Wigard – European Commission

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.

Table of Contents

GL	OSSARY	IV			
1.	INTRODUC	TION	5		
	1.1.	OBJECTIVES &	& SCOPE		5
	1.2.	Approach		!	5
	Selection	n criteria			6
	1.3.	STRUCTURE			7
2.	REQUIREM	IENTS	8		
	2.1.	CREATE, USE	AND EXTEND	:	8
	2.2.	CHANGE MAN	NAGEMENT		9
	2.3.	USER MANAG	SEMENT		9
	2.4.	RELEASE			9
	2.5.	CREATE MAPI			9
	2.6.	MANAGE QU			0
	2.7.	COMMUNICA		1	0
3.	ANALYSIS	OF TOOLS	11		
	3.1.	CALLIMACHU	S	1	1
	3.2.	GINCO		1	7
	3.3.	RE3GISTRY		2	1
	3.4.	REGISTRY CO	RE	20	6
	3.5.	Skosmos		3	
		TEMATRES		3.	
	3.7.	VocBench		4	1
				List of Figures	
Fig	ure 1: Editir	ng a concep	t in Callimac	hus1	5
Fig	ure 2: Globa	al scores for	· Callimachu	s1	5
Fig	jure 3: Editir	ng a term in	Ginco	20	υ
Fig	ure 4: Glob	al scores for	Ginco	20	0
Fig			-	tess Restriction") which can be published in different formats in the control of	
Fig	ure 6: Glob	al scores for	Re3gistry	24	4
Fig	ure 7: "All p	properties" t	tab of a code	e list in the Environment Registry2	9
Fig	ure 8: Glob	al scores for	Registry Co	ore3	0
Fig	ure 9: Exam	nple of a voc	cabulary terr	m presented in Skosmos3	4
				· 3	
۰, ۵	,				•

Figure 11: TemaTres search terms suggestion / autocomplete
Figure 12: Example of vocabulary term in TemaTres
Figure 13: Example of vocabulary term in TemaTres, metadata view
Figure 14: Global scores for TemaTres
Figure 15: Relationships tab in VocBench
Figure 16: Global scores for VocBench
Figure 17: Global scores
Figure 18: Scores per tool and per category
List of Tables
List of Tables Table 1: Selection criteria, motivation, and measurement for tool selection
Table 1: Selection criteria, motivation, and measurement for tool selection
Table 1: Selection criteria, motivation, and measurement for tool selection
Table 1: Selection criteria, motivation, and measurement for tool selection
Table 1: Selection criteria, motivation, and measurement for tool selection
Table 1: Selection criteria, motivation, and measurement for tool selection

GLOSSARY

Term	Definition	
Authority list	Controlled vocabulary of descriptive terms designed to facilitate retrieval of information	
Code list	List of values in a predefined set that can be used in metadata and which help metadata creators in selecting from a set of descriptors.	
Concept	In this context, a word or code in a code list.	
Ontology	Formal naming and definition of the types, properties, and relationships of the entities in a domain.	
Term	A word or code in a code list.	
Thesaurus	A type of controlled vocabulary seeking to dictate semantic manifestations of metadata.	
Vocabulary	List of terms in a particular domain and pertaining definitions.	

1. Introduction

This document is part of TASK-06 "Tools and Methodologies" of <u>ISA</u>² Action 2016.07 "Promoting semantic interoperability amongst EU Member States", commonly known as <u>SEMIC</u>. This task aims to provide updates to the tools and methodologies developed by the SEMIC action. The current report's purpose is to provide practical guidance to less experienced organisations on the selection of the most appropriate tools for reference data management in general and code lists in particular.

This document will build on previous work carried out under the SEMIC action, which focused on the governance and management of data models as well as for tools for managing those¹. Deliverable "D06.01 – Guidelines for the use of code lists", under the same specific contract, also concerns code list management, but from a different point of view, as such the two reports complement each other. While D06.01 guides code list publishers and consumers on the subject of code list management and governance in a manner that is tool-agnostic, the present deliverable provides guidance on how to choose a suitable tool in line with the needs of each organisation.

1.1. Objectives & scope

The objective of this document is to provide guidance to owners and publishers of code lists on the selection of an appropriate tool for code list management depending on their needs or the requirements of their organisations. By code list management, we understand the entire lifecycle of a code list, including design, release, change management, extension, mapping, quality management, communication, etc.

As the needs of organisations vary greatly by area of activity, size, and purpose of code lists, the guidance offered by this report does not aim to provide a definitive approach to the selection of a code list management tool. It does however provide an overview of the main features such a tool should provide. While the findings of the report could apply to almost any organisation, the analysis considers public administration representatives as its main stakeholders.

When an organisation, a developer or any other person needs to work with code lists, they have to do so through various points of the code list lifecycle: design, release, change management, extension, retirement, etc. Even more, if they want to exchange information based on a code list, they might need to perform mappings with other code lists or transform the codes to an agreed-to format, etc.

1.2. Approach

This section defines the approach followed for the development of this report, which included:

- Determining the appropriate selection criteria for a solution to be included in the analysis;
- Using the aforementioned selection criteria to draw a list of solutions;

¹ https://www.slideshare.net/SEMICeu/semic-governance-and-management-of-data-specifications

- Determining which features would be the focus of the analysis after the selection of the solutions;
- Evaluating the solutions against the pre-determined list of features;
- Summarising the findings of the analysis and drawing appropriate conclusions.

Selection criteria

This sub-section explains the criteria defined for the selection of tools.

In order to make a relevant selection of code list management tools that can be used by public administrations in the Member States of the EU and EU institutions, a number of criteria act as pre-conditions for a tool to be considered for the report. The motivation for selecting these criteria and the indicators used for measuring the compliance of a tool with the criteria are provided in Table 1.

Table 1: Selection criteria, motivation, and measurement for tool selection

Criterion	Motivation	Measurement
	To make sure the tool is relevant to the target audience of the report.	
Open Source Software	, , , , , , , , , , , , , , , , , , ,	software should mention an
Maintenance & activity	To avoid inactive solutions and enjoy the benefits of an active user community.	Number of users, repository activity, publications about the tool.

The proven **use of the tool by a public administration** can provide the additional advantage of reduced costs (as a result of reusing a tool developed by/for another public administration), or of having a community of practice that includes other public administration representatives. The <u>Sharing and Reuse Framework</u>², a European Commission guideline on the improvement of public IT services through sharing, reuse and collaborative development of IT solutions, encourages the exchange of information among public administrations, in addition to reusing or sharing software.

The focus on **Open Source** tools is also in line with the Sharing and Reuse Framework's specifications for how public administrations can improve their service delivery, and with one of the underlying principles of the <u>European Interoperability Framework</u>³. Additionally, since some organisations have either an obligation to use

_

² https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2017-10/sharing_and_reuse_of_it_solutions_framework_final.pdf

³ https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

Open Source software or follow a "comply or explain" policy⁴ in this regard, delving into commercial solutions would have limited impact and usefulness to the target audience.

1.3. Structure

The remainder of this report is structured as follows:

- Section 2 delves into the functional requirements specifically examined through the analysis;
- Section 3 contains the solutions analysis based on the requirements identified in section 2;
- Section 4 summarises the findings of the work.

⁴ The "comply or explain" policy refers to the policy adopted in certain Member States of the EU that requires public administrations to use Open standards in all public administration IT solutions, or explain why they did not. One example of such a policy is that of The Netherlands: https://www.noraonline.nl/wiki/Beleid_open_standarden

7

2. REQUIREMENTS

This section contains an overview of the requirements and expectations public administrations might have regarding solutions for code list management. These have been collected over the years through the interactions of SEMIC with publishers of code lists in the EU institutions and Member States.

In the interest of supporting public administrations seeking guidance in the selection of an appropriate code list management tool to best serve their needs, the approach selected for this section focuses on the features of the tools and user experience aspects such as multilingual interfaces and ease of use, coupled with the different typical steps involved in the management of a code list.

Each sub-section features the relevant features from important points in the management and governance of a code list. The aim is to prepare a concrete overview of these required features to evaluate the solutions and determine their suitability in section 3.

2.1. Create, use and extend

The following features can be important to code list management software users:

- Installation process: the existence of a clearly identified and explained installation process for a tool (applicable to tools not available as a service);
- Web applications vs. stand-alone applications: Web applications usually
 provide the potential for collaborative work, and could be seen as easier to
 access than stand-alone applications, in addition to facilitating the sharing of
 existing resources;
- Multilingual user interface: as in most cases the users of the tool will not be English native speakers. In some countries, it is also required to support several national languages;
- Multilingual code list support: the tool should be able to support the management of multilingual code lists, such as the Named Authority Lists of the Publications Office. The support of standards such as SKOS-XL⁵, OASIS XLIFF⁶, or W3C's ITS⁷, which support multilingual labels for codes can cover this aspect;
- Search: simple or advanced search, auto-completion, metadata search function, support for SPARQL, SQL or other advanced query language, dropdown of applicable terms;
- Data import/export: export in different formats, preferably SKOS, according
 to compatibility with vocabulary metadata; import content from other tools in
 multiple formats, ideally at least SKOS/RDF and XML. API⁸ access, for
 example for importing code lists from other applications, or for exporting, is
 also important;

⁵ https://www.w3.org/TR/skos-reference/skos-xl.html

⁶ https://www.oasis-open.org/committees/xliff/

⁷ https://www.w3.org/TR/its20/

⁸ https://techterms.com/definition/api

- User support: guides, tutorials, usage methodologies, wikis, FAQs, etc.;
- Usability: No limitation on the number of terms or concepts, user friendliness and capacity for personalisation, usefulness of menus and complexity of creating or modifying content, look and feel, etc.;
- Visualisation of code lists: visual editing, visible hierarchy, tree structures, etc. for presenting and browsing code lists.

2.2. Change management

The following features can be important to code list solution publishers and consumers:

- Change request management: ticket collection and issue tracking, etc.;
- Change synchronisation: use of Web services to synchronise code list versions by pushing code lists directly to applications;
- Notify changes: ability to notify code list consumers of changes to a code list;

2.3. User management

- User profile management;
- Credential-based authentication: Distinguishing users through individual usernames and passwords;
- Role-based access: different autorization for users based on their roles, e.g. only admin can delete a concept, editors can only add a concept, etc.

2.4. Release

The following features can be important to code list solution users:

- Documentation: availability of sufficiently thorough documentation supporting at least the current version of the software;
- Versioning: showing the version number of a concept or entire code list;
- History: providing information about the progression of a code or the entire code list, support for a release calendar;
- Status attribution: the ability to mark a concept, group of concepts or entire code list as active, superseded or retired by attributing a status label;
- Providing licensing information: the feature of displaying licensing information at concept- or list-level as a way to encourage reuse and maintain legal certainty;
- The ability to retire or delete individual terms;
- The ability to retire or delete groups of terms or entire code lists.

2.5. Create mappings

The following features can be important to potential code list solution users:

- Managing relations between a code in one code list and a term from another vocabulary;
- Defining and creating relations between codes in a given code list and another Web resource such as a DPpedia dataset (this feature only applies for those users interested in the Semantic Web⁹);
- Defining different types of relationships between terms.

2.6. Manage quality

The following features can be important to code list solution users:

- Quality control: metadata quality is an important aspect of facilitating access
 to information and search¹⁰, both at code list level and for each term in the
 code list. This aspect could be ensured by the presence of a validator or entry
 field-level validation during the creation or updating of a code. A high level of
 metadata quality can make information more easily findable by specifically
 providing certain types of metadata such as description or date attributes;
- (Dis)allow duplicate terms: ability to disable the possibility of entering duplicate terms into a code list.
- Consistency control: test for erroneous relations and duplicate concepts within a single language;

2.7. Communicate

The following features can be important to code list solution users:

- Forums: Web pages where code list managers (publishers, editors, consumers) and solution owners (code list management software developers) can discuss features and bugs, future developments, etc.
- Support requests: the existence of one or more ways for solution users to request support from the developer or the user community;
- RSS feed: a way for solution owners to push news about the solution to interested parties that subscribe to the feed.
- Publications: blogs, use cases, mailing lists, videos, etc.

⁹ https://www.w3.org/standards/semanticweb/

¹⁰ http://dri.ie/sites/default/files/files/metadata-quality-control.pdf

3. Analysis of tools

This section analyses a list of tools that fulfil the basic selection criteria, as described in section 1.2, against the features listed in section 2. Each tool fulfils the three basic criteria of this analysis (being Open Source Software, having been used by a public administration entity, and having some degree of presence in the market and/or activity around it). The information was collected by means of desk research.

Each sub-section provides an overview of the main features of the solutions, as well as aspects related to its ownership, licensing information, and possible room for improvement or gaps.

3.1. Callimachus



<u>Callimachus</u>¹¹ is a content management system which enables content publishing via web pages, giving the possibility to export metadata in RDF. This Open Source solution is released under Apache License 2.0¹². It is regularly updated on GitHub¹³, where it has 5 contributors, 19 watchers, 79 favourites and 19 forks. In the public sector, it has been used by the US Environmental Protection Agency¹⁴.

Callimachus enables the creation of different types of content, including SKOS concepts. A concept can have the following properties:

- Label;
- Alternate label;
- Definition;
- Example;
- Scope;
- History;
- Related concept;
- Narrower concept;
- Image;
- Change notes.

Concepts can then be organized in folders, which can act as code lists within its metadata description. Users can delete a single concept or an entire code list (folder).

¹¹ http://callimachusproject.org/

¹² https://www.apache.org/licenses/LICENSE-2.0

¹³ https://github.com/3-Round-Stones/callimachus

^{14 &}lt;u>http://callimachusproject.org/videos/0.17/epa-success-story.xhtml?view</u>

Create, use and extend	
Installation process	✓
Web application	✓
Stand-alone application	×
Create new code list	✓
Edit existing code list	✓
Browse existing code list	✓
Multilingual interface	×
Multilingual vocabularies	✓
Search	✓
Data import formats	RDF, TTL, JSON, XML
Data export formats	RDF/XML, Turtle, JSON-LD
API access	✓ (RESTful API integration)
User support	✓
Usability	✓
Visualisation	×

Manage changes	
Change request management	×
Change synchronisation	×
Change notification	×

Manage users	
User management	✓
Credential-based authentication	✓
Role-based access	✓

Release	
Documentation	✓
Version number attribution	* (possibility to indicate it through change notes)
History	✓
Status attribution	×
Provide licensing information	×
Retire individual terms	✓
Retire groups of terms	×
Retire entire code list	✓

Create mappings	
Manage relationships between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web resource	✓
Different types of relationships	√ (only narrower and related)

Manage quality	
Quality control	×
(Dis)allow duplicate terms	×
Consistency control	×

Communicate	
Forums	√ (as discussion)
Support requests	✓
RSS feed	✗ (developers can create RSS¹⁵)
Publications	√ (videos, blog posts, etc.)

http://callimachusproject.org/docs/1.5/callimachus-for-webdevelopers.docbook?view#Create_an_RSS_feed_from_two_named_Atom_feeds

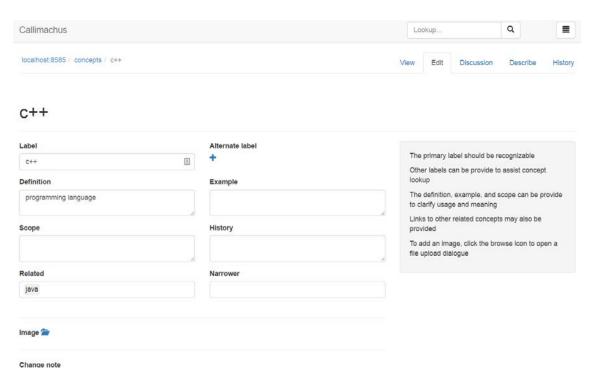


Figure 1: Editing a concept in Callimachus

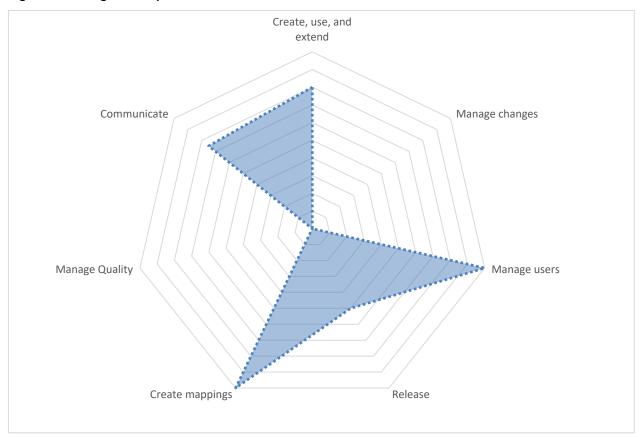


Figure 2: Global scores for Callimachus

Figure 2 presents a global view of how Callimachus scores for the different feature categories, each score as a proportion of the maximum score possible for its category.

As the figure shows, this tool is stronger for creating mappings and managing users, but lacks features related to quality management and change management.

3.2. Ginco



Ginco (Gestion Informatisée de Nomenclatures Collaboratives et Ouvertes¹⁶) is an application which allows creating and managing scientific and technical vocabularies (authority list, thesaurus, etc.). The application relies on the ISO 25964-1:2011 standard for the definition of the terms such as Concept and Term and it allows to export them in SKOS format.

Ginco is released as Open Source software under CeCiLL v2 license¹⁷. It is updated on GitHub¹⁸, where it has 8 contributors, 21 watchers, 31 favourites and 16 forks. In the public sector, it has been developed by the Ministry of Culture and Communication of France¹⁹, which publishes vocabularies online²⁰.

Create, use and extend	
Installation process	√ (relatively complex, but thoroughly explained by documentation)
Web application	✓
Stand-alone application	✓
Create new code list	✓
Edit existing code list	✓
Browse existing code list	√ (code lists are stored as file)
Multilingual interface	√ (yes, default is French)
Multilingual vocabularies	✓
Search	✓
Data import formats	SKOS, GINCO XML
Data export formats	RDF, TXT, GINCO XML

¹⁶ "Computerized Management of Collaborative and Open Nomenclatures" in English

_

¹⁷ http://www.cecill.info/licences/Licence CeCILL V2-en.html

¹⁸ https://github.com/culturecommunication/ginco

https://github.com/culturecommunication/ginco#what-is-ginco-

²⁰ http://data.culture.fr/thesaurus/

Create, use and extend	
API access	✓ (SOAP API access)
User support	✓
Usability	√ (customisable)
Visualisation	×

Manage changes	
Change request management	ж
Change synchronisation	×
Change notification	×

Manage users	
User management	✓
Credential-based authentication	✓
Role-based access	√ (administrator,expert,etc.)

Release	
Documentation	✓
Version number attribution	✓ (for code list)
History	✓
Status attribution	✓
Provide licensing information	×
Retire individual terms	✓

Release	
Retire groups of terms	✓
Retire entire code list	✓

Create mappings	
Manage relationship between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web resource	✓
Different types of relationships	✓

Manage quality	
Quality control	✓
(Dis)allow duplicate terms	✓
Consistency control	✓

Communicate	
Forums	√ (through suggestions)
Support requests	✓
RSS feed	×
Publications	×

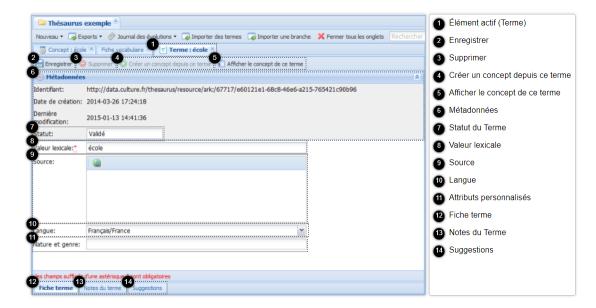


Figure 3: Editing a term in Ginco

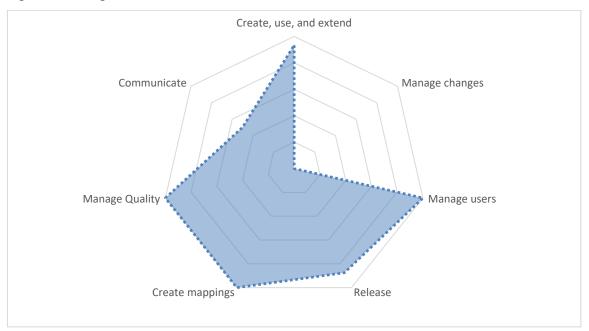


Figure 4: Global scores for Ginco

Figure 4 presents a global view of how Ginco scores for the different feature categories, each score as a proportion of the maximum score possible for its category. The scores indicate this tool is suitable for most aspects, with the exception of change management and, to a lesser degree, communication aspects.

3.3. Re3gistry

Re³gistry

The <u>Re3gistry</u>²¹ is an Open Source software released under European Union Public Licence - <u>EUPL</u> v.1.1²² and created to serve as a tool for managing and sharing reference codes in the context of <u>INSPIRE</u>²³. It helps ensure that concepts are correctly referenced in INSPIRE, but also other sectors. The solution has been used extensively by public administrations, including the European Commission and governmental organisations from Austria, Finland, France, Italy, Slovakia, etc. ²⁴

Create, use and extend	
Installation process	√ (relatively complex, but thoroughly explained by documentation)
Web application	✓
Stand-alone application	✓
Create new code list	✓
Edit existing code list	✓
Browse existing code list	✓
Multilingual interface	✓
Multilingual vocabularies	✓
Search	✓
Data import formats	CSV
Data export formats	HTML, XML, JSON, RDF, Atom, CSV
API access	✓ (RESTful API access)

²¹ Re3gistry: https://ec.europa.eu/isa2/solutions/re3gistry_en

²² EUPL: https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/eupl1.1.-licence-en_0.pdf

²³ INSPIRE Knowledge Base: http://inspire.ec.europa.eu/

²⁴ https://ec.europa.eu/isa2/solutions/re3gistry_en

Create, use and extend	
User support	✓
Usability	√ (customisable)
Visualisation	×

Manage changes	
Change request management	×
Change synchronisation	×
Change notification	×

Manage users	
User management	✓
Credential-based authentication	✓
Role-based access	√ (admin, user, etc.)

Release	
Documentation	✓
Version number attribution	✓
History	✓
Status attribution	✓
Provide licensing information	✓
Retire individual terms	✓
Retire groups of terms	✓

Release	
Retire entire code list	✓

Create mappings	
Manage relationship between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web resource	✓
Different types of relationships	✓

Manage quality	
Quality control	✓
(Dis)allow duplicate terms	×
Consistency control	✓

Communicate	
Forums	✓
Support requests	√ (bug reporting, support requests)
RSS feed	√ (Related to all INSPIRE initiatives)
Publications	√ (news items, videos, demos)

Authentication occurs either through Apache SHIRO (static user authentication using a simple configuration file containing the list of users) or through the EU Login system.

When importing data, the files need to be compressed in .zip format.

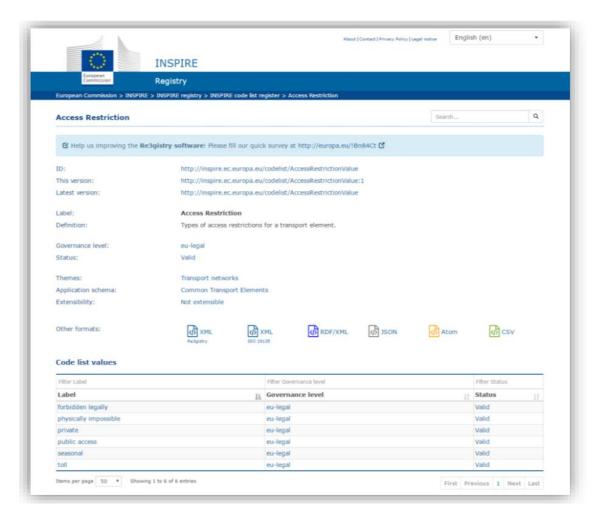


Figure 5: Example of code list ("Access Restriction") which can be published in different formats in Re3gistry

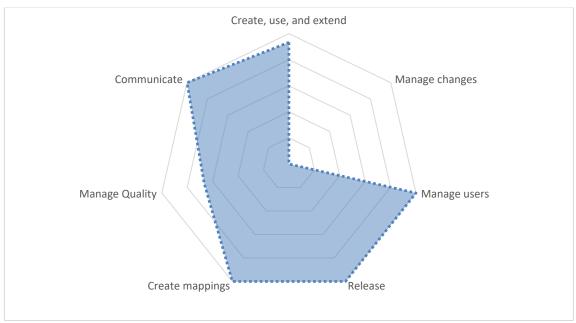


Figure 6: Global scores for Re3gistry

Figure 6 presents a global view of how Re3gistry scores for the different feature categories, each score as a proportion of the maximum score possible for its category. This tool's strongest aspects are user management, mapping creation, release

management, and communication. It also has a relatively high quality management score, but is lacking in the change management area.

3.4. Registry Core

Registry Core is an Open Source solution the UK government contracted. It serves as the basis for the Core Reference Data Management Platform provided by Epimorphics, which supports and documents Open Source standards itself. The GitHub²⁵ repository for this solution shows constant activity, with 10 watchers, 9 favourites and 5 forks. The solution is released under Apache license and it can be installed following a relatively complex process²⁶, or implementations can be contracted, for instance through the UK government's Digital Marketplace²⁷.

The solution has different implementations by public administrations. For instance, the <u>Environment Registry</u>²⁸ is a Web application available online, featuring user management. Users that are simply interested in accessing the existing code lists do not need any credentials, but those wishing to add or change entries do need a username and password.

Code lists existing within the solution display as a simple Web page containing important details such as a code list's URI, a description of it, and a table featuring the name, notation, description, types and status of the concepts in the code list. Remarkably, change management is covered by the possibility to label the concepts in a code list as "experimental" or "stable". These labels can also apply to entire code lists or registers. The "All properties" tab for a register, for instance, brings up a table containing information like description, label, owner, but also versioning information and date when it has last been modified, as shown in Figure 7.

Create, use and extend	
Installation process	* (not required)
Web application	✓
Stand-alone application	×
Create new code list	✓
Edit existing code list	✓
Browse existing code list	✓
Multilingual interface	×
Multilingual vocabularies	×
Search	✓

²⁵ https://github.com/UKGovLD/registry-core

_

²⁶ https://github.com/UKGovLD/registry-core/wiki/Installation

²⁷ https://www.digitalmarketplace.service.gov.uk/g-cloud/services/991341327991934

²⁸ https://environment.data.gov.uk/registry/_def

Create, use and extend	
Data import formats	RDF syntaxes Turtle and RDF/XML, JSON-LD, etc.
Data export formats	RDF, RDF/XML, JSON-LD, CSV, etc.
API access	✓
User support	√ (e-mail, online ticketing)
Usability	Somewhat complex to install without help of contractor.
Visualisation	✓

Manage changes	
Change request management	×
Change synchronisation	×
Change notification	×

Manage users	
User management	✓
Credential-based authentication	✓
Role-based access	✓

Release	
Documentation	✓
Version number attribution	✓
History	✓
Status attribution	✓

Release	
Provide licensing information	✓
Retire individual terms	✓
Retire groups of terms	✓
Retire entire code list	✓

Create mappings	
Manage relationship between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web resource	✓
Different types of relationships	✓

Manage quality	
Quality control	✓
(Dis)allow duplicate terms	×
Consistency control	×

Communicate	
Forums	✓
Support requests	✓
RSS feed	×
Publications	×

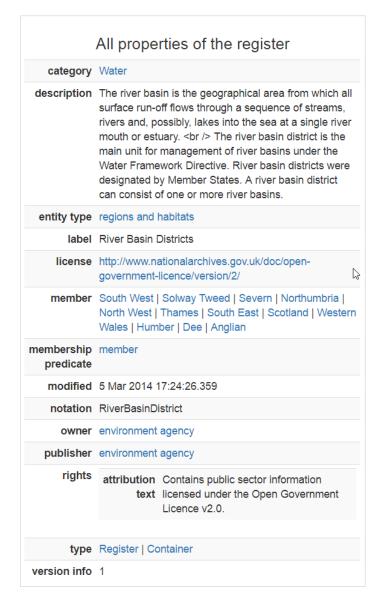


Figure 7: "All properties" tab of a code list in the Environment Registry

Individual concepts in a code list receive their own Web page and assigned URI. Each concept page contains information about the concept, such as label, notation, same as, and type. The "same as" field suggests the possibility to map the relationships between concepts in one code list and those in other vocabularies. Concept pages also feature a tab called "History", which contains information about the versions through which a term has passed, and its validity (both start date and end date if applicable).

Other features include the possibility to use an API or perform SPARQL queries and checking whether a URI is registered, both accessible without requiring users to register or log in to the system.



Figure 8: Global scores for Registry Core

Figure 8 presents a global view of how Registry Core scores for the different feature categories, each score as a proportion of the maximum score possible for its category. As the figure indicates, this tool is most useful for user management, release management, and mapping creation. It is most lacking in change management and quality management.

3.5. Skosmos

Skosmos²⁹ is an Open Source software for publishing controlled vocabularies, distributed under the MIT licence³⁰. The National Library of Finland is the institution behind it. It has an active community on GitHub³¹, where it is frequently updated and followed by 20 watchers, favourited 92 times and forked 29 times.

Interested developers can install Skosmos and run their own instance, but the installation involves a relatively complex procedure, assuming they have a Linux machine with Apache and PHP installed and enabled. It also assumes the developer has a SPARQL endpoint with SKOS vocabulary data. However, typical users do not need to install the developer version, and can access Skosmos through a browser.

Create, use and extend	
Installation process	√ (not required for browsing)
Web application	✓
Stand-alone application	✓
Create new code list	✓
Edit existing code list	✓
Browse existing code list	✓
Multilingual interface	✓
Multilingual vocabularies	✓
Search	✓
Data import formats	SKOS
Data export formats	RDF/XML, TURTLE, JSON-LD
API access	✓
User support	✓
Usability	✓

²⁹ http://skosmos.org/

³⁰ http://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf

³¹ https://github.com/NatLibFi/Skosmos

Create, use and extend	
Visualisation	✓

Manage changes	
Change request management	×
Change synchronisation	×
Change notification	×

Manage users	
User management	✓
Credential-based authentication	×
Role-based access	×

Release	
Documentation	✓
Version number attribution	✓
History	✓
Status attribution	✓
Provide licensing information	×
Retire individual terms	✓
Retire groups of terms	✓
Retire entire code list	✓

Create mappings	
Manage relationship between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web resource	✓
Different types of relationships	✓

Manage quality	
Quality control	×
(Dis)allow duplicate terms	×
Consistency control	×

Communicate	
Forums	✓
Support requests	✓
RSS feed	×
Publications	√ (demo, scientific articles)

Other features:

- Structured concept display;
- Visualised concept hierarchy;
- Multilingual user interface: English, French, Italian, Finnish, Swedish, German, Spanish, Polish, Norwegian, Japanese, etc.;
- Linked Data access to underlying vocabulary data.

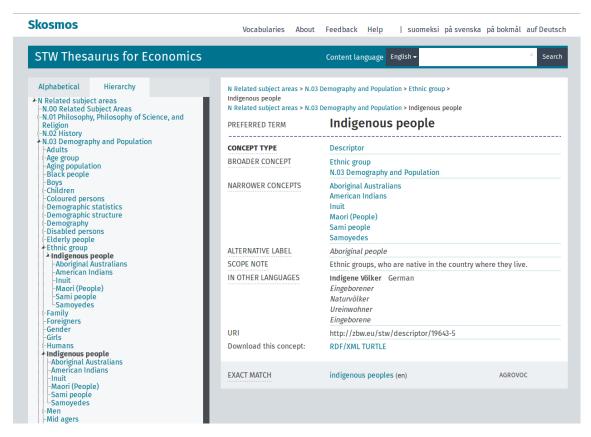


Figure 9: Example of a vocabulary term presented in Skosmos

Figure 9 shows an example of what a concept might look like in a code list in the Skosmos solution. As shown, the user can choose between the hierarchical and the alphabetical view for the full code list, and a significant amount of information is presented regarding related terms, related terms from other vocabularies, etc.



Figure 10: Global scores for Skosmos

Figure 10 presents a global view of how Skosmos scores for the different feature categories, each score as a proportion of the maximum score possible for its category.

The figure indicates Skosmos is particularly suitable for creating, using, and extending code lists, but also for creating mappings. It is however not a suitable tool for user management, quality management, nor change management.

3.6. TemaTres



TemaTres³² is an Open Source application to manage and exploit controlled vocabularies, thesauri, taxonomies, and other types of formal representations of knowledge. TemaTres was originally developed by the Library and the University of Buenos Aires in Argentina and released under the GNU General Public License 2.0³³. The solution has an active GitHub page³⁴ with 6 contributors, 11 watchers, 41 favourites and 29 forks.

TemaTres supports distributed management models, ensuring consistency and integrity of data and relationships between terms. Its features cover data traceability and quality control in the context of a controlled vocabulary.

There are more than 400 known cases of TemaTres being used to manage, publish, share, reuse vocabularies, thesauri, glossaries, code lists, etc. Many of these use cases come from public administrations such as: the Australian Government's Interactive Functions Thesaurus (AGIFT), the Argentinian Government's controlled vocabulary on the policy of the Labour Ministry, The National Agricultural Thesaurus of Brazil, The Monument Thesaurus Wales, and many others.

TemaTres is ideally suited for those publishers managing code lists on the web and those that need a long-term perspective and a dedicated management system. TemaTres is described as having a simple, functional user interface and sophisticated search capabilities.

The tool features a multilingual interface, specifically in English, Spanish, Italian, German, French, Portuguese, Dutch, Catalan, Chinese.

Regarding its usability, it is relatively simple and straightforward, but it does seem aimed at more experienced users, who can understand its alerts³⁵.

Create, use and extend	
Installation process	✓

³² http://www.vocabularyserver.com/index.html

³³ http://aims.fao.org/activity/blog/tematres-22-open-source-web-tool-manage-formal-representations-knowledge

^{34 &}lt;a href="https://github.com/tematres/TemaTres-Vocabulary-Server">https://github.com/tematres/TemaTres-Vocabulary-Server

³⁵ http://accidental-taxonomist.blogspot.be/2016/02/free-taxonomy-management-software.html

Create, use and extend	
Web application	✓
Stand-alone application	×
Create new code list	✓
Edit existing code list	✓
Browse existing code list	✓
Multilingual interface	✓
Multilingual vocabularies	✓
Search	✓
Data import formats	TXT, CSV, SKOS
Data export formats	XML, TXT, SKOS-Core, JSON, JSON-LD, PDF
API access	✓
User support	✓
Usability	✓
Visualisation	✓

Manage changes	
Change request management	✓
Change synchronisation	✓
Change notification	✓

Manage users	
User management	✓
Credential-based authentication	✓
Role-based access	✓

Release	
Documentation	✓
Version number attribution	×
History	✓
Status attribution	✓
Provide licensing information	×
Retire individual terms	✓
Retire groups of terms	✓
Retire entire code list	✓

Create mappings	
Manage relationship between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web Resource	✓
Different types of relationships	✓

Manage quality	
Quality control	✓
(Dis)allow duplicate terms	✓
Consistency control	✓

Communicate	
Forums	✓
Support requests	✓
RSS feed	×
Publications	√ (demo, examples, blog posts)

Additional features:

- SPARQL endpoint (SPARQL Protocol and RDF Query Language);
- Meta-terms; define facets, collections or arrays of terms;
- Powerful web services to expose vocabularies;
- Web services allowing for integration with other platforms;
- Search terms suggestion, as shown by Figure 11.

TemaTres



Figure 11: TemaTres search terms suggestion / autocomplete

Figure 12 and Figure 13 show an example of a term in TemaTres, illustrating features such as the relationships between a term and others, the inclusion of versioning and history information, etc.

marital status

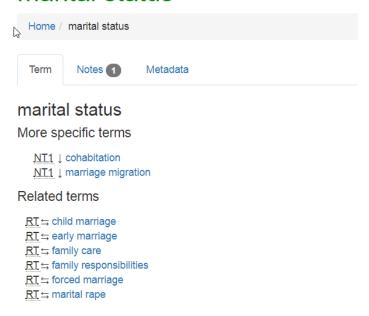


Figure 12: Example of vocabulary term in TemaTres

marital status

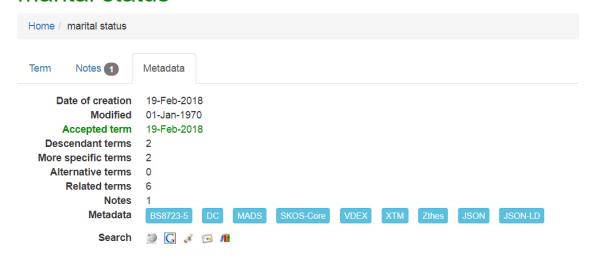


Figure 13: Example of vocabulary term in TemaTres, metadata view



Figure 14: Global scores for TemaTres

Figure 14 presents a global view of how TemaTres scores for the different feature categories, each score as a proportion of the maximum score possible for its category. The figure indicates almost perfect scores in most categories, albeit to a lesser degree for release management and communication.

3.7. VocBench



<u>VocBench</u>³⁶ is an editing and workflow tool that manages thesauri, authority lists and glossaries using SKOS. There are two active versions: VocBench 2 and VocBench 3. VocBench 3 enriched VocBench 2 with significant additional features, including versioning³⁷. VocBench is being used by various public sector organisations, including the European Commission, the Scottish Government and the Italian Government³⁸.

Create, use and extend	
Installation process	√ (detailed instructions)
Web application	✓
Stand-alone application	✓
Create new code list	✓
Edit existing code list	✓
Browse existing code list	✓
Multilingual interface	✓
Multilingual vocabularies	✓
Search	✓
Data import formats	RDF (SKOS, SKOS-XL), OWL
Data export formats	RDF (SKOS, SKOS-XL), OWL
API access	×
User support	✓

³⁶ http://vocbench.uniroma2.it/

38 http://vocbench.uniroma2.it/aboutus/

³⁷ Towards VocBench 3: Pushing Collaborative Development of Thesauri and Ontologies further beyond: http://ceur-ws.org/Vol-1937/paper4.pdf

Create, use and extend	
Usability	✓
Visualisation	✓

Manage changes	
Change request management	×
Change synchronisation	×
Change notification	×

Manage users	
User management	√ (Administration, group management)
Credential-based authentication	✓
Role-based access	✓

Release	
Documentation	✓
Version number attribution	✓
History	✓
Status attribution	✓
Provide licensing information	×
Retire individual terms	✓
Retire groups of terms	×

Release Retire entire code list ✓

Create mappings	
Manage relationship between concepts in a code list	✓
Define relationships between concepts in a code list and those from another Web resource	✓
Different types of relationships	✓

Manage quality	
Quality control	✓
(Dis)allow duplicate terms	×
Consistency control	✓

Communicate	
Forums	✓
Support requests	✓
RSS feed	ж
Publications	√ (scientific articles, guides)

Additional features:

- SPARQL endpoint Collaborative;
- Validation and publication workflow;
- Maintenance, validation, quality assurance;
- Extensive support infrastructure:

- o Detailed instructions for download, installation, configuration;
- User manual details all the operations that users can perform on the application;
- System administrator manual deals with various installation options and all the advanced features related to system customisation/reconfiguration that need to be performed under the hood;
- Developer manual provides behind-the-scenes information for VocBench developers, third party developers and anybody interested in how VocBench works;
- SKOS development pages providing insights on less-known aspects of SKOS development, best practices and typical modelling issues and how they are dealt with in VocBench.



Figure 15: Relationships tab in VocBench

Figure 15 shows the relationship section corresponding to a term as seen in VocBench³⁹. The figure also illustrates some other VocBench features, such as History.

-

³⁹ https://www.slideshare.net/Keizer/the-vocbench-project



Figure 16: Global scores for VocBench

Figure 16 presents a global view of how VocBench scores for the different feature categories, each score as a proportion of the maximum score possible for its category. VocBench scores relatively high in all categories, except for change management.

4. CONCLUSIONS

This section summarises the analysis performed in the previous sections, giving an overview of the tools and their features as seen from the point of view of the requirements a user might have at different points in the lifecycle of a code list.

Table 2 takes into account general requirements, including those regarding the actual usage of a code list and the possibility of extending it. Here, we see some solutions have some sort of required installation process, with a few of them having a relatively complex one for which organisations may require external help. However, many solutions differentiate between a developers' version of the solution, and a more basic one, for end-users simply interested in using the solution to manage code lists, and which is often available after a basic installation process and/or directly in the browser. Several solutions enable basic browsing of code lists without the need to actually install any software. Most solutions offer API access, opening the possibility to integrate them into other systems. A strong majority of solutions allow for multilingual vocabularies, and most of them also feature multilingual interfaces, enabling users to perform their work in a language they can comfortably use.

Table 2: Creation, usage and extension requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Installation process	✓	√ (relatively complex, but thoroughly explained by documentation)	but thoroughly	* (not required)	√ (not required for browsing)	✓	√ (detailed instructions)
Web application	✓	✓	✓	✓	✓	✓	✓
Stand- alone application	×	✓	✓	×	✓	√	√

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Create new code list	✓	✓	✓	✓	✓	√	✓
Edit existing code list	✓	√	✓	✓	✓	✓	✓
Browse existing code list	✓	√ (code list are stored as file)	✓	✓	✓	✓	✓
Multilingua I interface	×	√ (yes, default is French)	✓	×	✓	✓	✓
Multilingua I vocabularie s	✓	✓	✓	×	✓	✓	✓
Search	✓	✓	✓	✓	✓	✓	✓
Data import formats	RDF, TTL, JSON, XML	SKOS, GINCO XML	CSV	RDF syntaxes Turtle and RDF/XML, JSON-LD, etc.	SKOS	TXT, CSV, SKOS	RDF (SKOS, SKOS-XL), OWL

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Data export formats	RDF/XML, Turtle, JSON- LD	RDF, TXT, GINCO XML	HTML, XML, JSON, RDF, Atom, CSV		RDF/XML, TURTLE, JSON- LD	XML, TXT, SKOS- Core, JSON, JSON-LD, PDF	RDF/XML, TURTLE, JSON-LD
API access	√ (RESTful API integration)	✓ (SOAP API access)	✓ (RESTful API access)	✓	✓	✓	×
User support	✓	✓	✓	√ (e-mail, online ticketing)	✓	✓	✓
Usability	✓	√ (customisable)	√ (customisable)	✓ (Somewhat complex to install and configure)	✓ (customisable)	✓	✓
Visualisatio n	×	×	×	✓	×	✓	✓

Table 3 presents the situation regarding change management requirements users might have. It is one of the areas where the solutions tend to be lacking, with TemaTres being the only one that offers all three foreseen feature requirements. That said, change request management and

change notification are relatively easy to manage by code list publishers by using external solutions. Change synchronisation is likely to be lacking due to a relative technical complexity, in addition to code list consumers being less likely to accept automatic code list updates.

Table 3: Change management requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Change request management	×	×	×	×	×	✓	×
Change synchronisation	×	×	×	×	×	✓	×
Change notification	×	×	×	×	×	✓	×

Table 4: User management requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
User management	✓	√	✓	√	✓	✓	√ (Administration, group management)
Role-based access	✓	✓	✓	✓	×	✓	✓
Credential-based authentication	✓	✓ (administrator, expert, etc.)	√ (admin, user, etc.)	✓	×	✓	✓

Table 5 presents features related to release management. Particularly important here would be history, versioning, and status attribution. All analysed solutions offer some form of history records for each concept. Versioning is also provided by most solutions, along with status attribution. The possibility to feature licensing information for concepts is less common, potentially seen as less important because code list owners can simply provide licensing information covering the whole code list, not just a concept or group of concepts. Table 5 also provides information about the retirement features of the analysed solutions. As expected from such solutions, all of them feature the possibility to retire individual terms in a code list or the entire code list. More interesting is the aspect of deleting or retiring groups of terms, which is not possible in some solutions, like Callimachus or VocBench.

Table 5: Release requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Documentation	✓	✓	✓	✓	✓	✓	✓
Version number attribution	* (possibility to indicate it through change notes)	√ (for code list)	✓	√	✓	×	√
History	✓	✓	✓	✓	✓	✓	✓
Status attribution	×	✓	✓	✓	✓	✓	✓
Provide licensing information	×	×	✓	✓	×	×	×
Retire individual terms	✓	✓	✓	✓	✓	✓	✓

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Retire groups of terms	×	✓	✓	✓	✓	✓	×
Retire entire code list	✓	✓	✓	✓	✓	✓	✓

Table 6 contains information regarding the relationships between concepts, or between code lists and other vocabularies. All solutions feature some form of relationship management between the terms in a code list, even if it relates simply to designating narrower or broader terms or a basic "related" status.

Table 6: Mappings creation requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Manage relationship between concepts in a code list	✓	✓	√	✓	✓	√	✓
Define relationships between concepts in a code list and those from another Web resource	✓	✓	✓	✓	✓	✓	✓
Different types of relationships	√ (only narrower and related)	✓	✓	✓	✓	✓	✓

Table 7 contains information regarding quality management requirements. Issues like quality control are relatively well covered by most solutions, but features like the possibility to (dis)allow duplicate terms are rarer. Consistency control is a welcome feature for larger code lists, but only 3 out of 7 solutions mention it. Overall, quality management features are some of the less developed aspects of code list management solutions.

Table 7: Quality management requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Quality control	×	✓	✓	✓	×	✓	✓
(Dis)allow duplicate terms	×	✓	×	×	×	✓	×
Consistency control	×	✓	✓	×	×	✓	✓

Table 8 contains an overview of the communication features of the analysed solutions. The existence of forums where community members can debate and/or support each other, report issues or discuss difficulties encountered while using the solutions exist for almost all solutions. Alternatively, most solutions have some manner of accepting support requests, even if it is simply by reporting bugs or issues on GitHub or SourceForge. RSS feeds as a way to communicate with users are increasingly rare and unused, but most solutions mention a mailing list of some form.

Table 8: Communication requirements matrix

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Forums	√ (as discussion)	✓ (through suggestions)	✓	✓	✓	✓	✓

	Callimachus	Ginco	Re3gistry	Registry Core	Skosmos	TemaTres	VocBench
Support requests	✓	√	√ (bug reporting, support requests)	√	√	√	✓
RSS feed	(developers can create RSS)	×	√ (Related to all INSPIRE initiatives)	×	×	×	×
Publications	√ (videos, blog posts, etc.)	×	√ (news items, videos, demos)	×	√ (demo, scientific articles)	√ (demo, examples, blog posts)	√ (scientific articles, guides)

To sum up, organisations interested in creating or editing code lists can select an appropriate solution by examining the features of each solution against their own requirements. By selecting solutions which, as a baseline, are Open Source, active, and have previously been used in the public sector, this analysis ensures that almost any of the tools analysed here could be used for code list management with reasonable success. However, organisations using this report as guidance in selecting the most appropriate code list management software for their needs, should take into account the specific characteristics most relevant to their specific situation. For instance, an organisation lacking the technical skills to implement a more complex solution, might prefer avoiding those solutions which involve a complicated installation and configuration process. Alternatively, an organisation planning to create a code list that would have a lot of stakeholders and a lot of changes to implement to that code list would need to pay closer attention to the change management aspect.

Global scores

Figure 17 below illustrates the overall results of the tools' evaluation, based on collecting the total number of fulfilled requirements out of the total of 39 quantified features. The top three solutions, TemaTres, Re3gistry and Ginco, all fulfil more than 30 of the requirements, with TemaTres standing out as the solution meeting most requirements. Although some solutions earn better scores overall, this does not necessarily indicate that a higher-scoring solution is the optimal one to use for all organisations. Since organisations have different needs depending on whether they are code list publishers or consumers, whether their code lists are frequently updated, whether external change requests are common, the score itself does not make a definitive statement about the case-by-case suitability of a solution for the work of a given organisation. Needs, technical context, existing resources vary, and therefore an interested organisation would be better off examining whether a solution fulfils their individual requirements as well, not just whether it has a high score.

Figure 18 represents a juxtaposition of each tool's scores per category. This chart allows for easy comparison of the features provided by each of the tools analysed, in particular for users looking for software to support them for a specific area of code list management. This figure indicates a relative dearth of tools that enable change management, while most do very well in the areas of code list creation, use, and extension, as well as in user management, release management, and mapping creation.



Figure 17: Global scores

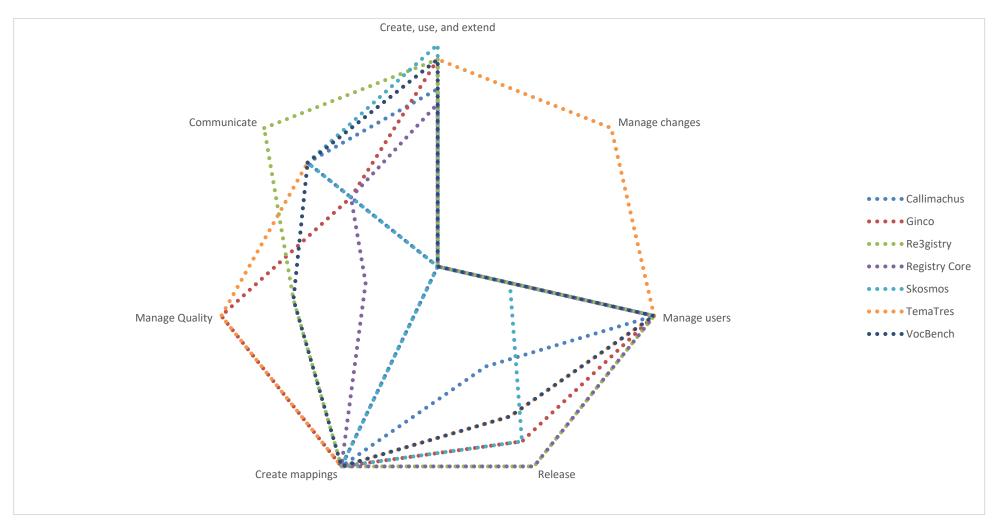


Figure 18: Scores per tool and per category