ShowVoc: a Thorough Platform for Publishing and Browsing Linked Open Datasets

Armando Stellato[1,2] [0000-0001-5374-2807], Manuel Fiorelli[1,2] [0000-0001-7079-8941], Tiziano Lorenzetti[2] [0000-0001-5676-8877], Andrea Turbati[2] [0000-0002-6214-4099], Willem van Gemert[3], Denis Dechandon[3], Anikó Gerencsér[3], Enrico Francesconi[4,5] [0000-0001-8397-5820]

[1] University of Rome "Tor Vergata", Department of Enterprise Engineering, via del Politecnico 1, 00133 Roma, Italy
manuel.fiorelli@uniroma2.it, stellato@uniroma2.it
[2] Lore Star srl, Italy
<name>.<lastname>@lorestar.it
[3] Publications Office of the European Union, Luxembourg
<name>.<lastname>@publications.europa.eu
[4] European Parliament, Luxembourg
[5] Italian National Research Council, Italy
enrico.francesconi@cnr.it

Abstract. ShowVoc is a web-based, multilingual, platform for publication of linked datasets of different nature and scope: OWL ontologies, SKOS/SKOS-XL thesauri, OntoLex-Lemon lexicons, generic RDF datasets and linksets (EDOAL, XKOS). Born in the context of the ISA2 European programme for the development of digital solutions for interoperable cross-border and cross-sector public services, ShowVoc aims at providing a one-stop shop for maximizing the diffusion of semantic and lexical resources as Linked Open Data. To this end, ShowVoc combines traditional data provisioning following LOD policies, an advanced browsing interface and inter-dataset functionalities (e.g. global search, navigation of dataset relationships/alignments, translation API benefiting from multilingual datasets and linksets) into a thorough platform for data provisioning. A metadata registry completes the offer combining different metadata vocabularies into an advanced catalog that can be inspected through a convenient user interface and LOD best practices. ShowVoc's mission, requirements and its collocation within the SoTA are discussed in the paper.

Keywords: Dataset Catalogs, Linked Open Datasets, Metadata repositories

1 Introduction

The Semantic Web [1], which is being built according to Linked Data [2] best practices, is based on the decentralized publication of disparate but interlinked datasets that together form a huge global graph. Although resolvable URIs and query-through-discovery are the defining access mechanism for a machine-accessible Web and the focus is on linking records, there is still a need, especially for humans, for a coarse-grained

perspective made of browsing, querying and visualization capabilities over the published resources. Discovery by link traversal - a la “follow your nose” - is closely related to people surfing the Web in search of information. If we take this analogy seriously, then we should consider people's reliance on search engines as an entry point to the Web. Although semantic web search engines are not as common as they could be, there has been a proliferation of dataset catalogs, both in specific domains and across the web, which play a similar role.

In this paper, we present ShowVoc, a platform for dataset publication and exploitation, which addresses both needs: it allows for the publication of datasets with resolvable URIs and a more sophisticated browsing experience than simple subject pages while offering a fully-fledged data portal for linked datasets.
ShowVoc can be seen as a companion to VocBench 3 [3], a platform for dataset development and maintenance, inheriting many of its features, such as its advanced multimodel support. However, while most of the operations in VocBench 3 deal with individual datasets, ShowVoc adds a number of cross-dataset operations that rely on managing multiple datasets. These include global search, translation and alignment management, which are based on the idea that multiple datasets contribute to a sort of giant virtual reference for terminology and translation.

ShowVoc is open source and made available under the BSD-3-Clause license. The project official web site is `https://showvoc.uniroma2.it/`. Source code and deployment artifacts are hosted on Bitbucket at `https://bitbucket.org/art-uniroma2/showvoc`.

2 History and Motivations

ShowVoc was not born as the direct, programmed outcome of a project aimed at its current intended target.

In 2017, the authors were involved in a European action called “Public Multilingual Knowledge Infrastructure” (PMKI), funded by a European ISA[2] programme[1] for the digitalization of Europe. The objective of PMKI was to contribute to overcoming language barriers within the EU by means of multilingual tools and services, focused on authoritative translation resources, which could feed and be used in combination with advanced translation services. The PMKI action was driven by recommendations from the European Interoperability Strategy (EIS[2]), thus fostering adherence to standards for describing language resources and the creation of an interoperability platform to manage and disseminate them. Another driving force was the European Interoperability Framework (EIF[3]), stressing the importance of multilingualism, accessibility, administrative simplification, transparency, and reusability of the solutions.

PMKI requirements thus mainly involved the acquisition, publication, dissemination, convenient visualization of lexical resources, with particular emphasis, at the content level, on multilingual and translation resources and, considering the stress on

[1] https://ec.europa.eu/isa2/
[2] http://ec.europa.eu/isa/documents/isa_annex_i_eis_en.pdf
[3] http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

fostering standards given by the EU action, the identification and adoption of a unifying model for lexical resources. Luckily, the time was right mature for the adoption of OntoLex-*lemon*, a lexicon model (including support for ontology-lexicon interfaces) developed with the exact intent of filling the representation gap for lexical resources in the Linked Data world, a model to which we actively participated both with contributions to the general specifications [4] and to its metadata characterization [5].

With the aim of fostering development and proliferation of resources in such newly-born standard, PMKI also funded support for OntoLex modeling within VocBench, a collaborative development platform for semantic resources, supported in turn by another parallel action of the ISA[2] programme and developed by the same team working on PMKI. This, together with converters from other popular standards for lexicons and terminologies, e.g. TermBase eXchange[4] (TBX), would have guaranteed a complete bootstrapping environment for OntoLex resources to be disseminated through PMKI.

Intuiting that such a combination of platforms could potentially become the focus of a true open-source and freely available ecosystem [22] for development and publication of Semantic Web resources, the authors suggested, for system design and implementation, to follow a wider plan that would meet PMKI requirements and, on the other hand, to think and prepare the architecture for a broader support for Semantic Web resources.

When PMKI was first put in pre-production for acceptance by different users of the European Commission, the system was appreciated for its innovative role as a dissemination platform for lexical and translation resources under the common umbrella of a Semantic Web community-driven model. However, not unexpectedly, many stakeholders, not limited in their comments by the scope of the PMKI mission, suggested the system to be extended to cover other sorts of datasets, including ontologies, thesauri and RDF datasets more in general which, in the end, was the mission already performed by VocBench for what concerns the collaborative development of such resources. These demands later led to the inclusion of the system (named in the meanwhile: “ShowVoc”) under the umbrella of the VocBench project (which has now been rebranded as “VocBench & ShowVoc”).

3 Related Work

We should probably start our discussion of related efforts on data portals with CKAN[5], which had established itself as a de facto standard, particularly in the public sector, with its rich API and support for catalogs federation. Within the scientific community, Zenodo[6] (based on the open-source software Invenio[7]) has established itself as the go-to solution for ensuring data persistence, similar to what arXiv[8] has achieved for preprint publication. Regarding the impact of archiving, the Open Archive Initiative [7] (OAI) is certainly of interest, especially for its metadata harvesting protocol (OAI-PMH).

[4] https://www.tbxinfo.net/
[5] https://ckan.org/
[6] https://zenodo.org/
[7] https://inveniosoftware.org/
[8] https://arxiv.org/

None of these solutions are specifically tailored to semantic web datasets, beyond the ability to store dumps as files. For this reason, we narrow down our focus on catalogs of semantic web datasets. LOV [8] is a catalog of Linked Data Vocabularies, while LOD Cloud[9] hosts, in addition to its eponymous figure, a catalog of the datasets associated with the former. There are also domain-specific catalogs, most notably BioPortal [9] for ontologies related to the biomedical domain. Today, the OntoPortal Alliance [10] has taken over BioPortal's original source code, which is being adopted by portals across various domains, such as agrifood [11] and biodiversity and ecology [12]. Within the field of solid Earth science, we mention a European initiative [13] using metadata and semantic technologies for integration and access of data from diverse sources.

Alignment management, which is addressed by many Semantic Web catalogs, including LOV and OntoPortal, can also be a use case in its own right. For example, the Alignment API [14] ships with a server that can handle an ontology network, with the ability to compute, retrieve, combine, and otherwise manipulate alignments between ontologies. In a related vein, the ELEXIS [15] project aims at linking (legacy) language resources via linked data and has developed a standard REST API for accessing a catalog of dictionaries. Both applications are covered by ShowVoc, as we will see later.

We conclude the section on related work by discussing the publication of linked data. Pubby[10] implements resolvable URIs by querying a SPARQL endpoint. This software is now discontinued, but newer alternatives have emerged, such as LodView[11] and Loddy[12] The triple store Virtuoso [16] has even integrated this feature without the need for third-party software. Subject pages were even took as a paradigm for data editing, in systems such as TemaTres [17] or OntoWiki [18].

Subject pages are not always the best choice for browsing through your data. For example, SKOSMOS[13] became a popular choice for publishing a collection of SKOS thesauri with more sophisticated browsing capabilities, including search and indexing. For ontologies, the need for more organized documentation became apparent. This can be automatically generated from the ontology definitions themselves using tools such as LODE [19] or, more recently, WIDOCO [20].

4 Requirements

The design of ShowVoc has followed three main driving missions and their associated sets of requirements: 1) the original PMKI mission, 2) the role of a publication companion to the VocBench platform 3) the design “at large” of a Semantic Web data dissemination platform.

As mentioned earlier, the contribution of PMKI to ShowVoc, as of its original embodiment in such project, has been the mission to cover lexical and translation resources. This coverage has not been merely limited to showing and allowing browsing

[9] https://lod-cloud.net/
[10] https://github.com/cygri/pubby
[11] https://github.com/LodLive/LodView
[12] https://bitbucket.org/art-uniroma2/loddy
[13] https://skosmos.org/

of such resources. Due to its nature of one stop shop for authoritative, domain contextualized, term translations, the system has to provide efficient APIs (and an associated UI) performing cross-language retrieval across all of its hosted resources.

The association with the VocBench project brought in several of its already assessed requirements, which have been taken immediately on-board. Of the 15 requirements for VocBench listed in [3], several of them were reconsidered as-is, such as *R1. Multilingualism, R5. Data Scalability, R7. Adaptive Context and Ease-of-use, R8. RDF Languages Support, R9. Maintainability (Architecture and Code Scalability*. For sake of conciseness, we refer the reader to [3] for the full description of these requirements, while giving a broad overview of them here: R1. *multilingualism* concerns with the multilingual representation (fundamental in the multicultural and multilingual European scenario) on three different layers: user interface, content editing and content representation. *Data scalability (R5)* is concerned with the capability to manage even non-trivial amounts of data – such as those given by large thesauri or datasets – still with convenient, user-friendly, interfaces. *R7* deals with providing, on a convention-over-configuration approach, a smooth installation and configuration approach satisfying different palates, from users willing a click&go desktop tool to large organizations requiring complex deployment scenarios. *RDF Languages support* (R8) deals with the aim of providing dedicated support for different types of knowledge models, embracing ontologies, KOSe (thesauri, authority lists, terminologies..), lexicons and any combination of them. *Maintainability (R9)* covers the ability to meet new requirements, cope with changed environments and make future maintenance easy, without requiring continuous intervention on the system.

Mutatis mutandis, other requirements such as: *R3. Data Interoperability* and *Integrity* (enforcing correct use of standards and being robust with respect to non-strictly adherent inputs) *R4 Software Interoperability/Extensibility* (implementing and exploiting LOD best practices, other than standards for finding and accessing data on the web plus the possibility to have extensions for interacting with other systems), *R6. Under-the-hood data access/modification* (provide different layers for data access, from high level UIs to triple-oriented data inspection and SPARQL querying), *R11. Provenance* (proper representation of actions leading to the dataset as-is and of its constituents), *R15. Everything's RDF* (RDF should be used not only for the mere data; rather, as much as possible, for metadata both at the content and application level) have kept their general intension while being repurposed on ShowVoc's different mission, which does not include editing but encompasses dissemination and fruition to a greater extent (e.g. just to cite one aspect: paying particular attention to representing dataset relations other than modeling each dataset independently). Similarly, other requirements that were initially conceived with (collaborative) editing in mind, such as: *R12. Versioning Support* (originally dedicated to the creation of versatile data dumps, supporting a time-machine within VocBench for jumping across stable versions of the system and providing diffing of resources across these versions), *R13. Dataset-level Metadata Descriptions* (useful for navigating through existing resources across different projects, resolving links, supporting alignment strategies, etc.) *R14. Customizable User Interface* (with the definition of forms with customized fields for the creation of different types of individuals, later repurposed in two separated concepts: construction forms and customized

views, the latter making sense even in ShowVoc) have been rethought in terms of data provision and of the PMKI mission of a one-stop shop for accessing authoritative multilingual translations, interweaving LOD policies with global activities (e.g. global search, navigation of dataset relationships, translation API) involving both the whole set of hosted datasets and Linked Open Datasets on the Web.

Two requirements: *R2. Controlled Collaboration* and *R10. Full Editing Capability (RDF Observability and Reachability)*, being strictly connected to the development of semantic resources, have been discarded from the set of ShowVoc requirements.

The broader picture that was first introduced by the role opened by the PMKI project and later by the more mature view, at large, of ShowVoc as a Semantic Web dissemination platform, introduced two further requirements, also supported by concrete use cases and scenarios:

R16. Community Involvement. While delegating to other platforms (e.g. VocBench) the editing and maintenance of datasets, a dissemination platform must at least provide for the possibility to introduce new datasets and to update them. A requirement introduced since its first incarnation as PMKI is to allow for contributions from the community, in terms of mere reporting of existing resources, addition of their metadata to the registry up to considering the ingestion of their data.

R17. Serving as a thorough Open Dataset Catalog. Providing browsing facilities for hosted datasets may be a prerogative of some dissemination platforms, while others focus on acting as mere registries for facilitating the discovery of (and access to, when directly hosted) datasets. Indeed, serving all these tasks together should not be seen as a lack of focus: a catalog could act as a powerful dataset registry, ranging from locally hosted datasets (and distinguishing for each dataset whether it is the official loci for its distribution or a mirror, reporting as metadata where the official distribution can be found) to linked open datasets on the Web, providing for all of them seamless navigation optimized for each dataset, locally thanks to configurable solutions or, more in general, even by means of advanced queries informed by dataset metadata.

R18. Serving dataset through Linked Open Data publication standards and best practices. A number of RDF publication systems (see the Related Work section) implements Linked Data principles for disseminating Linked Open Datasets, covering aspects such as metadata discovery and provision by “following your nose”, content resolution and negotiation. Many of these systems focus on implementing the core resolution aspects, thus facilitating the provision of different media (e.g. different types of RDF serializations for data-oriented clients and HTML Web pages for Web browsers) for different consumers, disburdening deployers from most of the service configuration. Indeed, there is often often a neat separation between such tools and advanced browsing platforms, thus requiring a combined installation when both are needed, yet with waste of overlapping features. Serving Linked Datasets by following LOD principles should be indeed a feature of any advanced system.

R19. Integration within Data Portals. Data Portals often come with own presentation styles, presentation pages, logos, links to accompanying information, all aspects that deviate from the specific mission of a – however general – disseminaton platform. It is out of the scope of dissemination platforms (e.g. a catalog as described in R16) to

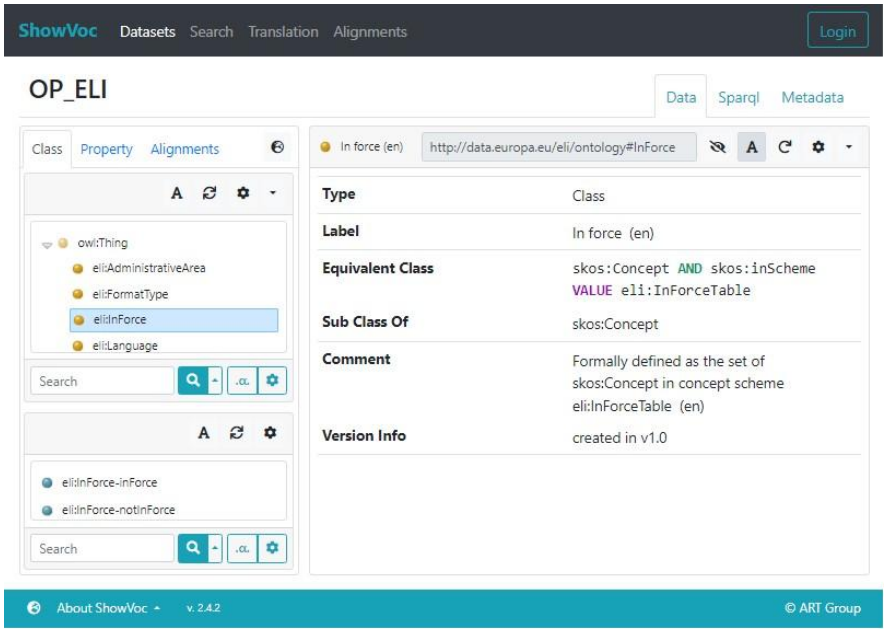

Figure 1: ShowVoc data view displaying the ELI (European Legislation Identifier) ontology.

satisfy all the requirements that a given portal may have. On the other hand, portals cannot and should not reinvent the wheel by replicating all the advanced functionalities that a Dataset Catalog may feature. Indeed, it should be upon catalogs to be integrable into portals in a way that keeps intact the presentation style of the latter.

R20. Provision of integrated material. An omni-comprehensive provision of data and information about a dataset should facilitate the production of associated material, such as derivative/associated forms of data (e.g. linksets [6] or ontolex lexicalizations [5]) and additional media (e.g. dataset documentation, diffing reports, images).

5 Architecture

ShowVoc has been designed as a single-page application, its frontend running inside a web browser that communicates with a back-end server through REST-like API.

The frontend is developed in TypeScript using the Angular framework and can be delivered to users by any web server or CDN (Content Delivery Network).

The backend server is based on Semantic Turkey [21], the same RDF services platform that powers VocBench 3. The platform, currently based on an opinionated combination of the Spring Boot[14] and PF4J[15] frameworks, supports the development and publication of services related to RDF data. Prebuilt services address multiple models and various concerns such as history, validation, and import/export. PF4J makes it easy

[14] https://spring.io/projects/spring-boot
[15] https://pf4j.org/

to deploy new services and extend the capabilities of existing ones by providing implementations of the extension points on which they depend. For example, the *export service* defines extension points that can be used to provide both the conversion logic to a particular serialization format (i.e., reformatting exporter) and the ability to deploy data to particular targets (i.e., deployer). Semantic Turkey features twelve extensions points making defining an equivalent number of extensible functionalities[16] and ships with implementations of these extension points for common use cases, but (as mentioned) new ones can be added to the system (see req. R4).

Semantic Turkey relies on the RDF4J framework to process RDF data and interact with triple stores (i.e., RDF database management systems), both in-process within Semantic Turkey or managed as separate processes. The latter option is the preferred method, allowing the use of enterprise-grade triple stores such as Ontotext's GraphDB[17]. Further aspects related to the architecture of the backend system and in particular concerning req. R9 have been widely discussed in [3].

6 Features Overview

We will introduce here ShowVoc and its main features, discussing relevance to the system requirements. ShowVoc is made available as a self-installing platform requiring only the existence of a Java Virtual Machine (v. >= 17 at the time of writing) within the running environment. The installation, in its core form, is very simple (R7) and limited to running the system which, upon its first invocation, will automatically deploy and initialize its data directory and ask the user to configure themself as administrator. As mentioned in the previous section, an external triple store can be optionally connected, with the possibility to use solutions for high data scalability (R5).

ShowVoc presents itself to the user with a (customizable) landing page and a menu for accessing its main areas: *dataset*, *search*, *translations* and *alignments* providing dedicated access to the hosted contents. The *dataset* entry opens up to a list of hosted datasets, which is in turn exploded onto different views (*data*, *sparql*, *metadata*, *downloads*, *tools*) offering targeted exploration of the selected dataset. The main view, *data* (see **Figure 1** for an overview) offers a split page with a data structure on the left (usually a tree, even though it depends on the nature of the dataset on user preferences) for searching and browsing through the dataset and a "resource view" on the right, a component exploding the description of each single selected resource within the dataset. This explosion mainly consists on a convenient visual representation of the selected resource through a series of predicate-object lists (i.e. groups of values aggregated under a same property, each value representing – with its aggregating property and the main described resource – a triple in the dataset), further aggregated into sections dedicated to similar properties (e.g. all properties dedicated to lexicalizing the dataset). This generic pattern is then structured with different sections according to the resource types defined in the core modeling vocabularies (RDF, RDFS, OWL, SKOS, SKOS-XL, OntoLex) having dedicated support in ShowVoc. Certain sections are common to

[16] https://semanticturkey.uniroma2.it/doc/dev/#extension_point_development
[17] https://www.ontotext.com/products/graphdb/

all resource types (e.g. the “type” section) while others depend on their nature (e.g. skos concepts have a scheme section detailed their membership to different schemes of the KOS they are part of). Furthermore, specific values, depending on their type and/or nature, can have their description produced by further expanding the exploration of the graph (e.g. blank nodes of type owl:Restriction can be expanded in convenient representations in Manchester syntax of the axioms they represent). The user interface is conveniently made adaptive with respect to the size (upon different dimensions) of the dataset being explored; e.g. a big thesaurus with a very large number of top concepts or even total absence of a hierarchy will see the user requested to switch to a search-based UI instead of merely representing all concepts in an unmanageable huge flat list; similarly, a lexicon could be browsed through 1 or 2 character indices or, again, forced into a search based interface, depending on the average size of these indices. Data scalability (R5), balancing functionalities and performance, is thus guaranteed even under the perspective of user experience.

Multilingual UI and content access and filtering. The platform features a multilingual (R1) User Interface, extensively supported in five languages: English, German, Italian, French and Spanish (the first three being maintained internally by the team, the latter two supported by the community, with more languages coming) and covering a range of 5000+ lexical entries (counting both mere labels of UI components and more elaborated content). Multilingualism is widely supported also at the content level, by allowing users to fine-tune the lexical content (e.g. labels, definitions and notes) they want to view through language selection, as well as the languages used for visually rendering the single semantic resources. Other forms of rendering are also accepted and configurable (e.g. concatenating the `skos:notation` with the existing labels and arbitrary formatting characters).

Contributions. A feature inherited from its first incarnation as PMKI, the ShowVoc dataset portal can optionally allow for contributions from community users (R16 and, in a sense, the initially discarded R2). These can request the addition of a new dataset in terms of an entry in the internal registry of known datasets or by hosting the dataset fully. This second option allows, in turn, for direct uploads of content, conversion non-RDF formats and the availability of a development environment within an associated VocBench instance. Whichever case, visiting users must perform a registration, providing relevant information about the dataset and the contribution (to any extent) must be vetted by a ShowVoc administrator, both in its final form (ready for publication) and its intermediate contribution in case of a resource to be converted, processed, edited.

Resource Resolution and Content negotiation. ShowVoc's use cases extend beyond cataloging third-party datasets, as it also addresses the needs of original dataset publishers (R18). These can set up ShowVoc as an advanced browser for their datasets; however, Linked Data rules require that entity identifiers be resolvable via HTTP, which is perhaps the defining characteristic of the Linked Data paradigm. ShowVoc supports this as well, with some endpoints that can be queried by a reverse proxy associated with the domain to implement content negotiation and generate different variants, including machine-readable serializations and human-friendly pages.

Multi-model support. ShowVoc manages arbitrary RDF datasets (R8), adding dedicated support for OWL ontologies, SKOS Knowledge Organization Systems and OntoLex/Lemon lexicons. In addition, ShowVoc is aware of various lexicalization models for grounding data in natural language, including RDFS labels, SKOS(-XL) terminological properties and OntoLex-Lemon lexical entries. At the user interface level, this flexibility is first visible in ShowVoc's resource view, which can display the description of any resource, divided into sections that roughly correspond to different properties. As such, the resource view can display any type of resource, but it can be specialized and made efficient for specific modeling vocabularies through a combination of customized templates (defining the prominent sections for different resource types), specialized sections (e.g., the one grouping class axioms), as well as dedicated support for specific mechanisms (e.g., proper rendering of class axioms in Manchester syntax). ShowVoc works seamlessly with different lexicalization models, which are taken into account when selecting the "labels" for displaying a resource (instead of its IRI or qname) or when populating the "lexicalizations" section of the resource view (which abstracts over the specific lexicalization model). ShowVoc also provides a number of views to browse the content of the dataset, depending on its nature, such as a class tree, instance list, property tree, concept tree, etc.

Dataset and vocabulary-level metadata. Proper metadata is considered critical for publishing a dataset according to the FAIR principles [23]. As such, ShowVoc manages a complete description for each dataset, including general metadata (e.g., title), customizable facets (e.g., category, organization), access metadata (e.g., SPARQL endpoint), structural metadata (e.g., URI space), and various metrics that provide insight into the richness of the available content both at the conceptual level (e.g., number of classes, concepts, etc.) and at the lexical level (i.e., regarding the degree of coverage of different natural languages). These metrics can be visualized as a table or as a chart of various types. Different versions of a dataset (R12) are also represented across multiple dimensions, encompassing different releases, different distributions, special evolving versions (e.g. the always-evolving linked open dataset providing resolution for its resources, or the master copy within a management system, if reachable) and combinations thereof (e.g. a certain distribution of a given release of a dataset).

ShowVoc manages the dataset descriptions using the Metadata Registry (MDR) component of Semantic Turkey. This in turn manages the available datasets as a DCAT [24] catalog, using a metadata profile based on a combination and interpretation of existing metadata vocabularies (e.g. DCTERMS, FOAF, VoID [6], LIME [5]) together with a small ontology addressing concerns (mostly related to access) not covered by the former. Such descriptions achieve the mission of robust data interoperability (R3) by enabling autonomous machine-2-machine communication, where most of the dataset metadata descriptors (R13) are targeted towards machines with the objective of optimizing content fruition and exploitation, taking into account the co-existence of several overlapping standards and clarifying, at the meta-level, how any dataset is modeled and structured. Data provenance (R11) is accounted for, both inside each dataset (separating, for instance, imported ontology vocabularies or different modules in multi-module datasets) and in the MDR (containing graph-distinct descriptions of all the hosted and

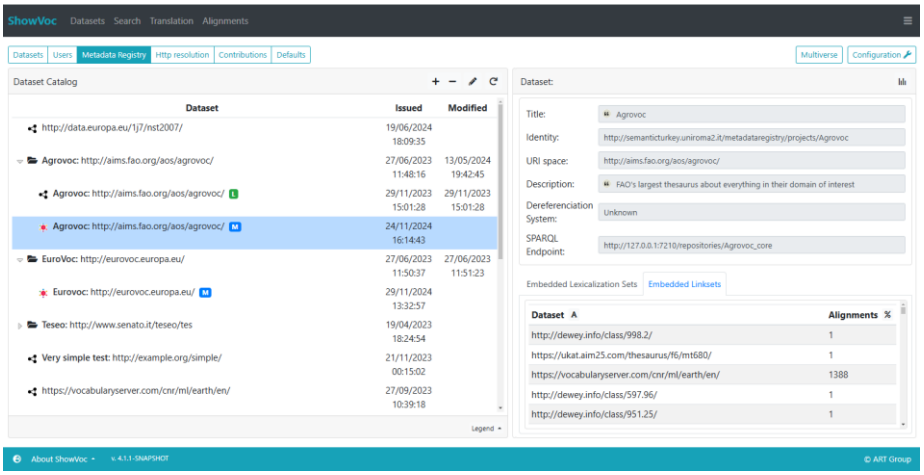

Figure 2: The Metadata Registry

known datasets). In figure 2 the view on the MDR is shown, with a few versions of the same datasets reported in a hierarchical view.

Seamless navigation between local and remote datasets. A user who encounters a reference to a resource outside the dataset being browsed can easily jump to it. If the resource belongs to another dataset in the same ShowVoc installation, the user interface automatically switches to that dataset and focuses on the target resource. External resources can also be displayed in a modal dialog populated with information retrieved by deferecentiation or a SPARQL endpoint, if known to the system by following an optimized view informed by the MDR (R17)

SPARQL querying. ShowVoc provides a SPARQL editor with syntax highlighting and completion that can be used to query individual datasets. It also supports federated queries involving other hosted datasets or remote SPARQL endpoints. The former can be done more efficiently if the chosen triple store is GraphDB, which has specific optimizations for local federation. (R6). Results can be downloaded in a variety of formats, and queries can be loaded from different scopes, e.g., system scope for general purpose queries or dataset-specific queries. Forthcoming evolutions include the possibility to user SPARQL for querying the MDR as well, in order to benefit from its full range of information in RDF (R15) without the mediation of service API.

Global Search. ShowVoc allows users to perform a full-text search across all hosted datasets. Matched entities, grouped by dataset, are displayed with their labels, IRIs, and `skos:note` specializations (which include definitions, examples, etc.).

Translation. Similarly to global search, translation allows users to lookup terms inside the datasets in the catalog, searching for a translation in one or more natural languages.

Alignments. ShowVoc keeps track of the alignments contained in each dataset, providing both a per-dataset and a global view of these alignments.

The former consists of an expanding tree whose roots are the datasets directly aligned with the current datasets. These can be expanded to show the datasets to which they are aligned, and thus to which the current dataset is indirectly aligned. Each node

is decorated by the number of links: when one of them is clicked, ShowVoc displays a paginated list of the correspondences (with some filtering features).

Global visualization of alignments is supported by a similar tree view as well as by graph visualization: nodes represent datasets and edges represent alignments. By clicking on a node or an edge, users can view metadata about a dataset or an alignment.

Distributions. ShowVoc provides multiple distributions of each dataset together with related materials (R20 and, again, R12). In the first place, these distributions can be used to provide downloadable data dumps of a (given version) of the dataset, together with the aforementioned alignments and any sort of files, including documentation.

Custom views. ShowVoc main resource exploration view, the "resource view", can be further customized to consider specific values requiring to be described by looking ahead in turn on their outgoing triples and further on along the graph, by performing several hops. This case can be summarized to resources i) being expressed through a URI (and not, for instance, through a blank node) for matters of resolvability on the Web but ii) not being considered totally independent, rather part of the global description, of other resources. An exemplification (not covering the full range of relevant cases though) is given by all reified lexical descriptors: e.g. skosxl labels and skos reified definitions represent lexical descriptors that, for the sake of being described in turn by metadata (e.g. providing the source or author of a definition) or in order to make them relatable to other resources (e.g. through `skosxl:labelRelation`) are modeled as IRI resources with a dedicated property pointing to their lexical content. ShowVoc provides customizable views (R14) able to specify the further exploration of the graph that is necessary in order to properly represent them, so that a concept decorated with a reified `skos:definition` does not show it as a linked IRI and shows instead its lexical representation. Further customization possibilities are possible, e.g. by showing tables (thus exploding the description of the linked value by showing more of its linked-in-turn characteristics through different fields of the table) and other types of widgets, such as maps (for geographical representation of points, areas and routes) and statistical widgets dedicated to series and series collections.

Multiverse and Embeddable views. ShowVoc can be embedded and integrated into hosting data portals (R19) by means of different parameterizations. The parameterizations allow for different "cuts" of the application to be embedded into different hosting pages and to be invoked separately. An additional feature, the "multiverse", allows for the possibility to spawn entire new copies (universes) of all the configuration space, including system, and project scoped settings, preferences and configurations. These new universes can give life to totally remodeled user interfaces, interaction modalities, presentation settings, etc.. that can be associated to a different view. For instance, in the instance of ShowVoc hosted by the European commission, some directorate generals requested for a more "narrative" (less of a "data browser" and more like a traditional web page) presentation of their thesauri, with properties being relabeled according to a given terminology adopted in their specific communities, headers and descriptions for sections, etc.. (all customization options available in ShowVoc, R14) while, in an embedded view that has been provided within the EU Vocabularies Portal[18], the thesauri

[18] https://op.europa.eu/en/web/eu-vocabularies

and ontologies have to be mandatorily presented in a data-centric, neutral way, using standard properties (and adopting their official names), thus a new universe has been created and associated with that view. The combination of embeddable views and the multiverse thus allow for very distinct experiences that can be associated with a single instance of the system, avoiding duplicated deployment and maintenance costs.

7 Impact and Comparison with State-Of-The-Art

First released in September 2021, ShowVoc is younger than its editing companion VocBench 3, which has become a reference platform for editing different types of semantic resources – ontologies, thesauri, lexica – since its launch in September 2017. Despite ShowVoc's relatively short history, we can point to some notable adopters. The Food and Agriculture Organization (FAO) of the United Nations adopted ShowVoc for the Caliper portal[19], which publishes statistical classifications as linked data; UN Statistical Department is following their lead with another dedicated instance. The Italian branch of LifeWatch ERIC – the European Research Infrastructure Consortium for biodiversity and ecology – used ShowVoc as a data publication platform supporting resolvable URIs. Last but not least, the Publications Office (OP) of the European Union (EU), which managed the development of the system, has deployed an instance of ShowVoc[20] "to support interested teams and professionals working for the EU institutions and agencies". As mentioned in the previous section. the Publications Office also integrated ShowVoc into the EU Vocabularies Portal[21] to provide an "advanced view" of the datasets content, complementing the portal's own capabilities.

Table 1 shows ShowVoc's coverage of the assembled requirements, comparing it to other relevant platforms. The purpose of this table is not to detail a thorough comparison of all the features exhibited by all platforms (in this sense, even ShowVoc itself is not adequately represented), rather provide a picture of whether and how some of the most representative state-of-the-art platforms addressed the said requirements.

The choice on which platforms to compare is restricted to those addressing linked datasets expressed using Semantic Web standards. The selection then fell on systems being either apical, domain independent, focal points for the discovery and retrieval of resources (e.g. LOV) or notable platforms adopted across different instances (in this sense, all of ShowVoc, OntoPortal and SKOSMOS sport some relevant adopters). Another characteristic of the selection is that the authors of this paper have first-hand experience with all of the systems, having been direct contributors of new features (on AgroPortal, with some features scheduled for being ported on the main project), of fixes and having hosted relevant instances of the platform (SKOSMOS for FAO) and having contributed content going through their contribution workflow (LegalHTML [25] and LIME on LOV).

Finally, the selected platforms offer different points of view upon the dissemination mission: LOV is focused on ontology vocabularies, offering discovery and access to

[19] https://caliper.integratedmodelling.org/caliper/browse/showvoc/
[20] https://showvoc.op.europa.eu
[21] https://op.europa.eu/en/web/eu-vocabularies

Table 1. Comparison matrix of relevant platforms from the SoTA wrt provided requirements

Req/Platform	ShowVoc	OntoPortal	SKOSMOS	LOV
R1. Multilingualism	config/5 UI langs	AgroPortal, content	mono/13 UI langs	No
R3. Robust Data Interoperability	Metadata is exploited to optimize content representation. Robust/complete interpretation of all formats	Datasets to be loaded have constraints at times stricter than the specifications	No	No
R4. SW Interop./Extens.	PF4J for extensions. Several connectors to other technologies/services	Google Analytics ReCaptcha, no plugin support	No	No
R5. Data Scalability	RDF4J middleware. Compl enterprise-grade triple store: GraphDB. Lucene indexing for cross-dataset retrieval	GOO object-triple mapping framework, persistence: 4store or AllegroGraph (AG), Redis cache, Solr for indexing	Generic SPARQL endpoint for access (sugg. Jena Fuseki). No optimization. (suggested use of a cache, e.g. Varnish)	Native: MongoDB SPARQL access through Jena Fuseki on RDF Export. Elastisearch for indexing
R6. Raw data manipulat.	SPARQL UI and Endpoint	SPARQL endpoint	No	No
R7. Adaptive Context and Ease-of-Use	Quick click&run setup. Easily scalable to diverse scenarios	Non-trivial installation, partially simplified by the release of a virtual appliance	Non-trivial config. of Fuseki spec. for SKOSMOS	
R8. RDF langs support	RDFS, OWL, SKOS, SKOS-XL, XKOS (partial), OntoLex	RDFS, OWL, SKOS, SKOSXL (agroportal)	SKOS (limited for SKOSXL, requiring SKOS labels)	RDFS, OWL
R11. Provenance	Data provenance	Dataset provenance,	No	No
R12. Versioning	Yes. Can load diffs produced by VB	Yes, with support for diffs	No	Yes
R13. Dataset Metadata	DCAT 3, VoID, Lime, VoAF + own	MOD 2.0 (DCAT 1 + OMV + VoID)	No	VOAF
R14. Customizable UI	Cust.Views + resview configurations	No	No	No
R16. Community Involv.	Contributions through links to published resource, uploading a RDF dataset, obtaining a workspace for importing/editing it	No	No; even internal admin. is not through UI; it requires editing data on the triplestore	submission link to published ontology
R17. Open Dataset Catal. and browsing	MDR-powered optimized LOD Browser	Local only	Local only	Local Catalog, no content browsing
R18. LOD Serving	Support for HTTP Resolution, Content Negotiation	URLs for accessing single resources	URLs for accessing single resources	No, resources link to published ontologies
R19. Data Portals Integr.	Embed. Integr./ Multiverse	Widgets	No	No
R20. Integr. Material	Alignments and versions automatically produced, documentation automatically produced through VB, further material.	Alignment and diff computation (but no vetting support). Links in metadata to other resources	No	No

them more than in mediating their fruition; conversely, SKOSMOS has a very tight focus on browsing/searching locally to the hosted thesauri; no discovery nor cross-dataset functionality is foreseen.

OntoPortal is probably ShowVoc's closest neighbor; born for hosting ontologies (hence its name), OntoPortal has recently embraced SKOS vocabularies; it has a tight focus on FAIR principles (e.g. it includes a FAIRness Evaluation tool) whereas ShowVoc is more adherent to embodying all Linked Open Data standards and best practices which, ultimately, imply meeting FAIR principles for semantic resources. In the table, we have skipped those requirements (from the original list for VB) that were not relevant for ShowVoc (R2 and R10) and those not particularly discriminating or observable among the presented systems (R9, R15).

8 Sustainability and Future Work

ShowVoc is currently supported mainly by the DIGITAL Programme[22] of the EU (see acknowledgements) and by a few user-parties interested in the evolution of the system. The maintenance roadmap includes at least two major releases each year and other minor/bugfixes releases on need. Two Google groups: user and developer (joined with those of VocBench) guarantee free support to a vibrant community that gathered around them: these groups now benefit from the active participation of users helping each other, thus not centralizing the role of developers. Furthermore, once in a while, a in-person event called "VocBench & ShowVoc Community Meeting" is held at the Publications Office of the EU, in Luxembourg, with active workshops for brainstorming new ideas and gathering desiderata directly from the user community. Participation is wide (in the range of >50 participants) and has usually to be contained due to logistics.

From its inception as PMKI to its evolved role of VocBench companion for publication and dissemination, ShowVoc has stretched far beyond the boundaries that define many similar initiatives. The next quantum leap is represented by gaining the role of centralizing data hub in large federations of data portals, thanks to an action carried on within ITINERIS[23]: Italian Integrated Environmental Research Infrastructures System, a project funded by the European Union – Next Generation EU PNRR- Mission 4 "Education and Research" – Component 2: "From research to business" – Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures", for building a network of Italian Environmental Research Infrastructures. This action will fund several improvements revolving around this data hub role, including the realization of several connectors to other existing dataset catalogs, enabling distributed search over connected data catalogs while using selected harvesting of selected resources for enabling ShowVoc advanced browsing and exploration capabilities.

Further directions include broadening the dedicated support for core modeling vocabularies (e.g. extending its support for XKOS statistical classifications), improving the publication workflow from VocBench to ShowVoc and further exploiting its linking metadata to broaden its possibilities as an authority-based translation system.

9 Conclusion

In this work, we have introduced ShowVoc, a web-based multilingual platform for publishing and consulting OWL ontologies, SKOS(-XL) thesauri, Ontolex-lemon lexicons and generic RDF datasets. ShowVoc features and impact on the dissemination of linked open datasets and its positioning within the State of the Art have been discussed. We hope this work will not only provide insights on this and related existing resources, but contribute to foster research and industry efforts towards the realization of progressively more advanced solutions for data discovery and interoperability.

[22] https://digital-strategy.ec.europa.eu/activities/digital-programme
[23] https://itineris.cnr.it/

Acknowledgements. ShowVoc has been originally designed by the University of Rome Tor Vergata and is now maintained and evolved by Lore Star srl in the context of the Digital Europe Programme, under management of the Publications Office of the EU in a provision contract with European Dynamics.

References

1. Berners-Lee, T., Hendler, J.A., Lassila, O.: The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American 284(5), 34-43 (2001)
2. Berners-Lee, T.: Linked Data. In: Design Issues. (Accessed 2006) Available at: https://www.w3.org/DesignIssues/LinkedData.html
3. Stellato, A., Fiorelli, M., Turbati, A., Lorenzetti, T., van Gemert, W., Dechandon, D., Laaboudi-Spoiden, C., Gerencsér, A., Waniart, A., Costetchi, E., Keizer, J.: VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons. Semantic Web 11(5), 855-881 (Jan 2020)
4. Cimiano, P., McCrae, J.P., Buitelaar, P.: Lexicon Model for Ontologies: Community Report, 10 May 2016. Community Report, W3C (2016) https://www.w3.org/2016/05/ontolex/.
5. Fiorelli, M., Stellato, A., Mccrae, J.P., Cimiano, P., Pazienza, M.T.: LIME: the Metadata Module for OntoLex. In Gandon, F., Sabou, M., Sack, H., d'Amato, C., Cudré-Mauroux, P., Zimmermann, A., eds. : The Semantic Web. Latest Advances and New Domains (Lecture Notes in Computer Science). Proceedings of the 12th Extended Semantic Web Conference (ESWC 2015), Portoroz, Slovenia, May 31 - 4 June, 2015. 9088. Springer International Publishing (2015), pp.321-336
6. Fiorelli, M., Stellato, A., Rosati, I., Fiore, N.: Process-Level Integration for Linked Open Data Development Workflows: A Case Study. In Garoufallou, E., Vlachidis, A., eds. : Metadata and Semantic Research (Communications in Computer and Information Science). 16th Research Conference, MTSR 2022, London, UK, November 7–11, 2022, Revised Selected Papers. 1789. Springer, Cham (2023), pp.148-159
7. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets with the VoID Vocabulary (W3C Interest Group Note). In: World Wide Web Consortium (W3C). (Accessed March 3, 2011) Available at: http://www.w3.org/TR/void/
8. Lagoze, C., Van de Sompel, H.: The open archives initiative: building a low-barrier interoperability framework. In : Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke Virginia USA, June 24-28, 2001, pp.54-62 (2001)
9. Vandenbussche, P.-Y., Atemezing, G.A., Poveda-Villalón, M., Vatant, B.: Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. Semantic Web 8(3), 437-452 (December 2016)
10. Salvadores, M., Alexander, P.R., Musen, M.A., Noy, N.F.: BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. Semantic Web 4(3), 277-284 (2013)
11. Jonquet, C., Graybeal, J., Bouazzouni, S., Dorf, M., Fiore, N., Kechagioglou, X., Redmond, T., Rosati, I., Skrenchuk, A., Vendetti, J.L., Musen, M., Alliance, m.o.t.O.: Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology. In Payne, T.R., Presutti, V., Qi, G., Poveda-Villalón, M., Stoilos, G., Hollink,

L., Kaoudi, Z., Cheng, G., Li, J., eds. : The Semantic Web – ISWC 2023 (Lecture Notes in Computer Science). 22nd International Semantic Web Conference, Athens, Greece, November 6–10, 2023, Proceedings, Part II. 14266. Springer, Cham (2023), pp.38-58

12. Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé Yeumo, E., Emonet, V., Graybeal, J., Laporte, M.-A., Musen, M.A., Pesce, V., Larmande, P.: AgroPortal: A vocabulary and ontology repository for agronomy. Computers and Electronics in Agriculture 144, 126-143 (2018)
13. Kechagioglou, X., Vaira, L., Tomassino, P., Fiore, N., Basset, A., Rosati, I.: EcoPortal: An Environment for FAIR Semantic Resources in the Ecological Domain. In : Proceedings of the Joint Ontology Workshops 2021. Episode VII: The Bolzano Summer of Knowledge. co-located with the 12th International Conference on Formal Ontology in Information Systems, and the 12th International Conference on Biomedical Ontologies, vol. 2969 (2021)
14. Bailo, D., Paciello, R., Michalek, J., Cocco, M., Freda, C., Jeffery, K., Atakan, K.: The EPOS multi-disciplinary Data Portal for integrated access to solid Earth science datasets. Scientific Data 10 (2023)
15. David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C.: The Alignment API 4.0. Semantic Web Journal 2(1), 3-10 (2011)
16. McCrae, J.P., Tiberius, C., Khan, A.F., Kernerman, I., Declerck, T., Krek, S., Monachini, M., Ahmadi, S.: The ELEXIS interface for interoperable lexical resources. In : Proceedings of the sixth biennial conference on electronic lexicography (eLex). eLex 2019 (2019)
17. Erling, O.: Virtuoso, a Hybrid RDBMS/Graph Column Store. Data Engineering Bulletin 35(1), 3-8 (2012)
18. Gonzales-Aguilar, A., Ramírez-Posada, M., Ferreyra, D.: TemaTres: software para gestionar tesauros. El profesional de la información 21(3), 319-325 (2012)
19. Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S.: OntoWiki – An authoring, publication and visualization interface for the Data Web. Semantic Web 6(3), 215-240 (2015)
20. Peroni, S., Shotton, D., Vitali, F.: Making Ontology Documentation with LODE. In : Proceedings of the I-SEMANTICS 2012 Posters & Demonstrations Track, Graz, Austria, September 5-7, 2012, pp.63-67 (2012)
21. Garijo, D.: WIDOCO: A Wizard for Documenting Ontologies. In : The Semantic Web – ISWC 2017. ISWC 2017 (Lecture Notes in Computer Science), vol. 10588, pp.94-102 (2017)
22. Pazienza, M.T., Scarpato, N., Stellato, A., Turbati, A.: Semantic Turkey: A Browser-Integrated Environment for Knowledge Acquisition and Management. Semantic Web Journal 3(3), 279-292 (2012)
23. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K.,

Zhao, J., Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(160018) (2016)

24. Albertoni, R., Browning, D., Cox, S., Gonzalez-Beltran, A.N., Perego, A., Winstanley, P.: The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake. Data Intelligence, 1-37 (December 2023)
25. Stellato, A., Fiorelli, M.: LegalHTML: A Representation Language for Legal Acts. In Pesquita, C., Jimenez-Ruiz, E., McCusker, J., Faria, D., Dragoni, M., Dimou, A., Troncy, R., Hertling, S., eds. : The Semantic Web (Lecture Notes in Computer Science, LNCS). Proceedings of the 20th Extended Semantic Web Conference (ESWC 2023), Hersonissos, Crete, May 28 - June 1, 2023. 13870. Springer, Cham (2023), pp.520-537
26. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology 1st edn. 1. Morgan & Claypool (2011)