Semantic Web 0 (0) 1–13
IOS Press

How to Create and Use a National Cross-domain Ontology and Data Infrastructure on the Semantic Web

Eero Hyvönen
Aalto University, Department of Computer Science, Finland, and
University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Finland
Semantic Computing Research Group (SeCo) (http://seco.cs.aalto.fi)
E-mail: eero.hyvonen@aalto.fi

Abstract. This application report presents a model and lessons learned for creating a cross-domain national ontology and Linked (Open) Data (LOD) infrastructure. The idea is to extend the global, domain agnostic "layer cake model" underlying the Semantic Web with domain specific and local features needed in applications. To test and demonstrate the infrastructure, a series of LOD services and portals in use have been created in 2002–2022 that cover a wide range of application domains. They have attracted millions of users in total suggesting feasibility of the proposed model. This line of research and development is unique due to its systematic national level nature and long time span of some twenty years.

Keywords: Semantic Web, Linked Data, Web Services, Infrastructures, Portals

1. Extending the Layer Cake Model

The Semantic Web (SW) sees the Web as an interlinked collection of data (Web of Data) instead of only a space of interlinked hypertext documents, Web of Pages. The idea was proposed in the 90's by Tim Berners-Lee [1], and first recommendations for the SW[1], such as the Resource Description Framework (RDF), were developed before the millenium. The recommendations constitute the W3C "layer cake model" [2, 3] on top of XML, the lingua franca of the WWW, and lay out a new basis of shared semantics for interoperability of data. Founded on using first order predicate logic, the semantics of the SW [4] are independent of application domains and natural languages. This makes the model suitable for dealing with the versatile data underlying the Web.

To develop applications, the layer cake model is not enough: domain and application specific infrastructures based on shared W3C standards and best practices are needed, too. These can focus on specific domains, such as medicine, biology, cultural heritage, or geography on an international level. However, in practice one also has to deal with national level issues and data available that are represented using national languages, data models, vocabularies, and are created using conventions of local legacy systems. For example, Cultural Heritage (CH) data in different countries is often nationally specific calling for adapted local solutions for representing and using the data.

Most of the international infrastructure work [5] is focused on collaborations on particular application domains. In contrast, this paper concerns the question: *How to Create a National Cross-domain Ontology and Linked Data*

[1] https://www.w3.org/standards/semanticweb/

Infrastructure and Use It on the Semantic Web. This problem is addressed by presenting, discussing, and evaluating approaches and living laboratory experiments developed in Finland during 2002–2022. Presenting lessons learned in this particular endeavour is hopefully useful in a more general setting, as similar challenges are likely to be faced in other countries, too.

The paper is organized as follows: In section 2, elements needed for a national SW infrastructure are first introduced. The idea and lessons learned in developing a national ontology and a LOD infrastructure are then presented in sections 3 and 4, respectively. After this, applications of the infrastructures are discussed: as a proof-of-concept, a model is presented that has been used for creating a series of in-use data services and semantic portals that have had up to millions of users. Finally, contributions of the work are summarized and related works discussed. This application report presents the first consolidated account of this line of research and development, summarizing works reported before in some 480 papers available on the Web[2].

2. Elements of a SW Infrastructure

Fig. 1 depicts components that are arguable needed in the developing a national SW infrastructure—according to the experiences to be reported in this paper. The system is based on the domain agnostic W3C *Web Standards and Best Practices* (on the left below in the figure) of publishing Linked Data[3] [6]. *Data Models* for metadata [7] are needed for representing knowledge of different applications domains, populated by resources taken from shared domain *Ontologies and Ontology Services* for interoperability. The ontologies should be made openly available and easy to access for interoperability and re-use, based on shared ontology services/libraries; cf. [8] for a survey of such systems. In the same vein, data services for publishing LD datasets, preferably using, e.g., open Creative Commons licenses, are needed for making re-use of data possible and easy. Also *Applications* of Linked Data are part of the infrastructure connecting the system to its end users. For making all this possible, *Software Tools* are needed for aggregating the distributed heterogeneous data from legacy and other data silos involved, and for extracting and linking (disambiguating) entities and relations from data records and textual descriptions [9]. Also tools for data publishing and analysis are needed, as well as tooling for developing new applications for the end users. Here using the FAIR principles[4] for creating Findable, Accessible, Interoperable, and Re-usable (R1) data are recommended. Obviously, the FAIR principles above are compatible with the linked data principles[5] and best practices[6] of the W3C.

For developing, maintaining, and using the infrastructure in a sustainable way a *Human Infrastructure* is needed (on the left in Fig. 1), too. This involves, e.g., educating people about the technology, introducing SW courses in university curricula, and production of documentations and learning materials for the community using national languages. In the Finnish case, for example, online materials have been created[7], a Finnish text book about SW was produced [10], and hackathons[8] were organized on using the data and tools.

3. Ontology Infrastructure

In the early 2000's, the focus in SW research was on ontologies [11], arguably the "silver bullet" of the SW [12]. Accordingly, a series of projects called "FinnONTO" in 2003–2012 were conducted[9] in Finland.

---

[2]http://seco.cs.aalto.fi/publications/
[3]https://www.w3.org/TR/ld-bp/
[4]The FAIR principles: https://www.go-fair.org/fair-principles/
[5]https://www.w3.org/DesignIssues/LinkedData.html
[6]https://www.w3.org/TR/dwbp/
[7]See, e.g., the self-study video lecture course "Linked Data Technologies for Cultural Heritage and Digital Humanities: Introducing the Semantic Web in Video Lectures" at https://seco.cs.aalto.fi/teaching/sw-introduction/.
[8]E.g., as part of the Helsinki DH hackathon series:
https://www2.helsinki.fi/en/helsinki-centre-for-digital-humanities/helsinki-digital-humanities-hackathon
[9]https://seco.cs.aalto.fi/projects/finnonto/

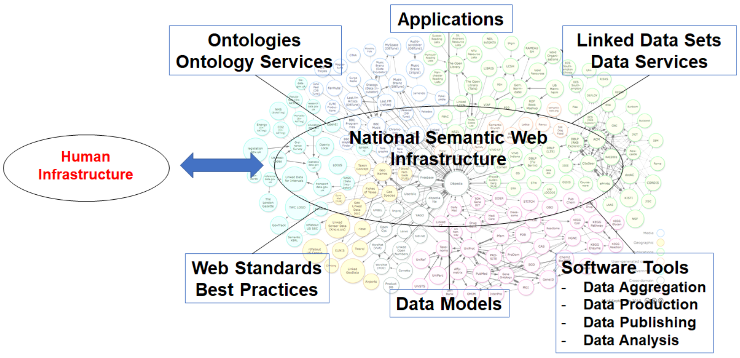

Fig. 1. Elements needed for a national Semantic Web infrastructure

A National Effort The goal in the FinnONTO initiative [13] was to develop a national ontology and content infrastructure, based on W3C standards, that would be cross-domain, multilingual (Finnish, Swedish, and English) and openly available. The consortium behind the initiative included finally some 50 companies and public organizations that represented a wide spectrum of functions of the society, including libraries, health organizations, cultural institutions, government, industry, media, and education. By integrating the interests of several functions of the society it was was possible to acquire substantial and long-standing funding for the infrastructure and application development work.

The initiative produced 1) metadata models contributing to national standards[10], 2) ontologies [14] to be used for populating the metadata models, 3) a living laboratory called ONKI of public ontology services [15], and 4) tools for metadata creation and application development, such as Skosify [16] for SKOS vocabulary quality assessment and the SAHA editor [17] for managing RDF repositories. The infrastructure was tested by using it in case studies in different application domains, including e-culture [18, 19], e-health [20], e-government [21], e-learning [22], and e-commerce [23].

A central goal of FinnONTO was to create an interlinked cloud of national ontologies [24] based on existing thesauri that were already used in different areas of the society. The rationale for this was that metadata available in national databases had already been catalogued using these thesauri, which would make it much easier to develop applications. According to the FinnONTO vision, the ontologies should be served not only through human readable browser interfaces, but also as centrally managed national ontology services using REST APIs. In this way, common functionalities of the services, such as (semantic) autocompletion [25] , URI fetching, and query expansion [26], could be shared on a national level, and everybody would get access to up-to-date versions of the ontologies. This kind of collaboration would be cost-efficient on a national level and gradually lead to better interoperability of the data catalogued in different organizations. Availability of the centralized services is needed especially for smaller organizations that do not have much expertise and resources for developing their own web services.

From Thesauri to Ontologies The FinnONTO project transformed 16 key national thesauri used in Finland into light-weight ontologies listed in Table 1. The transformation process was more ambitious than just transforming the traditional standard thesaurus format [27] into an RDF-based model. The thesauri were developed semantically a bit forward, using the OntoClean methodology [28] and RDFS, in the following ways [13]: 1) Multiple meanings of thesauri terms were disambiguated and relocated in `rdfs:subClassOf` hierarchies. For example, the concept of *child*, a unique concept in the underlying General Finnish Thesaurus YSA, can refer to the class of young people, to a family relation type, or a social class (superconcept of *street child*). These concepts should obviously be in located different places in an RDFS ontology. The concept was therefore split into several concepts in the corresponding General Finnish Ontology YSO [29]. 2) The thesauri that were transformed did not differentiate whether the standard Broader Term (BT) relation [27] means the part-of or hypernymy relation. This distinction was crafted manually in the ontologies. 3) The `rdfs:subClassOf` hierarchies were completed: all concepts were given at least one

[10]Such as the Public Recommendation for Geographic Metadata, Ministry of Internal Affairs, http://www.jhs-suositukset.fi/suomi/jhs158.

Ontology	Application Domain	# of concepts
YSO	General upper ontology	27 200
AFO	Agriculture and forestry	7000
JUHO	Government	6300
KAUNO	Literature	5000
KITO	Literary research	850
KTO	Linguistics	900
KULO	Cultural research	1500
LIITO	Economics	3000
MAO	Museum artifacts	6800
MERO	Seafaring	1300
MUSO	Music	1000
PUHO	Military	2000
TAO	Design	3000
TERO	Health	6500
TSR	Working and employment	5100
VALO	Photography	2000

Table 1

Linked ontologies of the KOKO cloud

superclass except the roots. 3) Inheritance of instanceship over subclass hierarchies was checked as specified by RDFS semantics, so that the hierarchies could be used for reasoning in, e.g., query expansion and when using faceted search in applications.

Linked Ontology Cloud KOKO The ontologies in Fig. 1 share lots of related and similar concepts [24]. The largest ontology YSO (27 200 concepts)—transformed from the most used thesaurus YSA in Finland of the National Library—shared lots of concepts with all other ontologies, in some cases more than 50%. This suggested that the ontologies should be linked together using YSO as the top ontology. This resulted in creating the Finnish linked ontology cloud called KOKO[11] where the top ontology concepts are refined by subconcepts of interlinked domain specific ontologies.

Lessons Learned A key problem to be solved in FinnONTO was that large cross-domain thesauri, especially the General Finnish Thesaurus YSA, could not anymore be maintained easily by its management team. Even if the team included people from different fields, the terminology related to specific areas needed deeper domain specific expertise than was available. Developing the interlinked KOKO ontology cloud mitigates the problem by distributing work on specific concepts to collaborative, domain specific ontology developer teams. However, in this model new problems arise pertaining to maintaining the linked ontology cloud and to coordinating the collaboration network [24]. These new challenges are now being tackled by the Finto collaboration network coordinated by the National Library. The FinnONTO initiative pointed out that lots of redundant work had been done in developing the thesauri in Finland as they shared lots similar concepts with each other. In the new, more coordinated KOKO model, redundant work can be better eliminated.

A challenge encountered in the ontologization process was that organizing the concepts into class hierarchies cannot in many cases represent correctly the meaning of the original terms that can be complex and fuzzy. The world cannot be represented fully using ontologies and there can be several ways in which this can be done. In spite of such challenges, the idea of adding more semantics needed for application development seems to be a better option that continuing using the original thesauri, whose semantics were too vague from a sofware development point of view in applications, such the semantic portals to be discussed in this paper. A strategic choice made in FinnONTO was to follow the wisdom articulated by Jim Hendler already in the late 90's in the SHOE project[12]: *A little semantics goes a long way.* In our case, the thesauri semantics were refined only a little using RDFS for interoperability and

[11] https://finto.fi/koko/fi/

[12] https://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html

to help development of web services. However, already this was a hand-full of work, as thousands of terms in the thesauri had to be manually checked and refined [29].

A mundane challenge of developing large vocabularies, at least in Finland, is how to convince the funding organizations, year after year, that this never ending work should be supported in a sustainable manner, not only as separate short-time projects. In our case, it took some ten years of project-based work before the KOKO ontology infrastructure and the current Finto.fi services could be funded in a more sustainable way by two Finnish ministries. The strategy taken in FinnONTO was to move forward in baby steps, and after each step show a demonstrator on how the ontologies can be applied in practise for creating something useful, such as the semantic Sampo portals [30] in use.

The idea of creating a “living laboratory” of ONKI ontology services [14, 15] on the Web turned out to be useful for deploying the infrastructure. The participating FinnONTO organizations were supported by the project in connecting their legacy systems to the APIs of ONKI for testing and evaluating the services. Finally, the “point of no return” was reached where pulling off the plug of the services was not an option anymore as the number of ONKI API users were counted already in hundreds.

The FinnONTO project series 2003–2012 started with a smallish one-year project, but eventually grew into a national effort of substantial size on the Finnish scale with tens of funding organizations involved. A reason for this was that in addition to public organizations, such as museums, libraries, and archives, also companies got interested in the technology, which convinced the main funding organization Tekes (called today Business Finland) that something useful and of monetary value is happening related to semantic web technologies. It is usually easier to get funding for technology development than for research in humanities.

The KOKO ontologies are based on keyword thesauri whose terms usually correspond to the classes. FinnONTO worked also on various “instance-based” ontologies, such as national geogazetteers, person and organization registries, biological taxonomies of species [31, 32], and nomenclatures and terminologies of medicine [20], such as Medical Subject Headings MESH[13]. Creating a national ontology infrastructure is a never-ending job and research goes on today, e.g., in the Linked Open Infrastructure for Digital Humanities initiative in Finland initiative[14] [33].

When developing ontology-based applications in FinnONTO, much of the time of the developers was “waisted” in cleaning and aligning the data from different organizations for interoperability. Obviously, it would be more cost-efficient do this work already when cataloging the data using ontology services. This would also enhance the quality of the linked data, which is a critical problem [34] on the SW. The local cataloguers know best their own data and should have the best interest in quality of their data. The motto for the FinnONTO work was therefore taken from a wisdom of Albert Einstein: *Intellectuals solve problems – geniuses prevent them*; a key goal of FinnONTO was to prevent interoperability problems rather than to solve them afterwards when the damage has already been done in cataloguing [35].

A major outcome of FinnONTO was the ONKI ontology server with its ontologies [15] that were publihied first in 2008. As a next step, the ONKI Light service[15] was developed and deployed in 2014 [36] by the National Library of Finland as the national Finto.fi service[16]. ONKI Light finally evolved in the open source Skosmos tool[17] in use in several other organizations in Finland and internationally[18]. ONKI Light was based on a SPARQL endpoint. This idea was to separate the data service fully from the user interface. This idea turned later useful when developing the Sampo model and Sampo-UI tool for semantic portals to be discussed later in this article.

The Finto.fi service has grown into a popular national open service. In 2019 it was used by 280 000 different users and its APIs were called 32 million times. The users include, e.g., museums, whose cataloging system get their keywords with URI identifiers from Finto. These developments suggest that the fundamental ideas of FinnONTO are feasible; they have actually made a paradigm change in Finland in developing and using linked light-weight ontologies on a national level instead of thesauri.

[13] https://finto.fi/mesh/fi/
[14] https://seco.cs.aalto.fi/projects/lodi4dh/
[15] https://seco.cs.aalto.fi/services/onkilight/
[16] Available at: https.//finto.fi
[17] https://skosmos.org/
[18] For a list of international services, see https://www.kiwi.fi/display/Finto/Skosmos-ohjelmisto0

4. From 5-star to 7-star Model

The SW infrastructure model of Fig. 1 includes a platform for publishing datasets and (re)using them via web services. A key component in LD publishing is the SPRAQL endpoint, but the platform should also support other functions [6]. The Linked Data Finland service LDF.fi[19] [37] was therefore developed in the "Linked Data Finland" follow-up projects[20] of FinnONTO.

LDF.fi has two user-groups: 1) For application developers, LDF.fi provides SPARQL endpoints and a suite of standard Linked Data (LD) services, including content negotiation, APIs for downloading datasets, LD browsing and editing, and additional tools for, e.g., data documentation and visualization. 2) For data publishers, the idea is to support and automate the data publishing process in the following way: The publisher creates a service description of the dataset and its schemas, using an extended version of the W3C Service Descripton recommendation[21]. Based on such metadata, LDF.fi then 1) automatically sets up the technical services, 2) generates a dataset "homepage" that explains the dataset, schemas, and 3) provides additional related services for querying, documenting, inspecting, and validating the data.

Linked data publications on the SW are typically evaluated with the W3C "5-star model"[22], using a quality scale analogous to evaluating hotels. In LDF.fi, the 5-star model is extended to a 7-star model: there are nowadays also a few 7 star hotels around[23]. The 6th star is given to a data publication if it includes not only the 5-star data but also the schemas of the data with documentation. This makes re-use of data easier. The 7th star is given to a data publication, if the publication includes some kind of evaluation that the data actually conforms to the provided schemas using, e.g., SHACL[24] or ShEx[25] [38]. The idea here is to encourage publishers to publish high quality data as data quality of LD is a severe issue on the SW.

Schemas can be documented automatically in LDF.fi for the human reader using a schema documentation generator, in our case SpecGen[26]. Datasets in the LD world often use schemas (vocabularies) for which definitions or descriptions are not available, but are embedded in the data itself. In order to find out how schemas are actually used in a dataset, including both published and unpublished schemas, a service vocab.at[27] was created that analyzes a given dataset from this perspective and creates an HTML document listing, e.g., statistics of vobaculary usage and raising up issues detected, e.g., if an IRI is not dereferenceable. The input for vocab.at is either an RDF file, a SPARQL endpoint, or an HTML page with embedded RDFa markup.

LDF.fi is implemented by a combination of the Fuseki SPARQL server[28] for storing the primary data and a Varnish Cache web application accelerator[29] for routing URIs, content negotiation, and caching. For deployment of applications with a data service (cf., e.g., the MMM system [39]) a microservice architecture with Docker containers[30] is used. Each individual component (the application, Varnish, and Fuseki) is run in its own dedicated container, making the deployment of the services easy due to installation of software dependencies in isolated environments. This enhances the portability of the services. The server environment of LDF.fi is provided by the CSC – IT Center for Science, a company of the Ministry of Education and Culture of Finland providing computational infrastructures for the national universities.

Lessons learned The Linked Data Finland platform has turned out to be useful for data-analytic research purposes and in developing applications (cf. Section 5). LDf.fi has been used for publishing some 100 linked datasets. Some of them are in use in semantic portal to application and via SPARQL querying combined with query editing and

[19]https://ldf.fi
[20]https://seco.cs.aalto.fi/projects/ldf/
[21]http://www.w3.org/TR/sparql11-service-description/
[22]https://www.w3.org/community/webize/2014/01/17/what-is-5-star-linked-data/
[23]Such as the Burj Al Arab in United Arab Emirates
[24]https://www.w3.org/TR/shacl/
[25]https://shex.io/
[26]https://bitbucket.org/wikier/specgen/wiki/Home
[27]http://vocab.at
[28]https://jena.apache.org/documentation/fuseki2/
[29]https://varnish-cache.org
[30]https://www.docker.com

scripting tools using the open CC BY 4.0 license. Some datasets are used only internally in related research projects, and for some datasets licensing policy of the data owners prohibits open use. LDF.fi hosts several instance-based ontologies, too, such as an RDF-version of the ca. 800 000 official Finnish geographical places based on data of the National Survey.

The LDF.fi service is still maintained by Aalto University and University of Helsinki that developed it on an academic project basis, but with the hope that some day it will be deployed and be maintained in a more sustainable way—this is at least what happened to the related ONKI/Finto ontology services. A step towards this is that in 2020 the concept of providing Linked Open Data services on a national level and LDF.fi were accepted on the new research infrastructure roadmap of the Academy of Finland as part of the larger initiative FIN-CLARIAH[31]. Here the idea is to combine—on a national level as in the CLARIAH initiative in the Netherlands—the work related to the pan-European infrastructure CLARIN[32], the research infrastructure for language as social and cultural data, and DARIAH[33], the infrastructure for arts & humanities scholars. In 2022, the work on creating the linked open data part of national FIN-CLARIAH infrastructure[34] infrastructure was started.

5. Sampo Model: Applying the SW Infrastructure

Table 2

Sampo Model Principles P1–P6

P1	Support collaborative data creation and publishing
P2	Use a shared open ontology infrastructure
P3	Make clear distinction between the LOD service and the user interface (UI)
P4	Provide multiple perspectives to the same data
P5	Standardize portal usage by a simple filter-analyze two-step cycle
P6	Support data analysis and knowledge discovery in addition to data exploration

When developing the Finnish SW infrastructure, applications testing and demonstrating its usability were constantly developed. This work evolved gradually into a general model for developing semantic portals, called the *Sampo Model*, and the *Sampo Series* of semantic portals and data services[35] [30]. The novelty of the Sampo model[36] lays in its attempt to formulate a set of re-usable design principles or guidelines for creating semantic portals, especially for Cultural Heritage applications and Digital Humanities research [40]. Based on six principles, the model is a kind of consolidated approach for creating LOD services and semantic portals, something that the field of the Semantic Web is arguably still largely missing [41].

The Sampo Model is an informal collection of principles for LOD publishing and designing semantic portals listed in Table 2. Principles P1–P3 can be seen as a foundation for developing data services; P4–P6 are related to creating semantic portals.[37]

The model is based on the idea of collaborative content creation (P1). The data is aggregated from local data silos into a global service, based on a shared ontology and publishing infrastructure (P2). The local data are harmonized and enriched with each other by linking and reasoning. In this model everybody can arguably win, including the data publishers by enriched data and shared publishing infra, and the end users by richer global content and services. The model argues for the idea of separating the underlying Linked Data service *completely* from the user interface via a

[31] https://www.kielipankki.fi/organization/fin-clariah/

[32] https://clarin.eu

[33] https://dariah.eu

[34] https://seco.cs.aalto.fi/projects/fin-clariah/

[35] See https://seco.cs.aalto.fi/applications/sampo/ for a complete list of "Sampo portals", videos, and further information.

[36] The model is called "Sampo" according to the Finnish epic Kalevala, where Sampo is a mythical machine giving riches and fortune to its holder, a kind of ancient metaphor of technology according to the most common interpretation of the concept.

[37] The model is called "Sampo" according to the Finnish epic Kalevala, where Sampo is a mythical machine giving riches and fortune to its holder, a kind of ancient metaphor of technology according to the most common interpretation of the concept.

SPARQL API. This arguable simplifies the portal architecture and the data service can be opened for data analysis research in Digital Humanities. For example, YASGUI[38] [42] interface for SPARQL querying and visualizing the results can be used, or Python scripting in Google Colab[39] and Jupyter notebooks[40] [43].

The general idea of principles P4–P6 is to "standardize" the UI logic so that the portals are easier to use for the end users and for the programmers to develop [44]. Principle P4 articulates the idea of providing different thematic *application perspectives* by re-using the data service. The application perspectives can be provided on the landing page of the Sampo portal system or be completely separate applications by third parties. According to P5 the application perspectives can be used by a two-step cycle for research: First the focus of interest, the target group, is filtered out using faceted semantic search [45–47]. Second, the target group is visualized or analyzed by using ready-to-use data analytic tools of the application perspectives. Finally, the Sampo model aims not only at data publishing with search and data exploration [48] but also to data analysis and knowledge discovery with seamlessly integrated tooling for finding, analysing, and even solving research problems in interactive ways (P6) [49].

The Sampo model principles are compatible with the FAIR principles for creating Findable, Accessible, Interoperable, and Re-usable data[41], but were developed in the context of publishing and using LOD. The principles P1–P6 can be used directly for creating semantic portals. However, its is also possible to apply them first to create an application domain specific framework and reuse it for developing different related application instances, which is arguably cost-efficient [50].

The Sampo model has evolved gradually in 2002–2022 via lessons learned in developing a series of semantic portals and LOD services, starting from **MuseumFinland – Finnish Museums on the Semantic Web**[42], **CultureSampo – Finnish Culture on the Semantic Web 2.0**[43] (online since 2009) [18, 19], and **BookSampo**[44] (online since 2011 with nearly 2 million annual users today) [51]. They demonstrated how CH content of tens of different kinds, both tangible and intangible CH content, can enrich each other. **WarSampo – Finnish World War II on the Semantic Web**[45] (online since 2015 with several new perspectives published in 2016–2019) [52] is a popular Finnish service that has had thus far over million users. A key idea in WarSampo is to reassemble the life stories of the soldiers based on data linking from different data sources. This biographical and prosopographical idea was a source of inspiration for several later biographical applications, including **BiographySampo – Biographies on the Semantic Web**[46] (online since 2018) [53] **Norssit Alumni** [54], **U.S. Congress Prosopographer** [55], and **AcademySampo**[47] (online since 2021) [56]. **NameSampo** [57] publishes data about over 2 million place names and places in Finland with old maps. The NameSampo project developed, based on the SPARQL Faceter tool [58] used in many earlier Sampos, the first version of the Sampo-UI framework [44] that has been used after this is in all Sampos, supporting implementation of principles P4–P6 from an UI point of view. Sampo-UI has also been re-used in the portal **Mapping Manuscript Migrations (MMM)**[48] (online since 2020) [39, 59] based on metadata about some 220 000 pre-modern manuscripts from the University of Oxford (U.K.), Schoenberg Institute (U.S.), and IRHT (France), in **FindSampo**[49] [60] (online since 2021) for supporting archaeology from a citizen science and metal detectorists' perspectives. **LetterSampo**[50] [61] is based on early modern epistolary metadata aggregated in the Early Modern Letters Online (EMLO) service[51] at the Oxford University, the CKCC corpus underlying ePistolarium[52] of the Huy-

[38]https://yasgui.triply.cc
[39]https://colab.research.google.com/notebooks/intro.ipynb
[40]https://jupyter.org
[41]https://www.go-fair.org/fair-principles/
[42]This application at https://museosuomi.fi got the Semantic Web Challenge Award at the ISWC 2004 conference.
[43]https://seco.cs.aalto.fi/applications/kulttuurisampo/
[44]https://kirjasampo.fi
[45]https://seco.cs.aalto.fi/projects/sotasampo/en/
[46]https://seco.cs.aalto.fi/projects/biografiasampo/en/
[47]https://seco.cs.aalto.fi/projects/akatemiasampo/en/
[48]https://seco.cs.aalto.fi/projects/mmm/
[49]https://seco.cs.aalto.fi/projects/sualt/
[50]https://seco.cs.aalto.fi/projects/rrl/
[51]http://emlo.bodleian.ox.ac.uk
[52]http://ckcc.huygens.knaw.nl/epistolarium/

gens Institute in the Netherlands, and correspSearch[53] service of the Berlin-Brandenburg Academy of Sciences. In addition, new Sampos are already in prototype phase: **LawSampo**[54] [62] publishes Finnish legislation and case law based on data from the Ministry of Justice in Finland. **ParliamentSampo**[55] [63] publishes LOD extracted from the materials of the Parliament of Finland (1907–2021), including knowledge graphs about over 960 000 Parliamentary debate speeches [64] and prosopographical data about the politicians' networks [65] in 1907–2021.

Developing these systems in use in a small university research group[56] would not have been possible without re-using the elements of the national infrastructure (Fig. 1) and developing them further step-by-step in a systematic way.

6. Discussion

This paper addressed challenges of extending the SW layer cake model for creating ontology and LOD infrastructures on national and domain specific levels. Lessons learned in developing Finnish ontology and LOD services 20012–2022 for practical applications were discussed. This work has utilized methods of design science [66–68] and action research [69], where the idea is to design artifacts, evaluate their value and utility, and to provide improvements in solutions. Rather than creating theoretical knowledge, design science applies knowledge. In this paper, infrastructure elements were designed, implemented, and applied to create the Sampo series of data services and portals as a proof-of-concept. They have had up to millions of end users, which suggests feasibility of the national infrastructures presented. The line of R&D presented is unique in its focus on different domains on a national level, longevity, and series of applications in use on the Semantic Web.

In contrast to current related ontology library systems [70] that typically focus on particular application domains, ONKI and Finto aimed at being a cross-domain ontology service on a national level. For example, the BioPortal [71] of Stanford University is focused on publishing biomedical ontologies. There are lots of LOD services and SPARQL endpoints around[57]. The novelty of the LDF.fi service lays on its 7-star model and the idea of integrating the data service with various online tools as well as leaning materials to support data reuse. Instead of being a focused data service for particular data, such as DBpedia for Wikipedias, the LDF.fi platform aims at being a cross-domain platform of datasets on a national level. The main application area of the presented infrastructure has been Cultural Heritage and Digital Humanities [40], especially in the 10's, although also systems for, e.g., e-health, e-government, and e-learning were developed.

During the past 20 years, the SW has evolved in phases [41] with a focus first on ontologies [11], then on Linked (Open) Data [6], and today on Knowledge Graphs (KG) [72]. The Sampo series reflects this development by showing a shift of focus form data publishing, based on shared ontologies and metadata vocabularies[58] (1. generation portals), to supporting the end-users of KGs with seamlessly integrated data-analytic tools and visualizations needed in areas such as Digital Humanities (2. generation systems). However, the series has also taken first steps forward towards 3. generation portals that can solve problems for the end users based on knowledge discovery, Artificial Intelligence, and computational creativity [49]. There are lots of related works pertaining to the different Sampo systems; discussing them is beyond the scope of this paper, but pointers to such works can be found in the referenced research papers.

The experiences reported in this paper indicate that creating and using a national semantic web infrastructure is useful from the data producers' and data users' points of view. However, creating and using linked data has its own challenges, too. More collaboration and agreements on data models and ontologies are needed for interoperability between the data producers, which complicates the publication process. Integration of SW technologies with legacy systems may be challenging, and there is lack of IT personnel competent in using SW technologies and tools.

[53]https://correspsearch.net
[54]https://seco.cs.aalto.fi/projects/lawlod/
[55]https://seco.cs.aalto.fi/projects/semparl/en/
[56]Semantic Computing Research Group (SeCo): https://seco.cs.aalto.fi
[57]https://www.w3.org/wiki/SparqlEndpoints
[58]https://lov.linkeddata.es/dataset/lov/

Creating linked data manually is costly but automatic methods may not be available and automation lowers data quality. Using structured semantic data and making the knowledge structures explicit to the end user in the UI calls for a new kind of digital data literacy and source criticism[59] from the end user [73, 74]. What the underlying data actually means is not always clear and issues of Big Data quality, such as completeness, veracity, skewness, uncertainty, fuzziness, and errors of data arise. However, in spite of the challenges, enriching data carefully with semantics, with one way or another, is in my mind the only way ahead towards creating a more and more intelligent Web in a cost-efficient way.

Acknowledgments Tens of people[60] at the Semantic Computing Research Group (SeCo) have contributed to the work presented in this paper, funded by over 50 organizations in Finland and beyond. Thanks to Marcia Zeng for the keynote talk invitation to the DCMI 2021 conference; this paper is a written account of that talk. The work of writing this paper is partly related to the Semantic Parliament (ParliamentSampo) project of the Academy of Finland, the EU project InTaVia: In/Tangible European Heritage[61], and the EU COST action Nexus Linguarum[62] on linguistic data science. CSC – IT Center for Science has provided computational resources for our projects.

References

[1] T. Berners-Lee, M. Fischetti and M.L. Dertouzos, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, 1st edn, Harper San Francisco, 1999. ISBN 0062515861.

[2] R. Goebel, S. Zilles, C. Ringlstetter, A.R. Dengel and G.A. Grimnes, What Is the Role of the Semantic Layer Cake for Guiding the Use of Knowledge Representation and Machine Learning in the Development of the Semantic Web?, in: *AAAI Spring Symposium: Symbiotic Relationships between Semantic Web and Knowledge Engineering*, 2008.

[3] J. Domingue, D. Fensel and J.A. Hendler, Introduction to the Semantic Web Technologies, in: *Handbook of Semantic Web Technologies*, J. Domingue, D. Fensel and J.A. Hendler, eds, Springer, 2011, pp. 1–41. ISBN 978-3-540-92913-0. doi:10.1007/978-3-540-92913-0_1.

[4] P. Hitzler, M. Krötzsch and S. Rudolph, *Foundations of Semantic Web technologies*, Springer, 2010.

[5] P. Brunet, L. de Luca, E. Hyvönen, A. Joffres, P. Plassmayer, M. Pronk, R. Scopigno and G. Sonkoly, Report on a European Collaborative Cloud for Cultura Heritage, 2022.

[6] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space (1st edition)*, Morgan & Claypool, Palo Alto, California, 2011. http://linkeddatabook.com/editions/1.0/.

[7] M. Zeng and J. Qin, *Metadata, Third Edition*, ALA Neal-Schuman, Chicago, 2022. ISBN 978-0-8389-4875-0.

[8] M. d'Aquin and N.F. Noy, Where to Publish and Find Ontologies? A Survey of Ontology Libraries, *Web Semantics: Science, Services and Agents on the World Wide Web* **11** (2012), 96–111.

[9] J.L. Martinez-Rodriguez, A. Hogan and I. Lopez-Arevalo, Information Extraction Meets the Semantic Web: A Survey, *Semantic Web – Interoperability, Usability, Applicability* **11**(2) (2020), 255–335. doi:10.3233/SW-180333.

[10] E. Hyvönen, *Semanttinen web. Linkitetyn avoimen datan käsikirja (Semantic Web. Handbook of Linked Open Data)*, Gaudeamus, 2018, p. 271. https://kauppa.gaudeamus.fi/sivu/tuotehaku/?action=search&search=semanttinen+web.

[11] S. Staab and R. Studer (eds), *Handbook on Ontologies (2nd Edition)*, Springer, 2009.

[12] D. Fensel, *Ontologies: Silver bullet for knowledge management and electronic commerce (2nd Edition)*, Springer, 2004.

[13] E. Hyvönen, K. Viljanen, J. Tuominen and K. Seppälä, Building a National Semantic Web Ontology and Ontology Service Infrastructure – The FinnONTO Approach, in: *Proceedings of the ESWC 2008, Tenerife, Spain*, Springer, 2008, pp. 95–109.

[14] J. Tuominen, M. Frosterus, K. Viljanen and E. Hyvönen, ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services, in: *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*, Springer, 2009.

[15] K. Viljanen, J. Tuominen and E. Hyvönen, Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI, in: *Proceedings of the ESWC 2009, Heraklion, Greece*, Springer, 2009, pp. 781–795.

[16] O. Suominen and E. Hyvönen, Improving the Quality of SKOS Vocabularies with Skosify, in: *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012)*, Springer, 2012.

[17] J. Kurki and E. Hyvönen, Collaborative Metadata Editor Integrated with Ontology Services and Faceted Portals, in: *Workshop on Ontology Repositories and Editors for the Semantic Web (ORES 2010) at ESWC 2010*, CEUR Workshop Proceedings, Vol. 596, 2010.

[18] E. Hyvönen, E. Mäkelä, T. Kauppinen, O. Alm, J. Kurki, T. Ruotsalo, K. Seppälä, J. Takala, K. Puputti, H. Kuittinen, K. Viljanen, J. Tuominen, T. Palonen, M. Frosterus, R. Sinkkilä, P. Paakkarinen, J. Laitio and K. Nyberg, CultureSampo – Finnish Culture on the Semantic Web 2.0. Thematic Perspectives for the End-user, in: *Museums and the Web 2009*, Archives & Museum Informatics, Toronto, 2009.

[59] https://ranke2.uni.lu/define-dsc/#%20,%20Universit%C3%A9%20du%20Luxembourg
[60] https://seco.cs.aalto.fi/people/
[61] https://intavia.eu/
[62] https://nexuslinguarum.eu/the-action

[19] E. Mäkelä, T. Ruotsalo and Hyvönen, How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo, *Semantic Web – Interoperability, Usability, Applicability* **3**(1) (2012), 85–109.
[20] O. Suominen, E. Hyvönen, K. Viljanen and E. Hukka, HealthFinland – a National Semantic Publishing Network and Portal for Health Information, *Journal of Web Semantics* **7**(4) (2009), 287–297.
[21] T. Sidoroff and E. Hyvönen, Semantic E-goverment Portals—A Case Study, in: *Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05*, 2005. http://www.seco.hut.fi/publications/2005/sidoroff-hyvonen-semantic-e-government-2005.pdf.
[22] T. Känsälä and E. Hyvönen, A semantic view-based portal utilizing Learning Object Metadata, in: *1st Asian Semantic Web Conference (ASWC2006), Semantic Web Applications and Tools Workshop*, 2004.
[23] M. Laukkanen, K. Viljanen, M. Apiola, P. Lindgren and E. Hyvönen, Towards Ontology-Based Yellow Page Services, in: *Proceedings of WWW2004 Workshop, Application Design, Development, and Implementation Issues, New York*, 2004.
[24] M. Frosterus, J. Tuominen, S. Pessala and E. Hyvönen, Linked Open Ontology cloud: managing a system of interlinked cross-domain light-weight ontologies, *International Journal of Metadata, Semantics and Ontologies* **10**(3) (2015), 189–201. http://dx.doi.org/10.1504/IJMSO.2015.073879.
[25] E. Hyvönen and E. Mäkelä, Semantic Autocompletion, in: *Proceedings of the First Asia Semantic Web Conference (ASWC 2006), Beijing*, Springer, 2006.
[26] J. Tuominen, T. Kauppinen, K. Viljanen and E. Hyvönen, Ontology-Based Query Expansion Widget for Information Retrieval, in: *Proceedings of the 5th Workshop on Scripting and Development for the Semantic Web (SFSW 2009), 6th European Semantic Web Conference (ESWC 2009)*, CEUR Workshop Proceedings, Vol. 449, 2009. http://ceur-ws.org/Vol-449/.
[27] J. Aitchison, A. Gilchrist and D. Bawden, *Thesaurus Construction and Use: A Practical Manual*, Aslib IMI, 2000.
[28] N. Guarino and C. Welty, Evaluating Ontological Decisions with OntoClean, *Communications of the ACM* **45**(2) (2002), 61–65.
[29] K. Seppälä and E. Hyvönen, Asiasanaston muuttaminen ontologiaksi. Yleinen suomalainen ontologia esimerkkinä FinnONTO-hankkeen mallista (Changing a Keyword Thesaurus into an Ontology. General Finnish Ontology as an Example of the FinnONTO Model), National Library, Plans, Reports, Guides, 2014. https://www.doria.fi/handle/10024/96825.
[30] E. Hyvönen, Digital Humanities on the Semantic Web: Sampo Model and Portal Series, *Semantic Web – Interoperability, Usability, Applicability* (2022), Accepted. http://semantic-web-journal.org/content/digital-humanities-semantic-web-sampo-model-and-portal-series.
[31] J.T. tuominen-et-al-avioand Nina Laurenne, M. Koho and E. Hyvönen, The Birds of the World Ontology AVIO, in: *The Semantic Web: ESWC 2013 Satellite Events*, Springer, 2013, pp. 300–301.
[32] J. Tuominen, N. Laurenne and E. Hyvönen, Biological Names and Taxonomies on the Semantic Web – Managing the Change in Scientific Conception, in: *Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011)*, Springer, 2011. doi:10.1007/978-3-642-21064-8_18.
[33] E. Hyvönen, Linked Open Data Infrastructure for Digital Humanities in Finland, in: *DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*, CEUR Workshop Proceedings, vol. 2612, 2020, pp. 254–259. http://ceur-ws.org/Vol-2612/short10.pdf.
[34] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, Quality Assessment for Linked Data: A Survey, *Semantic Web – Interoperability, Usability, Applicability* **7**(1) (2016), 63–93. doi:10.3233/SW-150175.
[35] E. Hyvönen, Preventing interoperability problems instead of solving them, *Semantic Web – Interoperability, Usability, Applicability* **1**(1–2) (2010), 33–37.
[36] O. Suominen, S. Pessala, J. Tuominen, M. Lappalainen, S. Nykyri, H. Ylikotila, M. Frosterus and E. Hyvönen, Deploying National Ontology Services: From ONKI to Finto, in: *Proceedings of the Industry Track at the International Semantic Web Conference 2014*, CEUR Workshop Proceedings, 2014, Vol 1383. ISSN 1613-0073. http://www.ceur-ws.org/Vol-1383.
[37] E. Hyvönen, J. Tuominen, M. Alonen and E. Mäkelä, Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets, in: *ESWC 2014: The Semantic Web: ESWC 2014 Satellite Events*, Springer, 2014, pp. 226–230. doi:10.1007/978-3-319-11955-7_24.
[38] J.E. Labra Gayo, E. Prud'hommeaux, I. Boneva and D. Kontokostas, *Validating RDF Data*, Synthesis Lectures on the Semantic Web: Theory and Technology, Vol. 7, Morgan & Claypool Publishers LLC, 2017, pp. 1–328. doi:10.2200/s00786ed1v01y201707wbe016.
[39] E. Hyvönen, E. Ikkala, M. Koho, J. Tuominen, T. Burrows, L. Ransom and H. Wijsman, Mapping Manuscript Migrations on the Semantic Web: A Semantic Portal and Linked Open Data Service for Premodern Manuscript Research, in: *Semantic Web. Proceedings of the The 20th International Semantic Web Conference (ISWC 2021)*, Springer, 2021, Forth-coming.
[40] E. Gardiner and R.G. Musto, *The Digital Humanities: A Primer for Students and Scholars*, Cambridge University Press, New York, NY, USA, 2015, https://doi.org/10.1017/CBO9781139003865.
[41] P. Hitzler, A Review of the Semantic Web Field, *Commun. ACM* **64**(2) (2021), 76–83–. doi:10.1145/3397512.
[42] L. Rietveld and R. Hoekstra, The YASGUI family of SPARQL clients, *Semantic Web – Interoperability, Usability, Applicability* **8**(3) (2017), 373–383. doi:10.3233/SW-150197.
[43] M. Tamper, A. Oksanen, J. Tuominen, A. Hietanen and E. Hyvönen, Automatic Annotation Service: Utilizing a Named Entity Linking Tool in Legal Domain, 2019, Submitted article under evaluation.
[44] E. Ikkala, E. Hyvönen, H. Rantala and M. Koho, Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces, *Semantic Web – Interoperability, Usability, Applicability* **13**(1) (2022), 69–84. doi:10.3233/SW-210428.
[45] E. Hyvönen, S. Saarela and K. Viljanen, Application of ontology-based techniques to view-based semantic search and browsing, in: *Proceedings of the First European Semantic Web Symposium*, Springer, 2004.
[46] D. Tunkelang, *Faceted search*, Morgan & Claypool Publishers, CA, USA, 2009.

[47] Y. Tzitzikas, N. Manolis and P. Papadakos, Faceted exploration of RDF/S datasets: a survey, *Journal of Intelligent Information Systems* **48**(2) (2017), 329–364. doi:10.1007/s10844-016-0413-8.

[48] G. Marchionini, Exploratory search: from finding to understanding, *Communications of the ACM* **49**(4) (2006), 41–46.

[49] E. Hyvönen, Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery, *Semantic Web – Interoperability, Usability, Applicability* **11**(1) (2020), 187–193.

[50] E. Hyvönen, P. Leskinen and J. Tuominen, LetterSampo – Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data of the Republic of Letters (2021), Submitted for review. https://seco.cs.aalto.fi/publications/2020/hyvonen-et-al-lettersampo-2020.pdf.

[51] E. Mäkelä, K. Hypén and E. Hyvönen, BookSampo—Lessons Learned in Creating a Semantic Portal for Fiction Literature, in: *Proc. of ISWC-2011, Bonn, Germany*, Springer, 2011.

[52] E. Hyvönen, E. Heino, P. Leskinen, E. Ikkala, M. Koho, M. Tamper, J. Tuominen and E. Mäkelä, WarSampo Data Service and Semantic Portal for Publishing Linked Open Data about the Second World War History, in: *The Semantic Web – Latest Advances and New Domains (ESWC 2016)*, Springer, 2016, pp. 758–773.

[53] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen and K. Keravuori, BiographySampo – Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research, in: *Proceedings of the 16th Extended Semantic Web Conference (ESWC 2019)*, Springer, 2019, pp. 574–589.

[54] E. Hyvönen, P. Leskinen, E. Heino, J. Tuominen and L. Sirola, Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web, in: *Proceedings, Language, Technology and Knowledge (LDK 2017)*, Springer, 2017, pp. 113–119. https://link.springer.com/chapter/10.1007/978-3-319-59888-8_9.

[55] G. Miyakita, P. Leskinen and E. Hyvönen, Using Linked Data for Prosopographical Research of Historical Persons: Case U.S. Congress Legislators, in: *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018, Nicosia, Cyprus*, Springer, 2018.

[56] P. Leskinen and E. Hyvönen, Linked Open Data Service about Historical Finnish Academic People in 1640–1899, in: *Proc. of the Digital Humanities in the Nordic Countries (DHN 2020)*, CEUR WS Proceedings, 2020, Forth-coming.

[57] E. Ikkala, J. Tuominen, J. Raunamaa, T. Aalto, T. Ainiala, H. Uusitalo and E. Hyvönen, NameSampo: A Linked Open Data Infrastructure and Workbench for Toponomastic Research, in: *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Geospatial Humanities*, GeoHumanities'18, ACM, New York, NY, USA, 2018, pp. 2:1–2:9. ISBN 978-1-4503-6032-6. doi:10.1145/3282933.3282936.

[58] M. Koho, E. Heino and E. Hyvönen, SPARQL Faceter – Client-side Faceted Search Based on SPARQL, in: *Joint Proceedings of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop*, CEUR Workshop Proceedings, 2016, Vol. 1615. http://www.ceur-ws.org/Vol-1615.

[59] M. Koho, T. Burrows, E. Hyvönen, E. Ikkala, K. Page, L. Ransom, J. Tuominen, D. Emery, M. Fraas, B. Heller, D. Lewis, A. Morrison, G. Porte, E. Thomson, A. Velios and H. Wijsman, Harmonizing and Publishing Heterogeneous Pre-Modern Manuscript Metadata as Linked Open Data, 2021, Accepted, JASIST Special Issue.

[60] E. Hyvönen, H. Rantala, E. Ikkala, M. Koho, J. Tuominen, B. Anafi, S. Thomas, A. Wessman, E. Oksanen, V. Rohiola, J. Kuitunen and M. Ryyppö, Citizen Science Archaeological Finds on the Semantic Web: The FindSampo Framework, *Antiquity, A Review of World Archaeology* (2021), Accepted.

[61] J. Tuominen, E. Mäkelä, E. Hyvönen, A. Bosse, M. Lewis and H. Hotson, Reassembling the Republic of Letters – A Linked Data Approach, in: *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, CEUR Workshop Proceedings, vol. 2084, 2018, pp. 76–88. http://www.ceur-ws.org/Vol-2084/paper6.pdf.

[62] E. Hyvönen, M. Tamper, E. Ikkala, S. Sarsa, A. Oksanen, J. Tuominen and A. Hietanen, Publishing and Using Legislation and Case Law as Linked Open Data on the Semantic Web, in: *The Semantic Web: ESWC 2020 Satellite Events*, Lecture Notes in Computer Science, Vol. 12124, Springer, 2020, pp. 110–114. doi:10.1007/978-3-030-62327-2_19.

[63] E. Hyvönen, L. Sinikallio, P. Leskinen, M.L. Mela, J. Tuominen, K. Elo, S. Drobac, M. Koho, E. Ikkala, M. Tamper, R. Leal and J. Kesäniemi, Finnish Parliament on the Semantic Web: Using ParliamentSampo Data Service and Semantic Portal for Studying Political Culture and Language, in: *Digital Parliamentary data in Action (DIPADA 2022), Workshop at the 6th Digital Humanities in Nordic and Baltic Countries Conference, long paper*, CEUR WS Proceedinsg, 2022, Forth-coming. https://seco.cs.aalto.fi/publications/2022/hyvonen-et-al-semparl-dhnb-2022.pdf.

[64] L. Sinikallio, S. Drobac, M. Tamper, R. Leal, M. Koho, J. Tuominen, M.L. Mela and E. Hyvönen, Plenary Debates of the Parliament of Finland as Linked Open Data and in Parla-CLARIN Markup, 2021, Aalto Univerisity, SeCo Group, submitted.

[65] P. Leskinen, E. Hyvönen and J. Tuominen, Members of Parliament in Finland Knowledge Graph and its Linked Open Data Service, 2021, Aalto Univerisity, SeCo Group, submitted.

[66] S.T. March and G.F. Smith, Design and natural science research on information technology, *Decision Support Systems* **15**(4) (1995), 251–266. doi:10.1016/0167-9236(94)00041-2.

[67] A.R. Hevner, S.T. March, J. Park and S. Ram, Design science in information systems research, *MIS Quarterly: Management Information Systems* **28**(1) (2004), 75–105. doi:10.2307/25148625.

[68] K. Peffers, T. Tuunanen, M.A. Rothenberger and S. Chatterjee, A design science research methodology for information systems research, *Journal of Management Information Systems* **24**(3) (2007), 45–77. doi:10.2753/MIS0742-1222240302.

[69] R.M. Davison, M.G. Martinsons and N. Kock, Principles of canonical action research, *Information Systems Journal* **14**(1) (2004), 65–86. doi:10.1111/j.1365-2575.2004.00162.x.

[70] M. d'Aquin and N.F. Noy, Where to Publish and Find Ontologies? A Survey of Ontology Libraries, *Journal of Web Semantics First Look* **11**(0) (2012). doi:10.2139/ssrn.3198941.

[71] M. Salvadores, P.R. Alexander, M.A. Musen and N.F. Noy, BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF, *Semantic Web – Interoperability, Usability, Applicability* **4**(3) (2013), 277–284. doi:10.3233/SW-2012-0086.
[72] C. Gutierrez and J.F. Sequeda, Knowledge graphs, *Communications of the ACM* **64**(3) (2021), 96–104. doi:10.1145/3418294.
[73] T. Koltay, Data literacy for researchers and data librarians, *Journal of Librarianship and Information Science* **49**(1) (2015), 3–14. doi:10.1177/0961000615616450.
[74] E. Mäkelä, K. Lagus, L. Lahti, T. Säily, M. Tolonen, M. Hämäläinen, S. Kaislaniemi and T. Nevalainen, Wrangling with non-standard data, in: *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*, S. Reinsone, I. Skadiņa, A. Baklāne and J. Daugavietis, eds, CEUR Workshop Proceedings, CEUR-WS.org, Germany, 2020, pp. 81–96, Digital Humanities in the Nordic Countries, DHN2020 ; Conference date: 17-03-2020 Through 20-03-2020.