

On which legal regulations is a public service based? Fostering transparency in public administration by using knowledge graphs

Leila Feddou¹, Maximilian Raupach¹, Felicitas Löffler^{2 †}, Samira Babalou¹, Jonas Hoyer¹,
Marianne Mauch¹, Birgitta König-Ries^{1,3,4}

Abstract: Knowledge about digitization in the public administration is complex and scattered. Information about legal regulations, methods, processes, APIs, metadata and data standards, registers, and terminologies are spread across different platforms. Hence, it is difficult for employees, developers, and decision makers to figure out what data standards, formats, and APIs are relevant for the digitization of a specific public service. Training administrative employees and IT companies requires to gather and link the required knowledge first in a well-structured and accessible manner. We address this need for more shared and transparent knowledge in the digitization of public services in Germany. We propose a first version of an ontology (GerPS-onto) for public administration instantiated by one example german public service. We utilize semantic modelling and Linked Data technologies to enable appropriate data and process descriptions that are readable for humans and machines. We also demonstrate how an existing process description of a public service can be linked to existing terminologies and evaluate the resulting ontology using domain-specific competency questions that are translated to SPARQL queries.

Keywords: Knowledge Graph; Ontology; Public Service; IT competence; Public Administration

1 Introduction and Related Work

A solid knowledge base is essential for a variety of tasks in public administration. The current digital transformation requires dedicated training and innovative concepts to enhance IT competences in public administration and its stakeholders. Besides trainings in using new software applications, public administration must be empowered in new arising tasks being relevant for artificial intelligence, e.g., annotation tasks to support the training of machine learning models or logic tasks to define and classify domain specific terms into hierarchical structures. Setting up such a knowledge base is challenging in the e-Government domain in Germany as data is scattered and spread across various platforms. For instance, digitizing public services requires to first understand and analyze the processes and data, including required activities, actors, and legal foundations. However, in Germany, there are different

[†] Corresponding author: felicitas.loeffler@tfm.thueringen.de

¹ Heinz Nixdorf Chair for Distributed Information Systems, Friedrich-Schiller-University, Jena, Germany

² Thuringian Ministry of Finance, Ludwig-Erhard-Ring 7, 99099 Erfurt, Germany

³ Michael-Stifel-Center for Data-Driven and Simulation Science, Jena, Germany

⁴ German Center for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany

approaches for processual descriptions. One concept is based on the legal foundations, the Federal Information Management - FIM⁵. It is the result of the analysis of legal resources (norm analysis) and consists of three basic building blocks: public service descriptions, data fields, and processes. The need to provide a broad spectrum of actors access to an integrated, harmonized view on complex information and the need to provide easy-to-use ways for information modeling, integration and harmonization is not unique to public administration. In many other fields, in the last decade, Knowledge Graphs (KGs) [Ho21] have been proposed as a promising approach to meet both needs. They encode the semantics of the data by representing it as real world concepts and their relations and have, among others, the following benefits: (1) Independent development by different actors with different expertise and world interpretation, (2) enabling data integration by interlinking concepts and relations, and (3) allowing machine readability and interpretation. In this paper, we use the term KG to describe information on the schema and instance level, while ontology is used to describe schema level information (e.g., the term “process” is a concept in our ontology, the concrete process of a specific service is part of the KG). Even if KGs are primarily instruments for developers, KG applications offer a wide range of opportunities for further stakeholders. For instance, semantic search and question answering approaches integrate KGs to allow searches beyond keywords (e.g., on what legal basis is a public service based?), which are relevant for decision makers, administrative staff, and interested citizens.

Related work. The importance of the semantic description of e-Government knowledge has been intensively discussed in the last decade [CM09, HTW10]. In the context of public administration, a very early ontology [Of12] was developed with the aim of semantically annotating law texts to ease transforming them into process models. The ontology combines concepts from different perspectives: law application (e.g., notification), administrative actions (e.g., application), and the law text (e.g., legal consequence). [SEW13] propose an ontology in UML format for describing public administration processes based on a specific use case (vehicle registration). It relates relevant concepts such as Public service, Activities, or Law. However, both approaches do not follow any agreed upon standard (e.g., FIM), [SEW13] is not available in a semantic common format, and [Of12] does not capture the process-oriented nature of public services (e.g., process steps). There are already initiatives at the European Union (EU) level to provide common vocabularies. The Publications Office of the EU provides a catalog of vocabularies (e.g., EuroVoc⁶) and models (e.g., the European Legislation Identifier (ELI) ontology⁷). However, semantic descriptions of the public administration domain are, to the best of our knowledge, still missing. This aspect is included in the the e-Government Core Vocabularies⁸ that was released by the the ISA² Programme⁹ and that includes a collection of upper level ontologies for basic

⁵ FIM, <https://fimportal.de>

⁶ EuroVoc, <https://op.europa.eu/en/web/eu-vocabularies/thesauri>

⁷ ELI, <https://op.europa.eu/en/web/eu-vocabularies/eli>

⁸ e-Government Core Vocabularies, <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies>

⁹ ISA, https://ec.europa.eu/isa2/isa2_en/

concepts in public administration (e.g., public service or legal entity). Nevertheless, the proposed vocabularies offer only very high level concepts. There exist also country-specific initiative [HTP05, SB09, BC20, Be22] e.g., the Finnish Ontology for Public Administration Services (JUPO)¹⁰. The latter are mostly too specific and follow the properties and standards of the corresponding country. Overall, existing efforts are in most of the cases not process-oriented, do not follow a known standard, are either too granular or too general, and do not link to existing terminologies.

We aim to close this semantic gap and model the missing context knowledge for the digitization of public services in Germany. Following the FAIR data [Wi16] and 5 stars principles¹¹, we reuse existing terminologies and align our model with upper level ontologies. We also demonstrate the usage of these terminologies for an existing process to populate a first KG. In summary, our contribution is as follows: (1) A first version of an ontology for describing processes of German public services. (2) Alignment of our concepts to the e-Government Core Vocabularies, and a first identification of semantic gaps (e.g., missing concepts). (3) A KG for a specific public service: KG population by automatically transforming an XML-based format for processes into a semantic format. (4) Evaluation of the ontology using SPARQL queries corresponding to specific competency questions. In the following section, we introduce our proposed ontology, preliminary evaluation results, and a first demonstration of the populated KG of one process description of a digitized service. All results are publicly available in our GitHub repository¹².

2 GerPS-onto: An ontology for German public service processes

The primary goal of our proposed GerPS-onto¹³ is the semantic modeling of processes in the German public administration. For its development, we combined two methods: *Ontology Development 101* [No01] and *Conducting Literature Search for Artifact Reuse (CLeAR)* [Ca20]. At first, we defined the scope of the ontology by formulating possible competency questions. Next, we assessed existing ontologies for reuse. For this purpose, we used the following aspects from *CLeAR*, because the *Ontology Development 101* methodology contains only very generic specifications regarding reuse. We looked for suitable ontologies with specific keywords such as “BPMN”¹⁴, “Knowledge Graph”, “Public Administration”, “Public Service”, “Process”, “Ontology” in *Google Scholar*. This resulted in ten ontologies to be further analyzed. Afterwards, we applied a set of inclusion (e.g., based on BPMN) and exclusion (e.g., online unavailability) criteria to select candidates that are suitable for reuse. Based on these criteria the single result is the BPMN 2.0 based Ontology for Business Process Representation (BBO) [AAGK19]. To build our

¹⁰ JUPO, <https://finto.fi/jupo/en/>

¹¹ 5 stars open data, <https://5stardata.info>

¹² GerPS-onto GitHub repository, <https://github.com/fusion-jena/GerPS-onto>

¹³ GerPS-onto URL, <https://w3id.org/GerPS-onto/ontology/>

¹⁴ BPMN, <https://www.omg.org/spec/BPMN/>

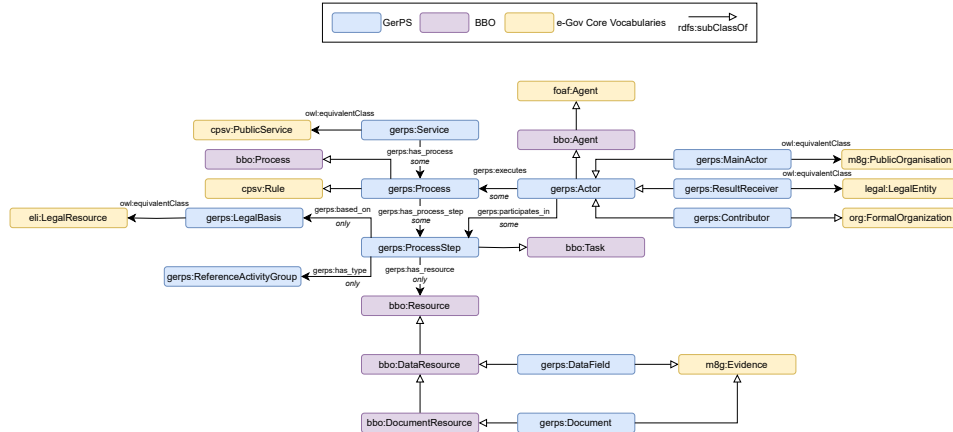


Fig. 1: Conceptual model of GerPS-onto

conceptual model, we identified important terms from our competency questions, precisely defined them, decided if they should be modeled as *property values*, *properties*, or *classes*, and mapped some classes and properties to BBO. For the latter step, we compared our competency questions to the ones provided by the selected ontology and analyzed the definitions of the BBO concepts and their conformity with our scope. The criteria, term definitions, and comparison of competency questions are available in GitHub. Figure 1 presents GerPS-onto. We reused the following BBO main classes: *Process*, *Task*, *Data Resource*, *Document Resource*, and *Agent* by defining them as superclasses of our added domain specific concepts (e.g., `gerps:Actor` `rdfs:subClassOf` `bbo:Agent`). We also reused two BBO properties (`gerps:has_resource` `rdfs:subPropertyOf` `bbo:has_resource` and `gerps:has_process` `rdfs:subPropertyOf` `bbo:has_process`). We added new properties to link the different domain specific concepts (e.g., `gerps:has_process_step`). The ontology was developed using *Protégé*¹⁵.

Ontology Alignment to e-Government Core Vocabularies. Apart from the BBO ontology, we also aligned our concepts to the e-Government Core Vocabularies of the EU. Therefore, we manually inspected a collection of ten available vocabularies to find classes corresponding to the GerPS-onto concepts. We then map them using two different properties (`owl:equivalentClass` and `rdfs:subClassOf`) to express either equivalence or generalization/specialization. In total, we mapped nine GerPS-onto classes (e.g., `gerps:Service` `owl:equivalentClass` `cpsv:PublicService`). However, for some other classes, we did not find any suitable candidates. This includes the core classes involved in modeling the actual business process reused from the BBO ontology (e.g., `gerps:ProcessStep`). For one of our new added GerPS-onto classes, (`gerps:ReferenceActivityGroup`), no candidate was found. It defines a concrete type of a specific process step, e.g., provide information or formally

¹⁵ Protégé, <https://protege.stanford.edu/>

examine the facts, and is a specific term of the FIM method. Overall, most of the existing vocabularies provide only high level concepts and do not model the underlying process.

Ontology Population and Evaluation. We populated the ontology with instances from one German public service¹⁶ to create a first KG. It is populated using sources in XML-based formats like *XProzess*¹⁷ and *XDatenfelder*¹⁸, for describing administrative processes and form fields involved in executing the specific service. The population is automatically done by parsing the XML-based files after manually mapping specific XML-nodes to our ontology classes. It is an example for the transformation of available domain knowledge (in non-semantic formats) into semantic formats. We evaluated the ontology by verifying its capability to correctly answer the competency questions expressed as SPARQL queries for one public service. The queries were correctly answered and are publicly available together with their corresponding results on GitHub. One example query with some results is shown in Figure 2 together with the concrete relations to other entities for one specific result.

The figure illustrates a SPARQL query interface and its results. On the left, a query editor shows a SPARQL query for finding legal foundations for a specific public service. The query is executed, and the results are displayed in a table under the heading 'Handlungsgrundlage'. One result is highlighted with a box and an arrow pointing to a detailed view on the right.

The detailed view on the right shows a graph of relationships between entities. The entities are 'Anzeige zur Beschäftigung ein...' (Process), 'Anzeige fachlich prüfen' (ProcessStep, SubProcess), and '§ 29 MuSchG' (LegalBasis). The relationships are 'has_process_step' and 'based_on'.

The metadata table on the right provides details for the process instance:

name	Anzeige zur Beschäftigung einer schwangeren oder stillenden Frau bearbeiten
has_leikaID	99006028261000
has_process_step	Zusätzliche Informationen einholen, Erhebung der Geldbuße veranlassen, Zuständigkeit prüfen, Notwendigkeit von Überwachungsmaßnahmen prüfen, Anzeige empfangen, Korrektur anfordern, Anzeige fachlich prüfen, Überwachungsmaßnahmen einleiten, Anzeige zurückweisen, Anzeige formell prüfen, Über Erhebung einer Geldbuße entscheiden, Auf zuständige Behörde hinweisen

Fig. 2: SPARQL query to obtain the legal foundations for a given public service and some corresponding results

3 Conclusion

We introduced GerPS-onto, an ontology for modeling processes of German public services. For this purpose, we reused concepts of an existing ontology modeling business processes and linked some of our concepts to the upper level e-Government Core Vocabularies. We populated the ontology with one German public service. In future work, we aim to model

¹⁶ FIM portal, <https://fimportal.de/detail/L/99006028261000>

¹⁷ XProzess, <https://www.xrepository.de/details/urn:xoev-de:mv:em:standard:xprozess>

¹⁸ XDatenfelder, <https://www.xrepository.de/details/urn:xoev-de:fim:standard:xdatenfelder>

missing knowledge and to extend the ontology with more specific concepts (e.g., technical information about standards, registers, and architectures). We hope that GerPS-onto can serve as a starting point for a comprehensive KG of digitization in public administration. We envision others to contribute to this KG by modeling other public services and linking their model to GerPS-onto.

Bibliography

- [AAGK19] Annane, A.; Aussenac-Gilles, N.; Kamel, M.: BBO: BPMN 2.0 based ontology for business process representation. In: ECKM 2019. volume 1, pp. 49–59, 2019.
- [BC20] Bastola, R.; Campus, P.: Developing domain ontology for issuing certificate of citizenship of Nepal. *Journal of Information Technology*, 2(02):73–90, 2020.
- [Be22] Benaddi, H.; Laaz, N.; Kettani, E. E.; Hannad, Y.: Ontology Model for Public Services in Morocco Based on 5W1H Approach: PSOM-eGovMa. *Procedia Computer Science*, 198:429–434, 2022.
- [Ca20] Campos, P. M.C. et al.: Finding reusable structured resources for the integration of environmental research data. *Environmental Modelling & Software*, 133:104813, 2020.
- [CM09] Charalabidis, Y.; Metaxiotis, K.: Ontology-based management of e-Government knowledge. In: *Social and Political Implications of Data Mining: Knowledge Management in E-Government*, pp. 221–234. IGI Global, 2009.
- [Ho21] Hogan, A. et al.: *Knowledge Graphs. Synthesis Lectures on Data, Semantics, and Knowledge 22*. Springer, 2021.
- [HTP05] Hinkelmann, K.; Thönssen, B.; Probst, F.: Referenzmodellierung für E-Government-Services. *Wirtschaftsinf.*, 47(5):356–366, 2005.
- [HTW10] Hinkelmann, K.; Thönssen, B.; Wolff, D.: Ontologies for E-government. In: *Theory and Applications of Ontology: Computer Applications*. Springer Netherlands, Dordrecht, pp. 429–462, 2010.
- [No01] Noy, N.; McGuinness, D. L. et al.: *Ontology development 101*. Knowledge Systems Laboratory, Stanford University, 2001, 2001.
- [Of12] Off, T.: *Durchgängige Verfolgbarkeit im Vorfeld der Softwareentwicklung von E-Government-Anwendungen: Ein ontologiebasierter und modellgetriebener Ansatz am Beispiel von Bürgerdiensten*. Doctoral thesis, Universität Potsdam, 2012.
- [SB09] Savvas, I.; Bassiliades, N.: A process-oriented ontology-based knowledge management system for facilitating operational procedures in public administration. *Expert Systems with Applications*, 36(3, Part 1):4467–4478, 2009.
- [SEW13] Schminck, A.; Eid-Sabbagh, R.-H.; Weske, M.: eGovernment Process Knowledge Ontology - Business Process Knowledge Interdependencies in the Public Administration. In: *INFORMATIK 2013*. volume P-220 of LNI. GI, pp. 722–735, 2013.
- [Wi16] Wilkinson, M. D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. 3(1):160018, 2016.