

Automated Consent Form Generation: Using Ontologies for Model Configuration

Maryam Mohammadi^{1,*}, Aliena Strathmann¹, Katja Politt^{2,1}, Annett Jorschick¹ and Hendrik Buschmeier¹

¹CRC 1646 'Linguistic Creativity in Communication', Bielefeld University, Bielefeld, Germany

²Rostock University, Rostock, Germany

Abstract

Legal and ethical requirements governing the processing of (personal) data have become increasingly complex, introducing substantial overhead for scientific research involving human participants. To address this challenge, we propose a modular, ontology-based framework and an initial implementation of an automated consent form Wizard that supports the generation of ethically sound and GDPR-compliant consent documents. The framework is grounded in Semantic Web technologies, employing data models that reuse established resources (e.g., the Data Privacy Vocabulary [1]) while developing lightweight, domain-specific ontologies where necessary. As a real-world use case, the system shows that exposing ontologies and Linked Data resources significantly enhances interoperability and findability, thereby improving data reusability in line with FAIR and Open Science principles.

Keywords

consent forms, taxonomy, ontology, linked data, modules

1. Introduction

Research across many domains, including linguistics, psychology, the social sciences, and other, inherently relies on data collected from human participants. Such data frequently qualify as personal or sensitive, imposing strict legal and ethical constraints on their collection, processing, and reuse. These constraints directly affect the adoption of Open Science practices and adherence to the 'FAIR' principles [2], particularly with respect to data sharing and reuse. The complexity of legal aspects poses additional challenges for researchers without specialized legal expertise, limiting their ability to design consent procedures that enable meaningful collaboration. Importantly, these challenges extend beyond documentation, influencing study design, participant information, and downstream data use. Legal and ethical requirements must therefore be integrated during planning but are often addressed ad hoc and remain disconnected from subsequent data management processes.

Recent work has therefore highlighted the need for research data management infrastructures that integrate automated support across the entire data lifecycle, from planning and collection to storage, sharing, and reuse, while explicitly accounting for legal and ethical constraints (e.g., [3]). Building on this vision, we introduce RUDI, "*Research-centered User-oriented Data Infrastructure*", developed within the INF project of a Collaborative Research Center (CRC 1646), *Linguistic Creativity in Communication* at Bielefeld University. RUDI aims to bridge the gap between legal requirements, user needs, and technical data management, leveraging Semantic Web technologies and ontology-based modelling.

The RUDI framework comprises two complementary, yet independently developed platforms: (i) a consent form "Wizard" application that supports the guided, automated step-by-step generation of consent forms, and (ii) a data management platform that governs data handling operations, such as search, access, and sharing, in accordance with the recorded consent. Both platforms are designed

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

*Corresponding author.

✉ maryam.mohammadi@uni-bielefeld.de (M. Mohammadi); aliena.strathmann@uni-bielefeld.de (A. Strathmann); katja.politt@uni-bielefeld.de (K. Politt); annett.jorschick@uni-bielefeld.de (A. Jorschick); hbuschme@uni-bielefeld.de (H. Buschmeier)

ORCID 0009-0007-8747-029X (M. Mohammadi); 0000-0002-4912-2653 (K. Politt); 0009-0004-0776-7113 (A. Jorschick); 0000-0002-9613-5713 (H. Buschmeier)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

to interoperate through a shared semantic layer grounded in ontology-based representations. In this paper, we focus on the first component, the *Consent Form Wizard*, presenting its ontology-engineering foundation. The ontologies establish the semantic groundwork for future integration not only with the data management platform, but also with external third-party systems and infrastructures.

The Wizard builds on established Semantic Web standards, reusing existing resources such as the Data Privacy Vocabulary (DPV; [1, 4]), and domain-specific lightweight ontologies where necessary. In particular, RUDI was initially developed for the field of (experimental) linguistics, which requires fine-grained modelling of specialized concepts with respect to research domain and demographic data. At the same time, the framework is designed with extensibility in mind: it supports the integration of other domain ontologies, thus enabling adoption by other disciplines through the incorporation of their respective vocabularies. In this way, the Wizard remains domain-agnostic at the architectural level.

Our application advances existing approaches by replacing static, manually curated templates with dynamically generated configurations that adapt expert-defined *legal* foundations in response to the needs of *users* including both researchers and study participants (for legal details, see [5]). The Wizard and its underlying ontologies provide a systematic, extensible, and machine-interpretable framework that unifies legal and technical requirements, supporting consistent, interoperable, and reusable data.

2. Dynamic Framework

Within our framework, the Consent Form Wizard generates study-specific consent forms and captures participant consent in a structured, machine-readable format. The framework relies on configurable templates and a modular architecture to accommodate evolving legal and domain-specific requirements.

2.1. Modular Interfaces

Linguistics serves as the initial application domain of the framework, requiring a modular design that accommodates heterogeneous data types and diverse participant populations. Linguistic research involves multimodal data (e.g., text, audio, video, and metadata) collected in varied research settings, making static legal templates insufficient. Studies may also involve vulnerable groups, such as children or individuals with impairments, requiring additional legal safeguards [6].

To address these challenges, the framework combines ontology-based representations, controlled vocabularies, and configurable templates to formally capture study properties and legal requirements in an interoperable and extensible manner. The model is structured into three interacting modules—*user*, *legal*, and *technical*, with clearly defined interfaces governing their interaction.

The user–legal interface supports two stakeholder groups: (i) *researchers*, who design studies and manage data, and (ii) *participants*, who provide data (“data subjects” under GDPR). It translates legal and ethical requirements into accessible, domain-specific guidance by providing researchers with structured information on obligations such as data minimization, while participants receive clear and understandable consent materials. Legal experts formalize linguistic use cases into reusable semantic templates and controlled vocabularies [7, 5], which provide the foundation for the technical layer.

The legal–technical interface operationalizes legal requirements through ontology-based configurations and rule-driven form generation. Structured templates encode study properties, dependencies, and constraints, enabling dynamic forms, automated validation, and context-specific guidance. The system also enforces legal requirements, such as guardian consent for minors, and generates privacy documentation based on the recorded study configuration. Future work will integrate automated identification of personal and sensitive data using the iVA decision process [8].

The technical–user interface emphasizes usability, transparency, and accessibility through a user-centred design guided by established usability principles [9, 10]. Researchers are supported by condition-based workflows, contextual help, and real-time generation of outputs, reducing perceived legal complexity. Participants receive clear, modular consent documents and can provide granular consent, with their choices represented as structured data for downstream data management.

2.2. Ontology Foundation

Semantic Web technologies enable the publication, interlinking, and reuse of structured, machine-readable data through ecosystems such as Linked Open Data (LOD) [11, 12]. This paradigm shifts platforms from isolated systems toward interoperable data infrastructures, where data can be semantically enriched, linked to external resources, and processed across applications. In line with Open Science principles, the Consent Form Wizard adopts Semantic Web standards and existing ontologies to support interoperability and reuse while ensuring that data sharing remains aligned with legal and ethical constraints, particularly for personal data.

Modelling Personal Data The increasing importance of personal data protection has motivated the development of machine-readable representations of legal concepts. The Data Privacy Vocabulary (DPV) [4, 13], developed by the W3C Data Privacy Vocabularies and Controls Community Group, provides a modular framework for representing personal data processing, legal bases, purposes, and related concepts through interoperable vocabularies. In our framework, DPV and its Personal Data (PD) extension [1, 14] form the basis for modelling personal and sensitive data, enabling the representation of data categories, legal constraints, and consent properties within the ontology layer. Where necessary, we adapt and extend these resources, for instance, by incorporating language-specific labels (i.e., German) required for the user interface, while maintaining compatibility with the underlying standards.

Modelling Linguistics Data Linguistic research requires domain-specific extensions for specialized study designs, data types, and metadata not covered by existing resources. To address this gap, we develop *eXperimental Linguistics* (XLing)¹, a lightweight and extensible ontology tailored to experimental linguistics. Its concepts are derived from empirical domain analyses [7] and refined with input from legal experts. Rather than hard-coding domain knowledge, XLing uses Semantic Web standards to enable interoperability and integration with the Linked Data ecosystem. The ontology is primarily implemented using SKOS [15], complemented by RDFS for class hierarchies and Dublin Core Terms for metadata representation. XLing concepts are aligned with the following standards and resources:

I. General standards: We employ ISO standards such as ISO 639-3 for language codes [16] and ISO 3166-1 for countries [17], ensuring consistency with widely used identifiers.

II. Domain-specific resources: For specialized categories, we relied on domain-specific ontologies to ensure semantic consistency and interoperability. In particular, health-related items are linked to external ontologies and controlled vocabularies from the [WHO International Classification of Diseases \(ICD\)](#), supporting alignment with standardized medical terminology. When necessary, additional relevant ontologies were identified via [BioPortal](#). Furthermore, items related to personal data were mapped to corresponding concepts in the Personal Data module of the Data Privacy Vocabulary (DPV; [1, 14]), ensuring compliance with established privacy semantics.

III. Linguistic infrastructures: To facilitate integration within the language resources community, XLing entries are further aligned with [CLARIN vocabularies](#), supporting interoperability with existing research infrastructures.

Through these links, the platform situates its internal representations within the global Linked Data ecosystem, enabling semantic enrichment, cross-domain integration, and reuse across systems.

3. Consent Form Wizard

The Consent Form Wizard demonstrates a concrete, practical application of Semantic Web technologies to reduce workload and complexity of administrative tasks for researchers. It guides researchers through a step-by-step workflow in which they specify study properties relevant to legal, ethical, and data-sharing requirements, along with contextual metadata for future reuse.

¹While XLing is still under active development, it is planned for independent release as an open-source taxonomy after completion of the project, thereby enabling continued extension, maintenance, and broader reuse.

3.1. Consent Generation

During form completion, the platform continuously evaluates the evolving study description in real time. Based on this semantic evaluation, it dynamically adapts the workflow by introducing additional questions or enriching generated documents whenever required by applicable legal or ethical constraints. This enables context-sensitive interaction: as researchers refine their study design, the Wizard incrementally refines the required information, ensuring that follow-up questions are only triggered when semantically and legally necessary. Output documents are generated from configurable XML-based templates and rendered in parallel as live previews within the user interface.

The screenshot displays the 'INF Consent Form Wizard 1.1' interface. The top navigation bar includes 'My Studies', 'Info', 'Suggestions / Support', and a 'DE' language selector. The main area is split into two panels. The left panel, titled 'General Study Information: Who Is Involved?', contains a form with the following fields: 'Name of the project leader' (with 'Project Leader' entered), 'Phone number of the project leader' (with a placeholder 'Please enter phone number here'), 'Email address of the project leader' (with 'project@leader.de' entered), a radio button question 'Is the contact person the same as the project leader?' (with 'No' selected), 'Name of the contact person' (with 'Contact Person' entered), 'Phone number of the contact person' (with '123456' entered), and 'Email address of the contact person' (with 'contact@person.de' entered). The right panel shows a live preview of the generated consent form. It features the 'UNIVERSITÄT BIELEFELD' logo and contact information for 'Contact person for inquiries: Name: Contact Person, Tel.: 123456, E-mail: contact@person.de'. The form content includes a title '1 - Information for Participants', a greeting 'Dear Participant,', a thank you message, a 'Purpose of the Study' section stating it's a demo instance, 'Responsible Parties for This Study' listing the Faculty of Linguistics and Literary Studies and the Faculty of Sociology, 'Details of the Study Procedure' mentioning 2 sessions, and a list of tasks for participants. It also includes a statement about anonymous data collection and a 'Benefits and Risks' section.

Figure 1: Screenshot of dynamically generated consent form based on researcher input.

Once the form is completed and validated by the system, researchers can export the generated output as a bundle of PDF documents, intended for distribution to participants to ensure transparency and formal consent. In addition, the study configuration can be fully exported as structured JSON, enabling reproducibility and reuse in subsequent or adapted studies.

The platform is designed to be fully configurable with respect to study properties, workflow structure, and evaluation rules, allowing adaptation to different research domains. To support this, we are developing a fully interactive editor for configuration templates, enabling domain experts from other research areas, together with legal advisors, to define the structure and semantics of input and output forms. To ensure semantic consistency and interoperability, the platform supports the import and reuse of controlled vocabularies from RDF-based taxonomies and ontologies (see Section 2.2).

3.2. Usability Test

As part of an iterative development cycle [10], we conducted a small pilot usability study with four linguistics researchers using real-world study scenarios. The study aimed to evaluate the initial Wizard release and identify potential improvements before broader deployment.

Participant Selection and Test Scenarios: The pilot participants were selected to cover two types of testing scenarios: (i) detailed studies involving extensive options for personal data collection, and (ii) simpler studies with no personal data requirements. The goal was to assess the Wizard's effectiveness across varying study complexities. Each group included one early-career (PhD) and one experienced researcher (postdoc or professor), who generated consent forms based on realistic study scenarios.

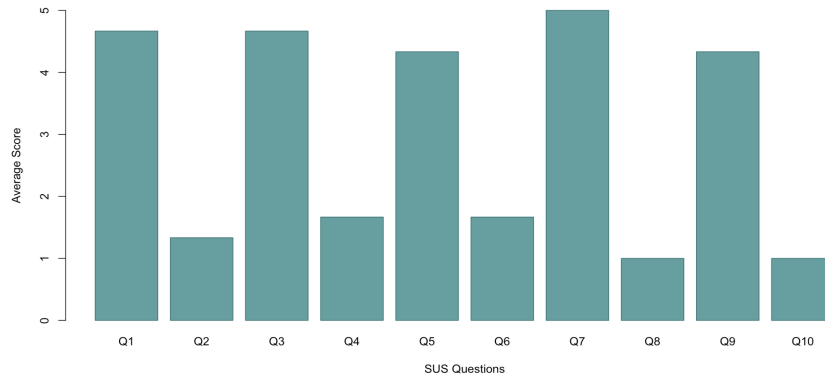


Figure 2: Average SUS item responses across participants.

Testing Method: Participants used the Consent Form Wizard web application on their own devices while their interactions were observed. Evaluation combined user observation [18], the System Usability Scale (SUS) questionnaire [19], and semi-structured interviews covering content, functionality, learnability, and overall experience. Each session lasted approximately one hour.

Results: The SUS scores (92.5, 90.0, 85.0, and 90.0) indicate high perceived usability, although results are based on a small sample. Participants consistently rated positive usability aspects highly and negative aspects low, indicating coherent perceptions across different expertise levels. Furthermore, the participants found the system intuitive and practical, highlighting the clarity of the workflow. Notable positive remarks beyond the evaluation framework included the overall layout of the design and information flow of the application, as well as the integration of ethical considerations beyond purely legal requirements.

3.3. Future Outlook

Beyond current consent generation, the Wizard is designed as a foundation for a semantically grounded data management infrastructure. Future extensions will enable (i) participant-side consent management, allowing granular and machine-readable consent decisions on data use, and (ii) ontology-based linking of consent records with dataset metadata to support policy-aware access, sharing, and reuse. Together, these components transform the Wizard from a document-centric tool into a consent-aware infrastructure for Open Science workflows.

4. Summary

In this paper, we presented a modular, ontology-driven framework, generating legally and ethically compliant consent forms. Developed with expertise from linguistics, law, and computational modelling, the Wizard platform assists researchers in adapting consent documentation to study-specific requirements. The system relies on standard ontologies and controlled vocabularies to ensure semantic consistency and interoperability within infrastructures. Currently in a pilot phase, the platform is being iteratively refined toward supporting structured consent management, including storage, modification, withdrawal, and lifecycle tracking of consent records. Future extensions will link consent specifications with dataset metadata to enable policy-aware access, controlled sharing, and reusable research data. This positions the platform as a semantically grounded infrastructure component for consent-aware data management in Open Science environments.

5. Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): SFB 1646/1 2024 – 512393437, project INF.

Declaration on Generative AI

During the preparation, the authors used GPT-4 for grammar and spelling check. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] B. Esteves, D. Golpayegani, G. P. Krog, H. J. Pandit, J. Flake, P. Ryan, Data Privacy Vocabulary (DPV), version 2.2, Final Community Group Report, World Wide Web Consortium, Wakefield, MA, USA, 2025. URL: <https://w3c.github.io/dpv/2.2/dpv/>.
- [2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3 (2016) 160018. doi:10.1038/sdata.2016.18.
- [3] A. Jorschick, P. T. Schrader, H. Buschmeier, What can I do with this data point? Towards modeling legal and ethical aspects of linguistic data collection and (re-)use, in: *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024, ELRA and ICCL, Torino, Italy, 2024*, pp. 47–51. URL: <https://aclanthology.org/2024.legal-1.8>.
- [4] H. J. Pandit, A. Polleres, B. Bos, R. Brennan, B. Bruegger, F. J. Ekaputra, J. D. Fernández, R. Gachpaz Hamed, E. Kiesling, M. Lizar, E. Schlehahn, S. Steyskal, R. Wenning, Creating a vocabulary for data privacy: The first-year report of data privacy vocabularies and controls community group (DPVCG), in: *Proceedings of the 18th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2019)*, Springer, Rhodes, Greece, 2019, pp. 714–730. doi:10.1007/978-3-030-33246-4_44.
- [5] A. Strathmann, M.-L. Joppek, M. Mohammadi, K. Politt, P. T. Schrader, A. Jorschick, H. Buschmeier, Modelling legal compliance in a consent wizard application as part of a research-centered and user-oriented data infrastructure, in: *Joint Workshop on Legal and Ethical Issues in Human Language Technologies (LEGAL 2026) and Computational Approaches to Language Data Pseudonymization, Anonymization, De-identification, and Data Privacy (CALD-pseudo 2026)*, ELRA, Palma, Mallorca, Spain, 2026, pp. 1–8.
- [6] P. T. Schrader, M.-L. Joppek, Datenbasierte forschung und einwilligungsunfähige. zulässigkeit der verarbeitung von daten minderjähriger und geistig eingeschränkter personen, *Zeitschrift für Datenschutzrecht* 2025 (2025) 613–618. URL: <https://beck-online.beck.de/?vpath=bibdata/zeits/ZD/2025/cont/ZD.2025.613.1.htm>.
- [7] M. Mohammadi, K. Politt, A. Jorschick, Assessing data management and compliance in large research collaborations via knowledge bases: A semi-structured interview approach, *F1000Research* 15 (2026) 37. doi:10.12688/f1000research.173178.1, [v1; peer reviews: 2 ‘approved with reservations’].
- [8] M. Herklotz, L. Oberländer, iva: Ein interaktiver virtueller assistent von berd@bw zur aufbereitung von rechtsfragen im bereich open science, in: V. Heuveline, N. Bisheh (Eds.), *E-Science-Tage 2021: Share Your Research Data*, heiBOOKS, 2022, p. 306–313. doi:10.11588/heibooks.979.c13742.
- [9] J. Nielsen, Heuristic evaluation, in: J. Nielsen, R. L. Mack (Eds.), *Usability Inspection Methods*, Wiley, New York, NY, USA, 1994, pp. 25–62.
- [10] M. Matera, F. Rizzo, G. T. Carughi, Web usability: Principles and evaluation methods, in: E. Mendes, N. Mosley (Eds.), *Web Engineering*, Springer, Berlin, Germany, 2006, pp. 143–180. doi:10.1007/3-540-28218-1_5.
- [11] C. Bizer, T. Heath, T. Berners-Lee, Linked Data – The story so far, *International Journal on Semantic Web and Information Systems* 5 (2009) 1–22. doi:10.4018/jswis.2009081901.
- [12] J. P. McCrae, A. Abele, P. Buitelaar, R. Cyganiak, A. Jentzsch, The linked open data cloud, 2007. URL: <https://lod-cloud.net/>, accessed: 2026-04-20.
- [13] H. J. Pandit, B. Esteves, G. P. Krog, D. Golpayegani, J. Flake, Data Privacy Vocabulary (DPV) –

version 2.0, in: The Semantic Web – ISWC 2024, Springer, Cham, Switzerland, 2025, pp. 171–193. doi:10.1007/978-3-031-77847-6_10.

- [14] A. Polleres, B. Esteves, G. P. Krog, H. J. P. and, Personal Data Categories (PD): Extension to the Data Privacy Vocabulary (DPV), Final Community Group Report, World Wide Web Consortium, Wakefield, MA, USA, 2026. URL: <https://w3id.org/dpv/pd>.
- [15] A. Miles, S. Bechhofer, SKOS Simple Knowledge Organization System Reference, W3C Recommendation, World Wide Web Consortium, Wakefield, MA, USA, 2009. URL: <http://www.w3.org/TR/skos-reference>.
- [16] ISO, ISO 639:2023 – Code for individual languages and language groups, International Standard 639:2023, International Organization for Standardization (ISO), Geneva, Switzerland, 2023. URL: <https://www.iso.org/standard/74575.html>.
- [17] ISO/IEC, ISO/IEC 3166-1:2020 – Codes for the representation of names of countries and their subdivisions – Part 1: Country code, Standard 3166-1:2020, International Organization for Standardization (ISO), Geneva, Switzerland, 2020. URL: <https://www.iso.org/standard/72482.html>.
- [18] S. Riihiäho, Usability testing, in: K. L. Norman, J. Kirakowski (Eds.), The Wiley Handbook of Human Computer Interaction, John Wiley & Sons, Ltd, 2018, pp. 255–275. doi:10.1002/9781118976005.ch14.
- [19] J. R. Lewis, The system usability scale: Past, present, and future, International Journal of Human–Computer Interaction 34 (2018) 577–590. doi:10.1080/10447318.2018.1455307.