Modelling Legal Compliance in a Consent Wizard Application as Part of a Research-Centered and User-Oriented Data Infrastructure

Aliena Strathmann[†‡] **Marc-Levin Joppek**[§‡] **Maryam Mohammadi**[†‡] **Katja Politt**[†‡‖]
Paul T. Schrader[§‡] **Annett Jorschick**[†‡] **Hendrik Buschmeier**[†‡]
[†] Faculty of Linguistics and Literary Studies, [§] Faculty of Law, [‡] CRC 1646 'Linguistic Creativity in Communication', Bielefeld University, Bielefeld, Germany
[‖] Rostock University, Rostock, Germany
{firstname.lastname}@uni-bielefeld.de

Abstract

Recent research calls for data management infrastructures that explicitly operate within the bounds of ethical and legal constraints, and facilitate adherence to Open Science principles by integrating automated support for planning, collection, storage, use, reuse, and sharing of data within. Legal and ethical requirements of data processing have become increasingly complex, introducing administrative barriers to scientific research investigating data generated by human participants, which encompasses a vast majority of humanities research. In response to this, we present RUDI ("Research-centered User-oriented Data Infrastructure"), a modular framework grounded in an interdisciplinary approach informed by legal, computational and linguistic expertise. This paper introduces its first component: a configurable and dynamically adaptive consent form generator in the form of a wizard web application. We outline how legal aspects are modelled within, and highlight its concrete benefits for administrative aspects of research. Further, we discuss the contextualisation of data within the research domain by leveraging the use of standardised ontology within the framework.

Keywords: informed consent, data management, data life cycle, personal data, GDPR, linguistics, ontology

1. Introduction

Research involving data generated by human participants operates at the intersection of two normative commitments. On the one hand, Open Science practices seek to promote transparency and collaboration benefiting all of society, and 'FAIR' principles (Wilkinson et al., 2016) correspondingly dictate that research data should be Findable, Accessible, Interoperable, and Reusable. On the other hand, strict ethical and legal limitations on data collection, processing, management, use and re-use reinforced by the General Data Protection Regulation in the European Union (GDPR, 2016) and similar legislation adopted elsewhere (e.g., Brazil, Canada, Israel, California) aim to protect the privacy and autonomy of the individuals who are the 'data subjects', i.e., the original sources of said data. Both objectives are of considerable importance, but often in conflict with each other. Reconciling them presents a substantial challenge for researchers who collect, process, and store human-generated data.

In response, recent work has argued for data management infrastructures that embed legal and ethical requirements directly into user-friendly and comprehensive technical architecture that supports research workflows, integrates and promotes awareness of legal and ethical compliance aspects, and facilitates Open Science practices across research projects and throughout the entire data life cycle (e.g., Siegert et al., 2020; Kamocki and Witt, 2024; Jorschick et al., 2024).

In this context, we present 'RUDI' (Research-centered User-oriented Data Infrastructure), developed within the 'INF' project of the Collaborative Research Center CRC 1646 *Linguistic Creativity in Communication* at Bielefeld University, Germany, where heterogeneous study designs, data types, and participant populations across projects are the norm rather than the exception.

RUDI is an interdisciplinary infrastructure framework developed in close collaboration between legal, linguistic, and technical experts. Conceptually, it specifies how legal norms, ethical constraints, and research-specific requirements can be represented in a structured and machine-actionable way. Its central, practical goal is providing a comprehensive data management platform that implements these specifications and supports researchers in planning, data collection, storage, controlled access, use and re-use of human-generated data.

Core features of this data management platform should enable researchers to

(i) inform and automate the process of creating meaningful (i.e., GDPR-compliant and ethically sound) informed consent forms and related materials for participants,

(ii) collect, store, and access individual instances of participant consent pertaining to specific collected data points,

(iii) index and contextualise available data within a research domain in order to facilitate sharing

Proceedings of LEGAL2026 and CALD-pseudo 2026 @ LREC 2026, pages 73–80
12 May 2026.

and reuse in accordance with Open Science principles, and conversely

(iv) locate, retrieve and reuse available data within a research domain restricted to the boundaries of participant consent.

While originating within the field of linguistics and spanning the specific research domains of the CRC 1646, the platform is designed to be adaptable to the requirements of any field of research involving human participants.

In this paper, we present the first component of RUDI and the data management platform: the web-based consent wizard. The technical implementation of the principles outlined above adheres to established principles of interface design, in particular Nielsen's ten usability heuristics (Nielsen, 1994), and to ensure an efficient and user-friendly workflow. Additionally, we employ an iterative development process that incorporates continuous user testing and systematic integration of user feedback (Matera et al., 2006).

2. Consent Wizard Web Application

The first stage of the platform's implementation, mapping to its core feature (i), is the 'consent wizard', a dynamic and configurable web application that allows researchers to easily generate consent forms and other information material for participants that are tailored to the specific context and requirements of a study. They are both for participants to ensure informed consent as well as for the researchers to be informed of legal specifications pertaining to their research from the participants' perspective.

The consent wizard, pending user feedback integration as part of iterative development, is functionally implemented as a web application and (for now) stand-alone component of the data management platform.[1] It maps a set of legal properties to a study based on the researcher's inputs on a dynamically generated, questionnaire-style form, and automatically generates corresponding output documents for potential participants to review and sign.

The current implementation still requires the researcher's manual involvement both in having participants sign and then managing the resulting consent forms. At this point, researchers are the only immediate users of the wizard. Future versions will incorporate participants as users of the platform by providing them the possibility to fully or partially consent[2] from within the platform. This consent data is stored in a database, eliminating the need for manually distributing and managing (signed) legal documents.

The wizard application is where the vast majority of legal aspects of the framework are situated and modelled in. The following section details how this is implemented in practice.

3. Modelling Legal Compliance

With modularity as a core design principle of the platform infrastructure, the wizard's software architecture heavily relies on the use of configuration templates. This facilitates legal and ethical compliance by leveraging knowledge from law and ethics experts for configuration, as well as allowing for swift and comfortable adaptation of the wizard to future changes in legislation.

Dynamic form generation. The *core configuration template* of the wizard defines a set of legally and ethically relevant "properties" that a study may assume, with the choices informed by legal expert counsel. Examples of these properties include:

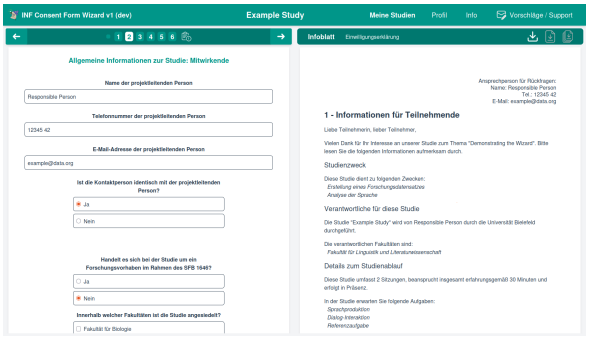
- processing of personal data according to the GDPR (see Section 3.2); represented as a binary value,
- permitted age groups of the participants; represented as a list of standardised string keys,
- participants under 14 years of age; represented as a binary value which may be programmatically inferred from the list of permitted age groups.
- whether legal guardian consent may be required for the participants; represented as a binary value.

A corresponding *form steps template* contains instructions for the program to dynamically assemble a questionnaire that maps the researcher's answers to this set of properties as concrete values, with the ability to conditionally render pages and questions (or other components) based on specific property values. Figure 1 exemplifies this: if the user specifies that the contact person of the project differs from the person responsible for the project, additional questions are displayed to collect the contact person's information.

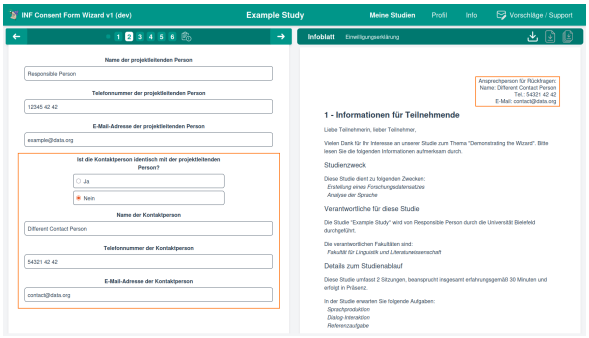
The wizard then uses these values to adapt text parts within the output documents intended for the participant, showing changes to the researcher in a live preview. The output forms, comprised of participant information and consent forms, are generated from configurable XML files based on document templates provided by a legal expert.

[1] https://purl.org/crc1646/RUDI-wizard

[2] Partially consenting means opting out of consent to specific data processing steps, e.g., publication or third-party sharing.



(a) The user specifies that the contact person is identical with the individual responsible for the project.



(b) The user specifies that the contact person differs from the individual responsible for the project.

Figure 1: Example demonstrating the consent wizard's conditional rendering of input form components. Additional input components are rendered in (b), and the output is adapted accordingly.

With every input, the wizard evaluates completeness and validity of the researcher's answers. If the form is determined to be sufficiently completed, the researcher is able to download the finalised consent and information forms in PDF format. In future versions it will be possible to make the output forms accessible to participants within the platform for review and electronic signing.

3.1. Modelling Legal Constraints

A third configurable template defines *autofill rules* based on legal constraints informed by expert counsel, and instructs the wizard to conditionally set and 'lock' certain properties based on present sets of property values per study instance.

Example of legal guardian consent. Study participation may require guardian consent under certain circumstances, e.g., for children or people with mental disabilities affecting their cognition (Schrader and Joppek, 2025). This is due to the fact that the GDPR requires the data subject's capacity to consent in order for consent to be effective. If the person giving consent is not capable of doing so, consent must be given by their legal representative. The GDPR does not define in detail when exactly the data subject is capable of giving consent and how this is determined. Essentially, it is important that the person giving consent is able to sufficiently understand the data processing that concerns them.

Usually, the researcher conducting the study must independently assess whether personal data are involved. Criteria such as the purpose, type, and scope of data processing as well as mental maturity can be included in the assessment. Nevertheless, it may not be possible to make an unequivocal judgment in individual cases. In cases of doubt, both the consent of the person concerned and that of their representative should therefore be obtained as a precautionary measure.

Fixed age limits may be considered as a possible solution. In the case of consent by minors, Art. 8 GDPR provides partial age thresholds for processing operations in certain contexts. However, due to the narrow scope of the provision, these requirements cannot be easily transferred to consent into research contexts and instead provide, at best, rough guidance. For other vulnerable groups, such as people with mental disabilities, such age limits do not apply. It therefore remains that the determination of capacity to consent is fundamentally case-specific.

Guardian consent within the wizard. The wizard explicitly asks whether the researcher presumes that legal guardian consent may be required, with an expandable information display summarising the legal situation outlined above. In our present *autofill rules* configuration template, the requirement of legal guardian consent is automatically set for participants under 14 years of age, or for participants under 18 years of age if the researcher has selected any special risks to the participants' mental or physical well-being associated with study-participation.

Based on this configuration setting, the autofill evaluation routine determines for each study instance whether either of the above conditions applies,[3] and locks the binary choice component to affirmative input, as shown in Figure 2. Furthermore, a corresponding explanation is displayed inside a yellow box and a clickable information display additionally outlines relevant legal situations where guardian consent may be required as described above. The resulting output documents are adapted accordingly, with additional forms for parents or legal guardians to sign.

[3]This mechanism is presently simplified, as we only have access to information pertaining to the study, not yet to information pertaining to the individual participant. Future versions that integrate participants as direct users of the platform will adapt this.

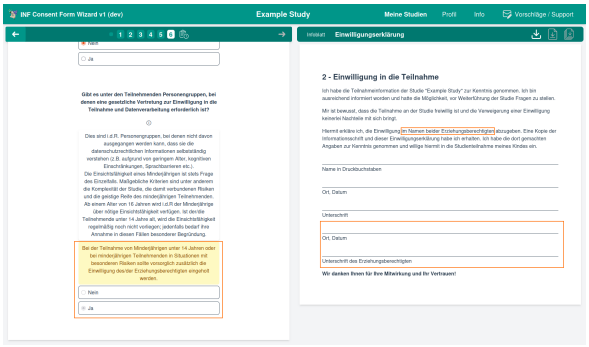


Figure 2: Example demonstrating a locked input component as a result of the autofill evaluation routine. The binary choice input is locked based on the researcher's prior input, with a corresponding explanation shown in a yellow box.

This same process can be applied to any number of explicit legal constraints based on an arbitrary set of determining parameters. Values that are conditionally (and automatically) set may correspond to (and override) parameter values that the user may otherwise set directly, as demonstrated in the example. In principle, they may also map to latent parameters without being directly bound to any specific input component; allowing full administrative control over setting conditions that determine the flow of information via the input form steps and study parameter templates.

3.2. Considerations for Personal Data

An important aspect of data management guidance for research projects in general, but especially for the consent wizard, is to clearly distinguish between *personal* and *non-personal* data in a way that is understandable to researchers. Understanding what is considered personal data is of central importance: the GDPR (2016) applies only when personal data are actually processed in accordance with Art. 2 para. 1 GDPR. This means, among other things, that the processing must be carried out in accordance with the principles of Art. 5 para. 1 GDPR, be based on a legal basis in accordance with Art. 6 para. 1 GDPR and the data subjects must be adequately informed about the data processing in accordance with Art. 13, 14 GDPR.

The term 'personal data' is defined in Art. 4 (1) GDPR. According to this, personal data are any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is someone who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Recital 26 GDPR specifies this as follows: to determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. Ultimately, whether an information is relatable to an individual must be determined from the controller's perspective. As a consequence, pseudonymised data cannot necessarily be regarded as personal data in every case (C-413/23 P, para. 86; of Justice of the European Union, 2025). Whether or not personal data is processed in a specific situation, often depends on the circumstances of the individual case. For reasons of transparency, obtaining consent under data protection law purely as a precautionary measure and providing data protection information in accordance with Art. 13, 14 of the GDPR without first reliably establishing whether personal data are in fact being processed should be avoided. Nevertheless, even in cases of exclusively anonymous data processing, consent to study participation is advisable for ethical reasons (e.g., Gauthier et al., 2010).

Data protection training courses or fact sheets are particularly suitable for communicating to researchers who are not legally experienced which data is considered personal data in individual cases. These measures can initially create a sound basic understanding. If there are still uncertainties in a specific case, these can be resolved through low-threshold counselling if necessary. In the long term, however, the aim is to minimise external input as much as possible: The categorisation of whether or not personal data is being processed should be computationally supported within the tool.

Personal data within the wizard. We currently determine whether personal data are processed by letting the researcher make this distinction via a binary selection. In cases of uncertainty, we refer the user to a third-party tool ('iVA'; Herklotz and Oberländer, 2022), which guides them through a four-step decision process. In future versions, this decision process will be integrated into the wizard's own questionnaire.

If personal data are determined to be processed, information forms are required to reflect

- which particular personal data is collected for the purpose of the research goal, and
- the period for which the data will be stored,

as well as further information concerning data processing, such as sharing and publication. When this is the case, the questionnaire steps of the wizard are expanded accordingly, and a separate information form on the processing of personal data is

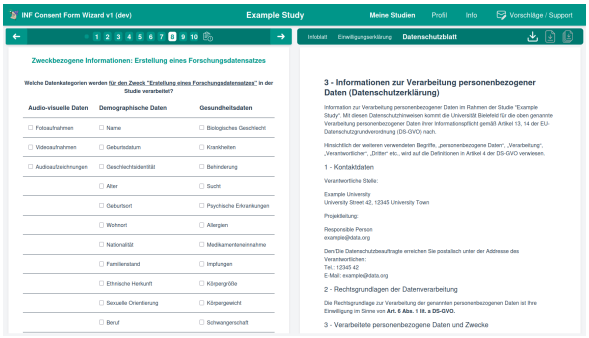


Figure 3: Example showing one of the extrapolated questionnaire steps if personal data is collected as part of the study, inquiring about specific personal data types collected and the data retention period per data processing goal. An additional information form with GDPR-specific information for the participant is populated with the additional input, and added to the set of output documents.

added to the output documents (see Figure 3) to comply with Art. 13, 14 of the GDPR.

3.3. Availing Legal Expertise to Users

Besides *researchers* who design studies and collect, use, and manage data and represent the main user group of the current stage of platform development, *users* also encompass study *participants* who provide data ('data subjects' in GDPR) in future development stages. For both user groups, we generally assume limited legal expertise.

For researchers, breaking down legal concepts into modular sets of questionnaire components and omitting any aspects that are irrelevant to their specific situation substantially reduces the burden of navigating legal considerations as laypeople. However, beyond just alleviating administrative workload by automating the consent and data management process, our goal is also to enhance awareness of legal aspects by embedding educational support seamlessly into the interaction process within the platform. To this end, the wizard features the ability for legal experts to embed extendable information displays into each questionnaire component via the form steps template: Figure 4 shows an example component that asks for (optional) information pertaining to the disclosure of the study's source of funding. When clicked, the display informs the researcher that the source of funding should be disclosed especially if there is reason to assume that this knowledge might influence the participant's decision. Another example of an information display is shown in Figure 2.

Additionally, the wizard's form steps template allows for full configurability in displaying conditional warnings and reminders: for example, a visually highlighted warning is shown that any *published*

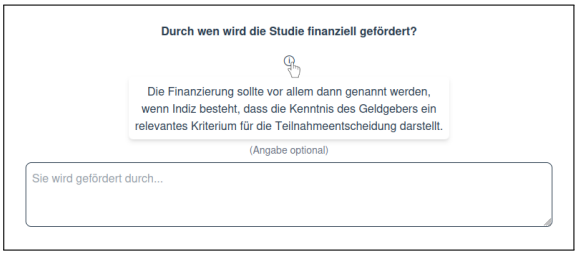


Figure 4: Example showing an information display for a text input component of the wizard questionnaire, pertaining to when disclosure of the study's source of funding is advisable. Clicking the info button displays the additional text box.

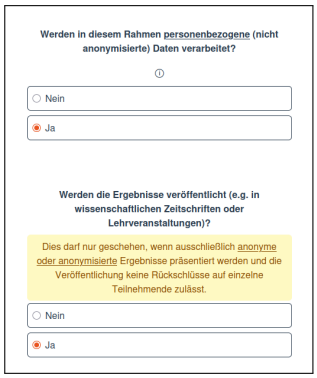


Figure 5: Example of a visual highlight warning, which is only displayed if both binary choice components shown in the image are answered affirmatively by the user.

results must always be anonymous or anonymised, and may never allow for drawing conclusions about the identity of individual participants (see Figure 5). This is only shown if the researcher selects that (a) personal data is being processed for the study, as well as (b) there is an intent to publish the results, e.g., in scientific journals, or use them within lectures or seminars. This is achieved by outfitting input components with warning text and corresponding 'warnIf' conditions within the form steps template, which are dynamically evaluated at runtime based on the current study parameter values.

This dynamic evaluation of study parameter conditions also allows for displaying situational suggestions for best practices and measures that may simplify the downstream data management process. Examples of this are yet to be implemented in the current version of the wizard; we are presently attempting to determine specific sets of personal data collection conditions that allow for anonymisation or pseudonymisation, which would both protect the data subject's privacy and avoid special consideration requirements under GDPR. Those conditions could then be embedded into the template and algorithmically flagged, and enable the application to inform the researcher of the possibility when given.

Another related feature that promotes researcher awareness of legal ramifications is the automated, conditional setting of values and locking of inputs based on a separate constraints template, as shown in the example already mentioned in Section 3.1: if a specific set of values set earlier dictates a future value, the researcher is informed by clear visual feedback and explanation (see Figure 2).

Finally, to benefit both user groups, great care is taken to translate legal concepts into easily understood (albeit legally accurate) language, tailored to their respective perspectives; this reflects the position of the European Data Protection Board (EDPB) that consent-related information must be formulated in clear and plain language that is understandable to the average person (Board, 2020, p. 18). The use of configurable templates where possible ensures a quick integration and addressing of user feedback from both groups, e.g., by amending phrasing that is perceived as confusing by users in either the study questionnaire or output forms.

4. Integrating Linguistic Perspective

Modular configurability of the platform facilitates the iterative development process and accommodates various legal and ethical aspects – but beyond that, it also provides crucial infrastructure for integrating aspects and requirements that are relevant to the specific research domain. Mapping studies to a modular set of data collection-related properties enables multifaceted downstream processing, including contextualisation of the collected data within a larger research ecosystem. Combined with integrated recording and storage of participants' individual consent choices, it allows for automated decisions regarding which operations (e.g., processing, sharing, publication, or controlled access) are permissible for specific datasets or even individual 'data points' (Jorschick et al., 2024).

This integrated approach sets our tool apart from prior approaches that implement standardised guidance, such as the 'DARIAH Consent Form Wizard' (Hannesschläger et al., 2020), which also supports template-based generation of consent forms, or 'Ethiktool' (Bendixen et al., 2025, 2026), which provides software-guided collection of information relevant for ethics review while generating participant information and privacy-related documents. Both tools treat the generated documents as the final output, rather than as one step in the data life cycle.

4.1. Leveraging Ontologies

Following the FAIR principles, we investigate an ontology-based implementation to enhance the findability of collected data. The use of inconsistent terminology risks creating broken links between related data and may result in data being lost in search processes. Principles of data visibility and reusability require that newly collected data are semantically linked to existing resources. This is particularly important here, given that RUDI comprises multiple modules which must work together consistently and integrate seamlessly with external resources. Ontological resources are thus shared across all modules. Mohammadi et al. (2026) identify the relevant data types and associated (meta)data within the linguistics domain. Although such information could be coded directly into the platform, we use ontologies, taxonomies and controlled vocabularies, to facilitate semantic alignment and connect our data to the broader semantic Web.

Given the growing need to address personal data considerations in different domains, numerous authors have proposed corresponding legal taxonomies (e.g., Pandit et al., 2019). Since linguistic demographic data substantially overlap with categories of personal data, we adopt and adapt established ontologies and standards wherever possible. For widely used classifications, we rely on ISO standards, including ISO 639-3 for language codes (ISO, 2023) and ISO 3166-1 for country codes (ISO/IEC, 2020). We also use domain-specific vocabularies such as BioPortal and the WHO International Classification of Diseases for medical information. To model personal data, we use the GDPR-aligned Data Privacy Vocabulary (DPV; Pandit et al., 2025; Esteves et al., 2025).

Within the context of the CRC 1646 research domains, we are developing the eXperimental Linguistics taxonomy (XLing), which defines a minimal set of field-specific concepts. Crucially, XLing entries are linked to CLARIN vocabularies to enable future integration and reuse. We employ established semantic web standards, including Resource Description Framework Schema (RDFs), Dublin Core Terms (dcTerms), and the Simple Knowledge Organization System (SKOS). Notably, these schemata have been implemented dynamically within the wizard application and will be extended to downstream components of the platform. This ensures that additions or changes in values can be integrated immediately at any stage, while practical usage of the platform can, in turn, inform further development of the ontologies.

5. Summary and Future Perspectives

In this work, we have introduced the present implementation of our consent wizard application, which is being developed as the first component of a more comprehensive, modular and configurable research data management platform. As a central, practical goal, this web-based platform is situated

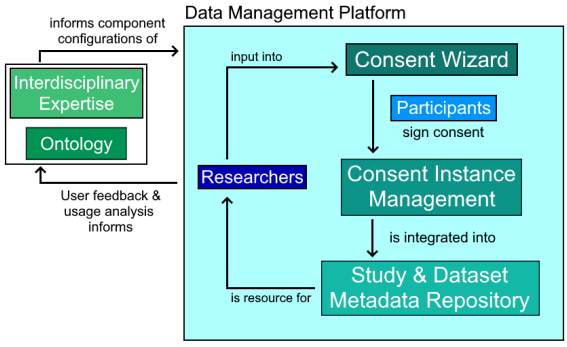


Figure 6: A visualisation of the planned data management platform, its components and its users situated within the Research-centered User-oriented Data Infrastructure (RUDI).

within our interdisciplinary infrastructure framework RUDI, which aims to support user-friendly linguistic research data management throughout the entire data lifecycle, by intuitively integrating the legal and ethical requirements of data management and facilitating Open Science practices and the FAIR principles. Figure 6 provides a simplified, general visual overview of how the wizard and the data management platform integrate into RUDI.

We have detailed how legal and ethical aspects of data processing are modelled and presented within the existing wizard application, and how they interface with linguistic research domain contextualisation aspects. Further, we have discussed the role of standardised ontology within RUDI, presented XLing as a taxonomy that is tailored to the specific research context of the CRC 1646, and how it integrates into the wizard component of the data management platform.

Further platform development. Pending server-side deployment, the consent wizard is to enter its first iterative development cycle with feedback from researcher users. The next concrete development milestone of the surrounding data management platform is the integration of an online database. Persistent, server-side storage enables implementing user management, which in turn enables (i) integrating participants as direct users, allowing for direct signing and storage of consent instances within the application and thus introducing the second core component of the platform, and (ii) sharing of study design templates between researchers. It also allows for collection and evaluation of usage (meta)data of the wizard and platform, laying the groundwork for further iterative platform development and meta-analysis within the surrounding research domain.

Following this, we plan to integrate the consent management capabilities into an open repository of metadata about empirical datasets; expanding the platform to act as a hub that facilitates discovery of legally reusable data and collaboration between the projects within its domain, thus fully encompassing the operationalisation of both legal compliance and Open Science practices.

General future goals of the project. While the platform is intended to be a stand-alone application, a general objective of the project is to establish compatibility with existing infrastructure supporting research workflows specific to Bielefeld University and, in the long term, with data infrastructures across Europe. An immediate goal is localisation of the wizard, the platform, and its configurations into English, which poses the challenge of creating legally accurate translations of, e.g., the wizard's input forms and output documents.

Pending completion of development, we intend to release the source code of the data management platform as a fully configurable and research domain-agnostic open-source project.

6. Acknowledgments

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

7. Bibliographical References

Alexandra Bendixen, Eva-Maria Berens, Thomas G. G. Wegner, Wolfgang Einhäuser, and Katarina Blask. 2025. Data (re-)use in human-participant research: Guided composition of informed consent forms. In *Abstracts of the 2nd Conference on Research Data Infrastructure (CoRDI)*. Zenodo.

Alexandra Bendixen, Thomas G. G. Wegner, and Wolfgang Einhäuser. 2026. Facilitating ethics application and review for interdisciplinary human-participant research via software-based guidance and standardization. In Bertolt Meyer, Ulrike Thomas, and Olfa Kanoun, editors, *Hybrid Societies: Humans Interacting with Embodied Technologies*, volume 1, pages 311–317. Springer, Cham, Switzerland.

European Data Protection Board. 2020. Guidelines 05/2020 on consent under Regulation 2016/679. Technical report, European Data Protection Board. Adopted on 4 May 2020.

Beatriz Esteves, Delaram Golpayegani, Georg P. Krog, Harshvardhan J. Pandit, Julian Flake, and Paul Ryan. 2025. Data Privacy Vocabulary (DPV), version 2.2. Final community group report, World Wide Web Consortium, Wakefield, MA, USA.

Janel Gauthier, Jean Pettifor, and Andrea Ferrero. 2010. The universal declaration of ethical principles for psychologists: A culture-sensitive model for creating and reviewing a code of ethics. *Ethics and Behavior*, 20:179–196.

GDPR. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council. *Official Journal of the European Union*, L 119:1–88.

Vanessa Hannesschläger, Walter Scholger, and Koraljka Kuzman. 2020. The DARIAH ELDAH consent form wizard. In *DARIAH Annual Event 2020: Scholarly Primitives*. Zenodo.

Markus Herklotz and Lars Oberländer. 2022. iVA: Ein interaktiver Virtueller Assistent von BERD@BW zur Aufbereitung von Rechtsfragen im Bereich Open Science. In Vincent Heuveline and Nina Bisheh, editors, *E-Science-Tage 2021: Share Your Research Data*, page 306–313. heiBOOKS.

ISO. 2023. ISO 639:2023 – Code for individual languages and language groups. International Standard 639:2023, International Organization for Standardization (ISO), Geneva, Switzerland.

ISO/IEC. 2020. 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.

Annett Jorschick, Paul T. Schrader, and Hendrik Buschmeier. 2024. 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*, pages 47–51, Torino, Italy. ELRA and ICCL.

Pawel Kamocki and Andreas Witt. 2024. Ethical issues in language resources and language technology – New challenges, new perspectives. In *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024*, pages 19–23, Torino, Italia. ELRA and ICCL.

Maristella Matera, Francesca Rizzo, and Giovanni Toffetti Carughi. 2006. Web usability: Principles and evaluation methods. In Emilia Mendes and Nile Mosley, editors, *Web Engineering*, pages 143–180. Springer, Berlin, Germany.

Maryam Mohammadi, Katja Politt, and Annett Jorschick. 2026. Assessing data management and compliance in large research collaborations via knowledge bases: A semi-structured interview approach. *F1000Research*, 15:37. [v1; peer reviews: 2 'approved with reservations'].

Jakob Nielsen. 1994. Heuristic evaluation. In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, pages 25–62. Wiley, New York, NY, USA.

Court of Justice of the European Union. 2025. Judgment of 4 September 2025, Case C-413/23 P. ECLI:EU:C:2025:645, para. 86.

Harshvardhan J. Pandit, Beatriz Esteves, Georg P. Krog, Delaram Golpayegani, and Julian Flake. 2025. Data Privacy Vocabulary (DPV) – version 2.0. In *The Semantic Web – ISWC 2024*, pages 171–193, Cham, Switzerland. Springer.

Harshvardhan J. Pandit, Axel Polleres, Bert Bos, Rob Brennan, Bud Bruegger, Fajar J. Ekaputra, Javier D. Fernández, Roghaiyeh Gachpaz Hamed, Elmar Kiesling, Mark Lizar, Eva Schlehahn, Simon Steyskal, and Rigo Wenning. 2019. 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)*, pages 714–730, Rhodes, Greece. Springer.

Paul T. Schrader and Marc-Levin Joppek. 2025. Datenbasierte forschung und einwilligungsunfähige. zulässigkeit der verarbeitung von daten minderjähriger und geistig eingeschränkter personen. *Zeitschrift für Datenschutzrecht*, 2025:613–618.

Ingo Siegert, Vered Silber-Varod, Nehoray Carmi, and Pawel Kamocki. 2020. Personal data protection and academia: GDPR issues and multimodal data-collections "in the wild". *Online Journal of Applied Knowledge Management*, 8:16–31.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3:160018.