trans-kom ISSN 1867-4844 https://www.trans-kom.eu
trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom 19 [1] (2026): 307–328

Tanja Wissik

Insights into data sharing practices in terminology work in institutional settings

Results from an exploratory study based on qualitative expert interviews

Abstract

This paper examines data sharing practices in institutional terminology work within an exploratory study based on 15 qualitative expert interviews. The findings reveal a variety of sharing methods within the community, with APIs gaining in popularity. In contrast, data repositories remain underutilized for sharing terminological data. The study highlights a shift in data accessibility: away from traditional methods, like only web-accessible databases and dictionaries, to terminological data that is increasingly downloadable, or accessible via APIs or computer-assisted translation tools. However, not all interview participants have adopted FAIR (Findable, Accessible, Interoperable, Reusable) data principles due to challenges related to data sharing, including legal and copyright issues, technical issues like incompatible data formats, and social and cultural factors. These challenges underscore the complexity of implementing FAIR data practices in institutional contexts and call for more targeted support to foster best practices regarding FAIR data sharing in the terminology community.

1 Introduction

Sustainable and FAIR data sharing, i. e. data sharing in line with FAIR data principles (making data Findable, Accessible, Interoperable and Reusable), is increasingly used in both academia and the public sector to optimize the reuse of data (e. g. ALLEA 2020).

FAIR data principles were introduced by Wilkinson et al. (2016) with an initial focus on digital research data with the goal of improving its findability, accessibility, interoperability and reusability for data processing machines. That has now been extended to include other types of digital assets, such as cultural heritage assets or government data. “Findable” means data should be described with rich metadata, should have unique identifiers and be indexed in a searchable resource (e. g., repository) to facilitate discovery; “Accessible” ensures that data can be retrieved and that metadata is available, even if the data itself is not openly accessible. Furthermore, in order to be accessible by machines, a standardized and if possible open communication protocol should be used; “Interoperable” focuses on using standard formats and vocabularies to enable data

integration across different systems; finally, “Reusable” means data should come with clear licences and metadata that allow others to use it effectively. Together these principles promote better data management and sharing in research and beyond (Wilkinson et al. 2016).

In the context of Open Science, data sharing and the definition and application of community practices for data sharing are very important, as described in the UNESCO Recommendations on Open Science (UNESCO 2021). Furthermore, data is increasingly processed by machines, and terminological data is no exception (Chiocchetti/Lušicky/ Wissik 2023). Consequently, the FAIR data principles are becoming increasingly relevant for terminological data. The application of FAIR data principles in the context of terminology work has been discussed by several authors including Vezzani (2022) and Vezzani and Di Nunzio (2023). There are several initiatives that promote the adoption of FAIR data principles, ranging from the creation of FAIRterm (Vezzani 2021), a tool for the compilation of FAIR terminological entries, to the testing (Elbaz et al. 2022) or the implementation of termbases following the FAIR principles (Di Nunzio et al. 2024). At the same time, challenges remain in the implementation of the FAIR data principles. For example, Vezzani, Di Nunzio and Costa (2023) pointed out possible limitations of ISO standards related to terminology in terms of FAIRification of terminological data.

If we broadly define data within the field of terminology as all (digital or digitized) materials and assets that are collected, generated and used during the process of compiling terminological resources, data is part of every aspect of terminological work. Starting from the compilation and use of specialized corpora in the documentation phase to the extraction of terms to the creation of the final terminological product (e. g., terminological dictionary or terminological entries in a database), and even the metadata describing the terminological resource, every part of the process involves data that can be shared.

In this article, data sharing is to be understood as the practice of making data available and accessible to other users (including machines) for reuse. This study aims to explore data sharing practices in terminology work in institutional settings and to determine if the data sharing practices in use are in line with FAIR data principles. Specific factors that influence data sharing practices in institutional settings will be examined. The article is structured as follows: following a theoretical introduction on data sharing and FAIR data principles, Section 2 provides a literature review on data sharing in terminology work or related fields (such as lexicography or language resource management). Section 3 describes the research method and Section 4 provides an overview of the results from the interview data, which are then discussed in Section 5. Lastly, Section 6 presents a few concluding remarks.

2 Data sharing in terminology work and related fields

When looking at prototypical terminology workflows, data sharing is most commonly found during the publication step (e. g., Kudashev 2013; Arndt et al. 2020; Chiocchetti/Lušicky/Wissik 2023), even though it is not called data sharing, rather publication or dissemination (e. g., Cerrella Bauer 2015: 339).

Terminological data can be published via online public databases or as restricted versions on an intranet or as online or printed glossaries and dictionaries (e.g., Drewer/Schmitz 2017: 182; Chiocchetti/Lušicky/Wissik 2023: 470). In these traditional forms of publication, terminology can be consulted by humans, but the actual underlying data is often inaccessible and cannot be further processed by machines, as discussed by Chiocchetti, Lušicky and Wissik (2023: 470). A similar picture was drawn by Tiberius et al. (2024) and Kosem et al. (2021) regarding lexicographical data. Most dictionaries are accessible online for consultation, but the reuse of the underlying data is often restricted. Most of the terminological or lexicographical data can be considered "shy data", i. e. data that you "meet in public places, but you can't take […] home with you," as defined by Cooper (2010). This category of data is often "web-accessible" (Cooper 2010), as in an online terminology database or an online dictionary, but the data itself is not downloadable and therefore cannot be further processed.

Other ways of accessing terminological resources are to allow other tools, such as computer-assisted translation tools, to access terminological data via web services and Application Programming Interfaces (APIs) (Roche/Alcina/Costa 2019) or to provide data downloads (Zorrilla-Agut/Fontenelle 2019: 147–149). Other possibilities include the publishing of terminological data in data repositories with metadata being registered in a catalogue (Lušicky/Wissik 2019; Andersen/Gammeltoft 2022: 253; Chiocchetti/Lušicky/Wissik 2023: 471). Besides general-purpose repositories such as Zenodo, there are specific repositories for language resources, such as the repositories run by the technical centres within the European research infrastructure CLARIN, the so-called CLARIN B-centres (De Jong et al. 2022: 45–48) or ELRC-SHARE repository (Smal et al. 2020), which is a repository for storing, browsing and accessing language resources provided by the public sector. Terminological data can also be published as linguistic linked open data (Roche/Alcina/Costa 2019; Cimiano et al. 2020; Di Buono et al. 2020; Martìn-Chozas et al. 2024).

Since terminology work involves not only the creation of terminological data but also the creation of other language resources, such as corpora or controlled vocabularies, language data management models (e. g., language resource lifecycle models) are also relevant to this discussion (e. g., Rhem 2016). In Rhem's (2016) language resource lifecycle model, sharing is placed in phase 7 under distribution and publication of the language resource. Rhem (2016) discusses publication through web interfaces, via API and as downloadable formats; the possibility of using data repositories as an additional publication channel is not discussed. For the sharing of lexicographic data, there is a separate lifecycle model (Klosa-Kückelhaus/Tiberius 2025), where data sharing occurs

in two phases: in phase 6 “preparation for publication” and phase 7 “maintenance and preservation,” which also includes the archiving of data. For lexicographic data, the use of APIs and data downloads were discussed, along with the more traditional methods of publishing like online and print dictionaries (Tiberius et al. 2024) or dictionary portals (Klosa-Kückelhaus/Tiberius 2025).

As for the sustainable and FAIR sharing of terminological data, there is a recently established (2022) annual international conference series, “Multilingual digital terminology today. Design, representation formats and management systems,” with a conference topic devoted to the evaluation of structural approaches to terminological resource design and discussion of the optimal design of terminological resources in order to better respond to FAIR data principles (Di Nunzio et al. 2022).

Vezzani and Di Nunzio (2023) argue that the role of terminologists in the digital age also includes tasks like preservation and management of digital resources, as well as ensuring the accessibility and reuse of terminological resources. However, there are few studies that examine the terminological workflow and how terminologists and other stakeholders (e. g., translators) handle these new tasks such as data preservation, data sharing, optimising resources for reuse, etc. in daily terminological workflows. A recent study on terminology management practices of court interpreters and legal translators showed that the sharing of terminological data is relatively uncommon in this group (Wissik/Lušicky/Chiocchetti 2024).

There are some recent studies on how language data in general are managed and shared in public administrations and SMEs (ELRC 2019; Smal et al. 2020; ELRC 2022). These studies found that there are several circumstances that negatively affect data sharing practices such as the “lack of recognition of the value of language data, lack of digital skills, lack of adequate language data management practices/plans” and legal issues (ELRC 2022: 21).

Since there is no systematic overview of data sharing practices in terminology work, particularly in institutional settings, a qualitative exploratory study based on expert interviews was carried out in 2023 to reveal insights on the current state of data sharing in terminology work and to explore if the circumstances that negatively affect data sharing practices reported by ELRC (2019, 2022) can also be found in institutional terminology work. The following sections describe the methodology of this study, present and discuss the data collected and offer some concluding remarks.

3 Method

This exploratory analysis was conducted to attempt to fill the gaps left by the aforementioned lack of insights into data sharing practices in terminology work in institutional contexts. The findings in this article are based on the analysis of 15 qualitative expert interviews (Meuser/Nagel 1991: 443) collected as part of a broader study

conducted in 2023 that investigated the influence of emerging technologies and paradigms, such as open data, on terminology work in institutional contexts (Wissik 2024, 2025a). In addition, publicly available documents and information regarding data sharing (e. g., information on available APIs, available download options) of the involved institutions were also analysed and taken into account. Some aspects of the project have already been published, namely sustainability aspects of terminology work (Wissik 2024), which is related to data sharing practices, and the impact of term extraction on terminology work (Wissik 2025a). Therefore, reference is made to these two publications in the method section.

Experts were selected according to the following specific inclusion criteria: experts had to be working in an institution[1] that (1) conducts practical terminology work and (2) maintains a publicly accessible terminology database or terminological dictionaries; and experts themselves had to be regularly involved in practical terminology workflows. An initial pool of potential interviewees was identified through networks such as the European Association for Terminology (EAFT), Rat für deutschsprachige Terminologie (RaDT), and the Conference of Translation Services of European States (COTSOES). Additionally, relevant international organisations, including specialized agencies within the United Nations, were added to the list. From this pool, 15 experts consented to participate in the study. The sample encompasses experts from diverse types of institutions, representing a balanced mix of roles within terminology workflows and a variety of languages, which provides a relatively clear picture of the actual landscape of terminology work in institutional settings, especially in Europe. The sample included the following institutions: 4 regional/state administrations in Europe, 2 European institutions, 2 international organisations, and 7 academic/research institutions in Europe (universities or academy-based terminology institutes and language centres). As per the participation requirements, all institutions maintain publicly accessible online terminology databases or dictionaries, ranging from large databases with over 400,000 entries to smaller ones with fewer than 10,000 entries. Furthermore, the sample covered all major types of terminology work: systematic, ad hoc, translation-oriented/text-based, preparatory for standardisation (e. g., terminology planning), proactive, and a posteriori terminology work. Roles represented across these institutions included directors of terminology units, terminologists, terminology managers/coordinators, technology managers, developers/IT experts and members of standardizing committees (Wissik 2025a).[2]

The 15 semi-structured interviews were conducted online, primarily in English, with two conducted in German, each lasting approximately one hour. The interview guide (Dorer et al. 2025: 84) was based on a previous study on terminology workflows by Chiocchetti and Ralli (2012, Attachment A) and was modified to suit the objectives of this study and published on Zenodo.[3] Besides general questions about the experts and the

1 Experts from commercial settings or freelancers were not included in this study.
2 For more details on the interview participants, see Wissik (2025a).
3 The interview guide is available on Zenodo: Wissik (2023).

institutions themselves, there were questions regarding the underlying methodology applied when conducting terminology work, the target users, practical terminology management and workflows, the use of technology, the adoption of terminology standards and questions related to research infrastructures. In total, more than 16 hours of interview data was recorded, transcribed and anonymised by removing identifiable details, such as institution names, names of specific terminology databases or custom in-house tools, research project titles, and specific languages (Wissik 2025a). For the analysis, a qualitative methodology was employed in this study, namely a thematic qualitative text analysis (Kuckartz 2014). The categories were developed through a combination of concept-driven and data-driven development of categories (Kuckartz 2019): First, a coding frame was developed with categories deductively derived from the literature and the interview guide. Subsequently, subcategories were derived directly from the data (Wissik 2025a). As described by Dorer et al. (2025: 85), in qualitative data analysis the grouping and categorisation of the data can be supported by computer-assisted qualitative data analysis software. For this study the open-source tool CATMA (Gius et al. 2023) was chosen to manually annotate the data with the established categories. CATMA is a browser-based tool that provides digital support for the manual annotation and analysis of texts. The tool allows for both concept-driven annotation as well as data-driven annotation (Schumacher 2019).

Before presenting the results, the potential limitations of the study should be considered. One limitation is the relatively small sample size of 15 qualitative expert interviews. However, as demonstrated by Guest et al. (2006), data saturation in qualitative research can already be reached with as few as 12 interviews, suggesting that additional interviews do not necessarily yield substantially more insights. Another limitation concerns a bias in the data toward institutions that maintain a public termbase, which means that institutions engaged in terminology work without a publicly accessible termbase are not represented in the sample.

4 Results

4.1 Sharing terminological data

Terminological data can be shared directly during the workflow, for example with expert reviewers or members of standardising committees, or data can be shared at the end of the workflow as the final terminological resource (e. g., dictionary or glossary). This section focuses on the latter.

4.1.1 Traditional methods of data sharing

A more traditional method of sharing terminological data is via an online terminology database. All of the institutions represented by the interview participants had one, as this was one of the inclusion criteria for the interviews. Another way of sharing data is through

specific search platforms or portals where users can search through all available resources, even if they are not in the same database.

> But we also have [Name of the portal] portal […] this is the portal of all dictionaries of [Name of Institute], […]. Here are general dictionaries, these are all general dictionaries, these are etymological dictionaries, historical and also terminological dictionaries. So, it is possible to search through all dictionaries at the same time. (INT 3)

Often the data is not only consultable online via the termbase or the portal but also downloadable, as confirmed by nearly all interview participants. Only one participant (INT 15) stated that the institute's data was not downloadable. Others mentioned that they share the data upon request:

> We get requests [from translators], but usually [they] are [for] a single term or a small quantity of terms. Sometimes we get [a] request for a whole export, for their translation memory [tool]. We do not have the terms in sentences, so they are not using it in translation memories but in their [termbase] etc. It is not quite frequent, but it happens from time to time. (INT 7)

The licences applied to these data downloads varied from open licences to more restricted licences. For some institutions, individual contracts needed to be signed in order to receive a data download. Sometimes, no licences were applied, or participants were not aware of the type of licences applied.

One participant (INT 6) specified that a contract needs to be signed in order to receive a data export.

4.1.2 Sharing data through APIs or computer-assisted translation tools

Another increasingly popular option is sharing terminological data via Application Programming Interfaces (APIs) (Wissik 2024), and more and more terminological data and dictionaries are actually shared via APIs (Roche/Alcina/Costa et al. 2019; Zorrilla-Agut/Fontenelle 2019; Tiberius et al. 2024; WIPO 2025). APIs are flexible and reusable interfaces that are also developer-friendly. They support a range of standard data formats, enabling interoperability and integration across diverse systems (FAIR Process Framework n. d.).

The set of interview questions included a specific question regarding APIs (Wissik 2024).

Several interview partners stated that their institutions have APIs in place. The most common scenarios for API use were:

(i) to share terminological data in terminological portals. Terminological portals are portals where you can seach in multiple terminological resources that come from diverse source: "[the API] is already made for this terminological portal" (INT 3);

(ii) to share data with whole institutions, e. g., for "download [by] other institutions. We have agreements and they show our data too. Not for the single personal user who goes to our website" (INT 8);

(iii) to share data with single users: “interested users can also use the REST […] (REpresentational State Transfer) application programming interfaces (APIs) to consult [Name of the termbase] data directly from third-party tools” (INT 10).

Terminological data is often shared directly through computer-assisted translation (CAT) tools (Caffrey/Valentini 2019; Wissik 2024), particularly in the field of translation-oriented terminology work. Data can be shared via a regular export of the termbase and subsequent import into the CAT tool of choice: “[...] we do an export [of the termbase] each month that is then made available in the computer-assisted translation environment” (INT 13). Or the data can be shared live within the CAT tool via plugins, for example:

> [W]e offer this plugin […] and this is a direct connection from Studio to the terminological database on a segment per segment [basis] and getting the results and all this is stored in the cache so there is no local term base. (INT 12)

One interview participant mentioned that they do not have APIs in place for their terminological data, but they have it in place for other data, so they are aware of APIs, they have implemented them for other data sets, but not for their terminological data:

> No, we haven’t, but we are completely open to that [API], we have done it on other projects. So, we have an API for example for our [Name of the project] project which is heavily used. […] You know I don't have a good answer in the sense that our terminology database doesn’t have an API. Probably it should. We would be happy to make one if we could find the time and we knew exactly what the requirements are. As I said, we have done that type of technology for another data set and we have seen it really valuable. It is so help for me when someone with programming capabilities come to me and asked me for access to data that I don’t have to write a script to generate data dump that they can do it themselves via an API, it’s so valuable for them and it*s so valuable for me as someone who is managing a data set. So, it’s actually a good question but hasn't been raised a whole lot recently but it has over the years. (INT 2)

4.1.3 Sharing data through data repositories

Another option is to share terminological data through data repositories, as mentioned in Section 2. All participants were asked a specific question related to repositories: “Is the terminology data stored/archived in a repository?”

Few participants mentioned that they archive their terminological resources in repositories: “[A] couple of years ago we did some kind of an archiving [operation] with some kind of text archive, if I [remember] correctly” (INT 1). Another expert mentioned that the deposited resources are also downloadable from the repository: “Actually you can [download] for some dictionaries. From the repository. But not for all, because of copyright issues” (INT 3). Most of the interview partners did not use repositories for data sharing (Wissik 2024).

4.1.4 Sharing data as linked (open) data

The emergence of the Semantic Web and, more recently, the Linguistic Linked Data initiative has opened up new perspectives and opportunities in terminology work,

particularly regarding the practical implementation of terminological resources (Roche et al. 2019) and the sharing of terminological data. Drewer and Schmitz (2017: 161) have already mentioned Linked Open Data in their book on terminology management. Similarly, Chiocchetti, Lušicky and Wissik (2023) discussed Linked Open Data as an opportunity to disseminate terminological resources. Since terminology resources are usually available in formats like TBX or XML, the data needs to be converted. There are several case studies that focus on converting terminology resources into Linked Open Data formats (Wissik/Declerck 2019; Declerck/Wissik 2020; Di Buono et al. 2020; Maroto 2024; Martín-Chozas et al. 2024).

As such, this study examined if terminological resources created in institutional settings are also shared as Linked Open Data. A specific question regarding publishing terminology as linked (open) data was asked during the interviews.

This question revealed that most of the interview partners were not familiar with linked (open) data and related formats such as RDF, and that they were not providing their data as linked (open) data. Some interview partners mentioned that they had received requests regarding specific data formats such as RDF or SKOS (INT1, INT 12): “No, RDF hasn’t really [been a question] […]. The question of SKOS has come up, sometimes people asking about [SKOS]” (INT 1). Furthermore, the data showed that the requests concerning RDF data are coming from researchers: “Yes, we have [received requests], especially from academia, from researchers. They would be interested that we provide RDF or some more open linked compatible data formats. We are not there yet” (INT 12). Furthermore, one interview partner (INT 12) mentioned that providing their data as RDF was discussed and put in the work programme but it is not a priority:

> We discussed this and [put] it in the work programme some years ago. But at the end it’s not really [a] priority for the internal services […]. Open linked data remains something that is nice to have but does not really bring a benefit to our first target users. So, it has [been repeatedly] postponed and somehow left aside. (INT 12)

4.2 Sharing questions and answers of terminology consultation services

The previous subsections looked at sharing terminology resources that are the result of either systematic or translation-oriented terminology work.

A specific case of terminology resources are the answers to requests made by the public through terminology consultation services (e. g., via e-mail or online forms) (Žagar Karer/Fajfar 2023). These answers are of course shared with the person that made the specific terminological request, but they are also often shared in an anonymized way with the wider public on a website, as part of a dictionary or terminology portal,[4] etc. Since these answers often require context to be completely understood, it is difficult to

[4] For example, in the Fran.si portal (Fran n. d.), which brings together language resources and portals developed by Fran Ramovš Institute of the Slovenian Language at the Scientific Research Centre of the Slovenian Academy of Sciences and Arts (ZRC SAZU) and also includes language and terminological questions and answers.

integrate them directly into the data model of a traditional terminological entry. Several ways of sharing these answers were mentioned in interviews.

> [Often these answers (in anonymised form) are published as unstructured data on a website] then we also publish it here [on our website], as a question and answer, but […] it's anonymised. (INT 3)

In other cases, Excel sheets were used to share questions and answers: "In our expert opinions that we have given to answer consultations we are also trying to give a bit more context in our databases. We have done that in Excel for a while […]" (INT 5).

Another option is to create a specific terminology database just for these questions and answers ("consultations" as one interview partner called them) because they do not fit into the traditional terminology database structure: "[…] because usually it's very, very complicated to put it into a single termbase entry, we have now a separate multiterm database now which is specifically made […] for these consultations" (INT 5).

However, some interview partners mentioned that they also publish the final results to user requests from the query system in their terminology database: "So users ask us about things we don't have in our database. So, we have to check for solutions. Do the file and publish it [in Name of the inhouse terminology database]" (INT 8).

4.3 Sharing corpus data

Using and creating corpora when doing terminology work is not new (Bowker 1996). Often existing corpora are consulted, as for example in ad hoc terminology work or when answering consultations, but specialized corpora are also created for specific domains when creating terminology databases from scratch, expanding them to new domains or creating specialised dictionaries (Wissik 2025b). But what happens to these corpora? Are they published or shared in some format? Experts were asked a set of questions related to corpora to gain insight on corpora in general; one question was especially relevant for the purposes of this study: "If you use your own corpora, do you publish them as well (e. g., licence)?"

Several methods used by institutions for handling corpus data emerged from the interviews. Some institutions publish the corpora they develop, depending on copyright (see also Section 4.4.1), to provide more resources for the language they support, as well as to assist in the development of technologies for the languages they work with, as the following excerpts from interviews with two participants (INT 2, INT 4) illustrate:

> [A]s part of our work in terminology we have built up parallel corpora, monolingual corpora and […] anyone doing machine learning or artificial intelligence or anything like that wants access to the data. […] We want to support [Name of less-resourced language], and one instance is technology, […] so number one is technology. So, if we, as part of our work, develop valuable data […] we make [the data] available to researchers [for] developing technologies on top of that and number two is to make our dataset available in an extremely useful way to non-technical users, who are non-scientists. (INT 2)

> We do publish a lot of our corpora […] we have the contemporary corpus of [a less-resourced language] which is about 100 million [words] of [this] contemporary [Name of less-resourced language] published over the last decade or so and we have a [Name of legislative parallel corpus] of materials from EU legislation and national legislation and they are both freely available on our website. (INT 4)

Some of the interview partners who create their own corpora do not publish them because they view them as internal working materials and not as a final resource: "Yes they are for internal use and there is one corpus per dictionary" (INT 9).

There was also a third scenario, a kind of in-between scenario, where the corpora are shared internally and also with external users through corpus management systems, although they are not published publicly (Wissik 2025b):

> [W]e store them [corpora] in SketchEngine, it's collaborative so you can share the corpus with other people in the organisation or outside the organisation so that's very useful and we typically leave it there. I mean we don't export […] Sometimes we will use them but we don't publish them or we don't, you know, otherwise store them except for SketchEngine where we have a licence and some storage. (INT 13)

The evidence from the interviews shows that there are institutions that share the corpora that are built as part of their work in terminology, and others that do not share them because they are considered internal working materials. If corpora are shared publicly, the form in which they are shared, again, often depends on copyright, as will be discussed in Section 4.4.1.

4.4 Challenges to data sharing

The challenges that emerged from the interview data can be grouped as follows: licences and legal issues; technological barriers and data formats; and cultural and social challenges.

4.4.1 Licences and legal issues

Licences and legal issues can be an obstacle in publishing corpora or terminological resources as already discussed by ELRC (2019, 2022). The interview data showed that licences and legal issues concerned corpus data as well as terminological data.

Some interview participants mentioned that they did not publish certain corpora because of copyright issues:

> The corpora you mean? We publish … Obviously it depends on copyright. The [name of corpus] I don't think we published this because […] it would have taken too long to get all the necessary permissions. We used a lot of web material which we harvested and we did not bother to follow up with the copyright holders to get permission to publish. (INT 4)

Sometimes the data is published but it is only consultable online and it is not for download because of copyright restrictions: "Some of the material is to download, some of them are there only to search, again depending on the copyright restrictions" (INT 4).

Legal issues can also affect sharing and reuse scenarios of terminological data sets: one reuse scenario might be covered by a specific licence, while others are not. One interview participant gave the example that the terminological data could not be reused for an NLP (natural language processing) task in machine translation, due to licence restrictions:

> The terminological data is more restricted and it is on my to-do list to address that question because I examined a PhD candidate recently [who was] working on machine translation […]. [O]ne of the issues was that one of the data sets not being usable in creating open datasets, so dataset for NLP tasks. So, in NLP you want the data that the task is based on available so that other people can rerun the task and change the parameters or something. So, the licence [for the terminological data] is not fully open. (INT 2)

4.4.2 Technological barriers and data formats

Compatibility issues between older and newer formats or systems can hinder efficient data sharing. Furthermore, formats that are not based on established standards in the field, or proprietary formats (Declerck/Wissik 2020), can be an obstacle to data exchange as well:

> In fact, this service is quite old, at the beginning of open data and all of this, we started doing this. And this format [custom XML][5] was established [a long time ago]. We know that sometimes it is not easy to incorporate it [the XML data] somewhere else, and it needs dealing with it and needs someone who knows internal programming, but it is a way to offer data. (INT 8)

Data conversion or mapping (Wissik 2024) between a specific tool format and a standard format such as TBX is often needed.

4.4.3 Social and cultural challenges

Language data such as terminological data, corpora or translation memories are a valuable asset on their own and individual researchers, the research community and other stakeholders in the public and private sector (e. g., freelance translators, translation departments at universities, language service providers, language technology developers) can benefit from data sharing practices. These benefits include but are not limited to higher quality of language data, more efficient use of resources, and a more competitive development of language technologies like language-centric AI (ELRC 2022). Despite these potential advantages, data sharing is not particularly widespread in the research community (Rakers et al. 2023) or in the public sector (ELRC 2022).

Regarding the terminology community specifically, not all the stakeholders in the terminology workflow are aware or fully convinced of the benefits of FAIR and sustainable data sharing and open data, and it takes time to make these changes: as Rakers et al. (2023) stated, “processes of cultural change are needed to reap the benefits of data sharing” and sustainable language data management. The interview data showed

[5] It is an XML format, but it is not compatible with standard formats like TBX or TEI.

that not all stakeholders in the terminology workflow were ready to share their data as open data:

> In case of the terminology […] the licence for the terminology bit [is] slightly restricted in that you can use it for personal use but you can't share it or reuse it. And the only reason for […] that at the time we were making it available the [Terminology Committee was] not fully ready for the idea of making their data freely available […] and fair enough, these things take time. (INT 2)

Some people involved in terminology work still need to learn and understand that publishing language data as open data or with open-access licences is especially beneficial for less-resourced languages:

> [T]hey [the glossary editors] have to understand why is it good for us for the [less-resourced language] to have open access […] the old editors, those who started 25 years ago, they don't understand the benefits of open access. (INT 14)

Other social and cultural factors were linked to the fact that terminology professionals primarily view human users as their target group and not machines (see Section 4.1.2), and that they underestimate the value of the compiled corpora for external users (see Section 4.3).

5 Discussion

In this section we will discuss the results of the study in light of the FAIR data principles. First, we will discuss the sharing of terminological data followed by the sharing of corpus data. Regarding findability, most terminological resources have not been assigned a global unique and persistent identifier, they only have identifiers at the local level as required by ISO 30042:2019 (see also Vezzani/Di Nunzio/Costa 2023), unless they are registered or indexed in a searchable source, such as a repository (e. g., the repositories provided in the CLARIN infrastructure), which automatically assigns a unique and persistent identifier to the deposited resource. However, as the interview results show, very few terminology professionals deposit their terminological resources in repositories. Another aspect of findability is that the data should be described with rich metadata, which is the case for most terminological resources that comply with ISO standards related to terminology. Furthermore, the metadata should clearly and explicitly include the identifier of the data they describe. However, if a resource does not have a persistent identifier, the corresponding metadata cannot include such an identifier.

The fact that terminological resources are downloadable does not automatically mean that they are accessible in terms of the FAIR data principles. Most downloadable terminological data do not comply with the FAIR data principles because they do not have persistent identifiers assigned, they are not accessible via standardized communication protocols and they do not have clear licences attributed to them and sometimes the data are not provided in a standard format such as TBX.

Several interview participants mentioned that they have an API in place. Terminological data that is provided via APIs is more likely to be FAIR because APIs are highly aligned with FAIR principles, offering strong support for findability, accessibility, interoperability and reusability. One of the requirements for being accessible is that the data is retrievable by a standardized communication protocol. APIs inherently use standard protocols like HTTP, fulfilling this requirement (FAIR Process Framework n. d.) They also support a range of standard data formats (FAIR Process Framework n. d.), enabling, for example, the access to terminological data directly within computer-assisted translation systems. Furthermore, the protocols used by APIs are usually as required by the FAIR data principles "open, free, and universally implementable" and "allow for an authentication and authorisation procedure, where necessary" (Wilkinson et al. 2016).

Regarding the interoperability aspect of FAIR, which focuses on using standard formats and vocabularies to enable data integration across different systems, most terminological resources follow relevant ISO standards even if not all of them were mentioned in the interviews explicitly. The ISO 16642:2025 *Management of terminology resources — Terminological markup framework*[6] provides the structural meta-model to ensure interoperability between different terminological data collections (Vezzani/Di Nunzio/Costa 2023), while in ISO 12620-1:2022, the data categories are specified. The data categories are also stored in the DatCatInfo, a Data Category Repository (DCR), assigning PIDs for each data category. Another ISO standard, ISO 30042:2019 Terminology eXchange Format, which is also crucial for interoperability and reusability (Vezzani/Di Nunzio/Costa 2023), was mentioned in the interviews, and most terminological data was provided in TBX. However, some interview participants also mentioned challenges related to data formats that can be an obstacle in sharing terminological data. Regarding the reuse aspect of the FAIR data principles, the terminological data should be provided with clear licences and metadata that allow others to use it effectively. The provision of rich metadata is not the problem, but the specification of licences can pose challenges, as the interview data showed.

With regard to sharing terminological data as Linked Open Data, among the interview participants, Linked Open Data did not play an important role. However, the adoption of Linked Open Data could be one possible step to implement the FAIR data principles because Linked Open Data can provide the structure for the terminological data to become FAIR (e. g., by assigning Uniform Resources Identifiers (URIs)) and the terminological data already provides the rich metadata, which is lacking in the LOD framework. Only the question of licensing needs to be solved by the terminology professionals. However, when transforming TBX data into LOD, a mapping of data categories and vocabularies is needed (see Vezzani/Di Nunzio/Costa 2023).

The study has also shown that the results of ad hoc terminology work, like the answers to terminology questions from users, are shared with the general public.

[6] With its previous version ISO 16642:2017 still in in place during the interview period.

However, this data is often shared as unstructured data on a website or in spreadsheets because its structure resembles more of a question-and-answer format and requires more context, making it difficult to include in the rigid data models of terminology databases. Often these resources are not even seen as "data," and therefore they are rarely shared according to the FAIR data principles.

Regarding sharing corpus data, three scenarios were identified in the interview data: in the first scenario, specialised corpora are published and openly shared, copyright permitting; in the second scenario, specialised corpora are shared internally or with a wider group through a corpus management system, but they are not published and openly shared outside of the corpus management system; and in the third scenario, specialised corpora are only used internally because they are considered an intermediate step in creating terminological resources. Thus, only in the first scenario the data need to adhere to the FAIR data principles because in the other two scenarios corpora are not shared publicly. It would be beneficial to raise awareness in the terminology community that corpora collected for the purpose of creating terminological resources may also constitute valuable language resources for other users and application scenarios when shared in accordance with the FAIR data principles.

From the interview data, several challenges related to FAIR data sharing emerged: licences and legal issues, technological barriers and data formats, and cultural and social factors. Licences and legal issues were particularly problematic for sharing corpus data. Furthermore, licences attributed to terminological data do not cover all reuse scenarios, e. g., in a specific case mentioned, reuse for an NLP task was not covered by the specific licence attributed. Another challenge mentioned involved compatibility issues between legacy and modern formats or systems, which affects interoperability of the data and often requires time-consuming data conversions or data mappings. The terminology community needs to be more open towards open data formats and/or standardised data formats, such as TBX, because not all resources were made available in standardised formats which makes it difficult to share data. Using domain-relevant data and metadata standards increases the reusability of resources and is also one of the reuse requirements for FAIR data sharing practices.

Another set of challenges concerned cultural and social factors. The interview data showed that not all participants were ready to share their data as open data, and not all of them applied FAIR data principles when sharing their data. All data created by participants were web accessible. However, not all data were openly available via download or APIs. Sometimes more restricted licences were used. In some cases, no licences were attributed, which, strictly speaking, means that downloaded data cannot be reused. As explained in the FAIR data principles (Wilkinson et al. 2016), the clarity of licensing status is essential for the reusability of data. As discussed by Rakers et al. (2023) regarding the research community: "A transition to FAIR and open data sharing can only be sustainable if it permeates all aspects of [a community] […] Therefore, this transition takes time and must start as early as possible." This is also applicable to the

terminology community, and there are different stakeholders that can support this transition, such as the institutions themselves, different associations and networks, research infrastructures and individual projects. Stakeholders can also function as "platform[s] for discussion and collaboration around cultural change" (Rakers et al. 2023). Another issue, closely connected to social and cultural factors, is that terminology professionals still primarily view human users as the main target group for their terminological resources. This perspective was reflected in the data, for example, by the fact that APIs are not implemented for terminological data, even if this technology is used for other data sets within the same institution. Greater awareness is therefore needed that terminology data should be provided not only for human users but also for machines. To facilitate this shift, the adoption of FAIR data principles is essential.

Another important aspect of strengthening FAIR data sharing practices, as outlined by Amelung et al. (2023), is to provide training to build the required skills and raise awareness about the benefits of FAIR data sharing practices. All this will bring "cultural change in the handling of [language] data" (Amelung et al. 2023) in the context of terminology work in institutional settings.

While insightful, it is difficult to generalise based on the results presented here due to limitations of the study mentioned in the method section. For instance, only institutions that maintained publicly accessible termbases or dictionary databases or portals were included in the sample. Therefore, terminology data sharing practices by institutions that do not maintain a public terminology database fall outside of the scope of this study. Furthermore, the small sample size of only 15 expert interviews and the focus on institutions in Europe, apart from 2 international organisations, make it difficult to draw conclusions about other types of institutions. However, since the 15 expert interviews covered the most important current approaches to terminology work and the most common scenarios where terminology work is performed in institutional settings, the study contributes to a deeper understanding of current FAIR data sharing practices in terminology work in institutional settings in Europe.

6 Concluding remarks

The available literature on terminology practices contains limited discussion on sustainable and FAIR data sharing practices in terminology work in institutional settings, which motivated the present study. This study examined data sharing practices, the application of FAIR data principles in terminology work in institutional settings and the related challenges based on qualitative data gathered through 15 expert interviews among stakeholders performing terminology work in institutional settings and the analysis of publicly available documentation.

The research has shown that there are different data sharing methods within the community. The study confirmed that APIs are becoming increasingly popular; on the other hand, data repositories are not widely used as a channel for data sharing in the

field of terminology work in institutional settings. The results of this study indicate a change in how terminology data is accessed and shared: it is not only publicly consultable in terminology databases and dictionaries, as it was in the past, but data can also be downloaded, accessed via APIs, via computer-assisted translation environments, and processed further. However, the study also showed that not all interview participants consistently applied FAIR data principles when sharing their data, for example through the absence of persistent identifiers or clear licences. In addition, the study reveals that Linked Open Data formats were not widely used among the interview participants. The study also identified a range of challenges to FAIR data sharing in terminology work in institutional settings, including legal and technical issues, data format constraints, and social and cultural factors such as the lack of awareness of the benefits of open licences and the limited recognition of machines as users of terminological resources, which largely align with the circumstances described by ELRC (2022: 21). To overcome those challenges, especially the social and cultural factors, it is important to raise awareness about the benefits of FAIR data sharing practices among the terminology community and to provide specific training activities to build the required skills for FAIR data sharing practices and to support a cultural change in handling of language data. Furthermore, it would be beneficial to adapt existing terminology workflow models to the requirements of open science and FAIR principles and integrate the depositing of data in a repository as an independent step into the terminology workflow models.

Acknowledgement

The author would like to thank all the interviewees for their participation. Without their support this study would not have been possible. The author would also like to thank the anonymous reviewers for their feedback to improve this article.

AI declaration

The author has used a generative AI tool (ChatGPT-5.2) for linguistic revision to check and improve the grammar, language, style. The author has then reviewed and edited the article as needed and takes full responsibility for the content.

References

[ALLEA] Natalie Harrower, Maciej Maryl, Beat Immenhauser, Timea Biro (eds) (2020): *Sustainable and FAIR data sharing in the humanities: Recommendations of the ALLEA working group E-Humanities*. Berlin: ALLEA – https://doi.org/10.7486/DRI.tq582c863

Amelung, Lisa et al. (2023): “NFDI white paper: Interim report reference.” Zenodo – https://doi.org/10.5281/zenodo.10101412

Andersen, Gisle; Peder Gammeltoft (2022): “The role of CLARIN in advancing terminology: The case of Termportalen – the national terminology portal for Norway.” Darja Fišer, Andreas Witt (eds): *CLARIN: The infrastructure for language resources*. Berlin/Boston: De Gruyter, 249–274

Arndt, Tamara; Silvia Cerrella Bauer, Mark Childress, Klaus Fleischmann, Beate Früh, Regine Hellerich, Rachel Herwartz, Ana Hoffmeister, François Massion, Peter Oehming, Uta Schmitt, Ayten Turhan, Ulrich Wachowius, Annette Weiland, Cecil A. Woas, Wolfgang Zenk (2020): “Module 5. Project and process management.” *Terminology work. Best practices 2.0*. Köln: Deutscher Terminologie-Tag, 1–24

Bowker, Lynne (1996): “A corpus-based approach to terminography.” *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 3 [2]: 27–52

Caffrey, Colm; Cristina Valentini (2019): “Applications of technology in the Patent Cooperation Treaty (PCT) translation division of the World Intellectual Property Organization (WIPO).” Minako O'Hagan (ed.): *Routledge handbook of translation and technology*. 1st edn. London/ New York: Routledge, 127–147

Cerrella Bauer, Silvia (2015): “Managing terminology projects.” Hendrik J. Kockaert, Frieda Steurs (eds): *Handbook of terminology*. Vol. 1. 2nd ed. Amsterdam: Benjamins, 324–340

trans-kom **ISSN 1867-4844**

trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

trans-kom wird ausschließlich im Internet publiziert: https://www.trans-kom.eu

Chiocchetti, Elena; Natascia Ralli (2012): Deliverable D3.2 report workflow adaptation for LISE – https://cordis.europa.eu/docs/projects/cnect/7/270917/080/deliverables/001-LISEdeliverable32.pdf

Chiocchetti, Elena; Vesna Lušicky, Tanja Wissik (2023): "Multilingual legal terminology databases. workflows and roles." Łucja Biel, Hendrik J. Kockaert (eds): *Handbook of terminology.* Vol. 3: *Legal terminology.* Amsterdam/Philadelphia: Benjamins, 458–484

Cimiano, Philipp; Christian Chiarcos, John P. McCrae, Jorge Gracia (2020): *Linguistic linked data: Representation, generation and applications*. Berlin: Springer

Cooper, Doug (2010): "When nice people don't share: 'Shy' data, web APIs, and beyond." Position presentation for panel discussion by the same title, organized during the 2nd International Conference on Global Interoperability for Language Resources (ICGL) at the City University of Hong Kong in January 2010 – http://sealang.net/archives/ICGL2010.pdf

De Jong, Franciska; Dieter Van Uytvanck, Francesca Frontini, Antal van den Bosch, Darja Fišer, Andreas Witt (2022): "Language matters." Darja Fišer, Andreas Witt (eds): *CLARIN: The infrastructure for language resources.* Berlin/Boston: De Gruyter, 31–58 – https://doi.org/10.1515/9783110767377-002

Declerck, Thierry; Tanja Wissik (2020): "Automated formalisation of unstructured terminologies." *Terminologie & Ontologie: Théories et Applications. Actes de la conférence TOTh 2020*. Université Savoie Mont Blanc, 26th & 27th November 2020. Chambéry Cedex: Presses Universitaires Savoie Mont-Blanc, 217–232

Di Buono, Maria Pia; Philipp Cimiano, Mohammad Fazleh Elahi, Frank Grimm (2020): "Terme-a-LLOD: Simplifying the conversion and hosting of terminological resources as linked data." Maxim Ionov, John P. McCrae, Christian Chiarcos, Thierry Declerck, Julia Bosque-Gil, Jorge Gracia (eds): *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*. Paris: European Language Resources Association, 28–35

Di Nunzio, Giorgio Maria; Geneviève Henrot, Maria Teresa Musacchio, Federica Vezzani (2022): "Preface of 1st International Conference on Multilingual Digital Terminology Today. Design, representation formats and management systems." *Proceedings of the 1st International Conference on Multilingual Digital Terminology Today.* (CEUR Workshop Proceedings 3161.) Padua: CEUR – https://ceur-ws.org/Vol-3161/#preface

Di Nunzio, Giorgio Maria; Eszter Papp, Federica Vezzani, Ellie Kemp (2024): "FAIR terminology meets CLEAR global." Apostolos Antonacopoulos, Annika Hinze, Benjamin Piwowarski, Mickaël Coustaty, Giorgio Maria Di Nunzio, Francesco Gelati, Nicolas Vanderschantz (eds): *Linking theory and practice of digital libraries. TPDL 2024.* (Lecture Notes in Computer Science 15178.) Cham: Springer – https://doi.org/10.1007/978-3-031-72440-4_17

Dorer, Brita; Anna Kuźnik, David Orrego Carmona, Cornelia Zwischenberger (2025): "Surveys and interviews." Anna Maria Rojo López, Ricardo Muñoz (eds): *Research methods in cognitive translation and interpreting studies.* Amsterdam: Benjamins, 69–91

Drewer, Petra; Klaus-Dirk Schmitz (2017): *Terminologiemanagement. Grundlagen – Methoden – Werkzeuge.* Berlin: Springer

Elbaz, Pascale; Elpida Loupaki, Giorgio Maria Di Nunzio, Federica Vezzani, Rodolfo Maslias (2022): "YourTerm: A collaborative & FAIR European project." *Translating and the computer (TC44).* Luxembourg: ASLING

ELRC (2019): *ELRC white paper – Sustainable language data sharing to support language equality in multilingual Europe: Why language data matters.* Saarbrücken: European Language Resource Coordination Consortium

ELRC (2022): *ELRC white paper – AI for a multilingual Europe: Why language data matters.* Saarbrücken: European Language Resource Coordination Consortium

FAIR Process Framework (n. d.): Factsheet – Step 6 application programming interface (APIs) – https://www.fairprocessframework.org/wp-content/uploads/2024/12/6.2-Application-Programming-Interfaces-APIs-FAIR-Process-Frameworkpdf.pdf

Fran (n. d.): Fran: Slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU – https://www.fran.si/

Gius, Evelyn; Jan Christoph Meister, Malte Meister, Marco Petris, Mareike Schumacher, Dominik Gerstorfer (2023): "CATMA 7 (Version 7.0)." Zenodo – https:/doi.org/10.5281/zenodo.1470118

Guest, Greg; Arwen Bunce, Laura Johnson (2006): "How many interviews are enough? An experiment with data saturation and variability." *Field Methods* 18 [1]: 59–82

ISO 16642:2025 (2025): *Management of terminology resources — Terminological markup framework*

ISO 12620-1:2022 (2022): *Management of terminology resources — Data categories.* Part 1: *Specifications*

ISO 30042:2019 (2019): *Management of terminology resources — TermBase eXchange (TBX)*

Klosa-Kückelhaus, Annette; Carole Tiberius (2025): "The lexicographic process revisited." *International Journal of Lexicography* 38 [1]: 1–12 – https://doi.org/10.1093/ijl/ecae016

Kosem, Iztok; Sanni Nimb, Carole Tiberius, Bob Boelhouwer, Simon Krek (2021): "License to use: ELEXIS survey on licensing lexicographic data and software." Zoe Gavriilidou, Lydia Mitits, Spyros Kiosses (eds): *Proceedings of the 19th EURALEX International Congress*. 7–9 September 2021, Alexandroupolis. Vol. 2. Komotini: SynMorPhoSe Lab, Democritus University of Thrace, 705–712 – https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202020-2021/EURALEX2020-2021_Vol2-p705-712.pdf

Kuckartz, Udo (2014): *Qualitative text analysis. A guide to methods, practice and using software.* London: Sage

Kuckartz, Udo (2019): "Qualitative text analysis: A systematic approach." Gabriele Kaiser, Norma Presmeg (eds): *Compendium for early career researchers in mathematics education*. Springer, 181–197

Kudashev, Igor (2013): *Quality assurance in terminology management: Recommendations from the Termfactory Project*. Helsinki: Unigrafia

Lušicky, Vesna; Tanja Wissik (2019): "Language resources in repositories and catalogues: Pilot study on their potential for LSP teaching and training." Ingrid Simonnæs, Øivin Andersen, Klaus Schubert (eds): *New challenges for research on language for special purposes. Selected proceedings from the 21st LSP Conference, 28–30 June 2017 Bergen, Norway*. Berlin: Frank & Timme, 327–342

Maroto, Nava (2024): "TeresIA. Spanish access portal to terminologies and artificial intelligence services." *Proceedings of the 3rd International Conference on Multilingual Digital Terminology Today (MDTT 2024)*. Granada, Spain, June 27–28, 2024

Martín-Chozas, Patricia; Thierry Declerck, Elena Montiel-Ponsoda, Victor Rodríguez-Doncel (2024): "Representing terminological data in the Semantic Web: A proposal based on OntoLex-lemon." *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 30 [2]: 285–321

Meuser, Michael; Ulrike Nagel (1991): "Experteninterviews – vielfach erprobt, wenig bedacht." Detlef Gerz, Klaus Karaimer (eds): *Qualitative-empirische Sozialforschung. Konzepte, Methoden, Analysen*. Opladen: Westdeutscher Verlag, 441–471

Rakers, Julia; Bernhard Miller, Julia Mohrbacher, Daniel Nüst, Torsten Schrade, Jörg Seegert, Christian Vater, Cord Wiljes, Holger Simon (2023): "Because data shall grow (and we with it)." *1st Conference on Research Data Infrastructure – Connecting Communities (CoRDI)*. Karlsruhe: Zenodo – https://doi.org/10.5281/zenodo.8333776

Rehm, Georg (2016): "The language resource life cycle: Towards a generic model for creating, maintaining, using and distributing language resources." *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16).* Portorož: European Language Resources Association (ELRA), 2450–2454

Roche, Christophe; Amparo Alcina, Rute Costa (2019): "Terminological resources in the digital age." *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 25 [2]: 139–145 – https://doi.org/10.1075/term.00033.roc

Roche, Christophe; Rute Costa, Sara Carvalho, Bruno Almeida (2019): "Knowledge-based terminological e dictionaries.The EndoTerm and al-Andalus Pottery projects." *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 25 [2]: 259–290

Schumacher, Mareike (2019): "CATMA." *forTEXT. Literatur digital erforschen* – https://fortext.net/tools/tools/catma

Smal, Lilli; Andrea Lösch, Josef van Genabith, Maria Giagkou, Thierry Declerck, Stephan Busemann (2020): "Language data sharing in European public services – Overcoming obstacles and creating sustainable data sharing infrastructures." *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille: European Language Resources Association, 3443–3448

Tiberius, Carole; Jelena Kallas, Svetal Koeva, Margit Langemets, Iztok Kosem (2024): "A lexicographic practice map of Europe." *International Journal of Lexicography* 37 [1]: 1–28 – https://doi.org/10.1093/ijl/ecad023

UNESCO (2021): *UNESCO recommendation on open science*. Paris: United Nations Educational, Scientific and Cultural Organization

Vezzani, Federica (2021): "La ressource FAIRterm: entre pratique pédagogique et professionnalisation en traduction spécialisée." *Synergies Italie* 17: 51–64

Vezzani, Federica (2022): *Terminologie numérique: conception, représentation et gestion*. Lausanne: Peter Lang

Vezzani, Federica; Giorgio Maria Di Nunzio (2023): "Multilingual digital terminology: Introduction to the special issue." *Digital Scholarship in the Humanities* 38 [1]: i1–i5 – https://doi.org/10.1093/llc/fqad028

Vezzani, Federica; Giorgio Maria Di Nunzio, Rute Costa (2023): "ISO standards for terminology resources management. Are they FAIR enough?" *Digital Translation* 10 [2]: 233–252

Wilkinson, Mark D. et al. (2016): "The FAIR guiding principles for scientific data management and stewardship." *Scientific Data* 3: 160018 – https://doi.org/10.1038/sdata.2016.18

WIPO (2025): WIPO Pearl Terminology API – https://wipopearl.wipo.int/en/api

Wissik, Tanja (2023): "Guideline for questions used for expert interviews in 'Study on terminology workflows in the 21st century'." Zenodo – https://doi.org/10.5281/zenodo.11144968

Wissik, Tanja (2024): "Dimensions of sustainability in terminology practices in institutional settings." *Terminology Science & Research* 27: 93–116

Wissik, Tanja (2025a): "Impact of automatic term extraction on terminology work: A qualitative interview study in institutional settings." *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 31 [1]: 110–135

Wissik, Tanja (2025b): "An infrastructural approach to terminology work: The case of research infrastructures." *Selected papers of the CLARIN Annual Conference 2024*. Linköping: Linköping University Electronic press, 1–12 – https://doi.org/10.3384/ecp216.01

Wissik, Tanja; Thierry Declerck (2019): “Using an infrastructure for lexicography in the field of terminology.” *Terminologie & Ontologie: Théories et Applications. Actes de la conference TOTh 2019.* Chambéry: Presses Universitaires Savoie Mont Blanc, 365–379

Wissik, Tanja; Vesna Lušicky, Elena Chiocchetti (2024): “Terminologiearbeit für das Gerichtsdolmetschen: Eine Umfrage unter österreichischen Gerichtsdolmetscher*innen.” *Fachsprache* 46 [1–2]: 20–39 – https://doi.org/10.24989/fs.v46i1-2.2247

Žagar Karer, Mocja; Tanja Fajfar (2023): “Terminological problems of terminology users: Analysis of questions in terminological counselling service on the Terminologišče website.” *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 29 [2]: 78–102

Zorrilla-Agut, Paula; Thierry Fontenelle (2019): “IATE 2. Modernising the EU’s IATE terminological database to respond to the challenges of today’s translation world and beyond.” *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication* 25 [2]: 146–174

Author

Tanja Wissik is a senior scientist at the Austrian Centre for Digital Humanities at the Austrian Academy of Sciences. Her research focuses on language resources, workflow research, infrastructures, and the didactics of digital methods. She has extensive experience teaching terminology, language technology, and digital methods in the humanities. She has led several research projects focused on the creation of language resources and serves as National Coordinator for Austria in the European research infrastructure CLARIN.
E-mail: Tanja.wissik@oeaw.ac.at

Empfehlungen

Transkulturalität – Translation – Transfer

Herausgegeben von Martina Behr, Julia Richter und Larisa Schippe

Tatsiana Haiden: **Translation and soft power – Translation culture of the Paul Zsolnay publishing house 1924–1938.** ISBN 978-3-7329-1061-8

Easy – Plain – Accessible

Herausgegeben von Silvia Hansen-Schirra und Chris Maaß

Silvia Hansen-Schirra / Chris Maaß (Hg.): **Text- und nutzerseitige Studien zu Leichter Sprache.** ISBN 978-3-7329-0701-4

Regina Stodden: **Automatic German Text Simplification: Data, Evaluation, and Models.** ISBN 978-3-7329-1216-2

Chris Maaß: **Leichte Sprache, Einfache Sprache, Leichte Sprache Plus. Aktuelle Entwicklungen von DIN bis KI.** ISBN 978-3-7329-1235-3

POLSKI.BEITRÄGE zur Geschichte und Gegenwart der polnischen Sprache im deutschsprachigen Raum

Herausgegeben vom Kompetenz- und Koordinationszentrum Polnisch (KoKoPol)

Anna Żurek: **Kommunikationsstrategien zweisprachiger Personen. Am Beispiel der Herkunftssprache Polnisch in Deutschland.** ISBN 978-3-7329-1097-7

vis-à-vis – Semiotik transdisziplinär

Herausgegeben von Thomas Sähn, Marie Schröer und Christian Sinn

Thomas Sähn / Marie Schröer / Christian Sinn (Hg.): **Semiotik als Handwerkszeug: perspectives pluridisciplinaires.** ISBN 978-3-7329-1165-3

Studien zu Fach, Sprache und Kultur

Herausgegeben von Ines-Andrea Busch-Lauer

Ines-Andrea Busch-Lauer (Hg.): **Wenn KI auf Fach und Sprache trifft.** ISBN 978-3-7329-1225-4

Forum für Fachsprachen-Forschung

Herausgegeben von Hartwig Kalverkämper

Eriberto Russo: **Nachhaltigkeit in der Landwirtschaft. Eine fachsprachliche Untersuchung.** ISBN 978-3-7329-1196-7

Claudio Di Meola / Joachim Gerdes / Livia Tonelli (Hg.): **Sprachvariation im Deutschen zwischen Theorie und Praxis. Fachsprachlichkeit, Inklusion, Didaktik, Übersetzung, Kontrastivität.** ISBN 978-3-7329-1042-7

Leonora Wollenhaupt: **Conceptual foundations of legal language. An analytical study of the terms nationality and ethnicity.** ISBN 978-3-7329-1151-6

Dorina Irimia: **Les verbes de la condamnation pénale. Initiation à la jurilinguistique appliquée au discours judiciaire. Avec une préface de Richard Ryan.** ISBN 978-3-7329-1203-2

Weidler bei Frank & Timme

Berliner Sprachwissenschaftliche Studien

Herausgegeben von Claudia Wich-Reif

Jessica Marwig: **Sprachliche Strategien zur Herstellung und Darstellung der Unternehmensidentität.** ISBN 978-3-89693-804-6

Europäische Beiträge zur Angewandten Germanistik

Herausgegeben von Jörg Meier

Roberta Rada / Katharina Zipser (Hg.): **Zum medialen Corona-Diskurs. Linguistische Annäherungen im interkulturellen Kontext mit Schwerpunkt Österreich – Ungarn.** ISBN 978-3-89693-805-3

Federico Collaoni: **Die deutsche Energiewende – Terminologie und Diskurs. Eine linguistische Untersuchung von den Ursprüngen bis zum Atomausstieg 2011.** ISBN 978-3-89693-791-9

Alle Bücher sind auch als E-Books erhältlich.

Frank & Timme

Wittelsbacherstraße 27a, 10707 Berlin • Telefon +49 (0)30 36 42 15 700 • info@frank-timme.de • www.frank-timme.de