

DATA MANAGEMENT PLAN FOR DH TRAINING

ADMINISTRATIVE INFORMATION

Project name: Writing and Reading Processes in Technical Software Documentation

Principal investigator/researcher: Justyna Dlociok

Supervisor and co-supervisor: Birgitta Meex and Bert Oben

Institution: KU Leuven

Project summary:

When producing content, a technical writer adapts its tone, style, complexity, and structure to the intended audience. However, the writing process nowadays involves other stakeholders and contributors, such as subject matter experts, technical illustrators, and companies with specific branding strategies or budgetary restrictions. Each of those stakeholders might have different expectations towards the produced information. The user wants it to be readable, findable, and understandable. For the writer, it needs to be consistent and reusable. The subject matter expert might foreground correctness, sometimes at the expense of style. If the company strives for a high-quality machine translation, the source text must be unambiguous and skillfully simplified. Fulfilling all of those criteria is a challenging task. However, as numerous studies show, applying specific language rules has been considered a possible remedy.

For instance, Byrne (2004) analyzed the impact of iconic linkage and textual cognetics on the usability of software documentation. He discovered that using parallel structures significantly improves the readability and comprehensibility of the user manuals. O'Brien (2003), and Shubert et al. (1995). drew similar conclusions. The former showed that controlled languages facilitate readability in complex texts; the latter proved that simplified English enhances the comprehensibility of complex instructions. In the context of machine translation, Roturier (2006) investigated the positive impact of controlled English on the comprehensibility, usefulness, and acceptability of machine-translated technical documentation for French and German users.

Similarly, Steensland & Dervisevic (2005) developed a set of rules that improved the source text and the quality of machine translation when fed into a language checker. Last but not least, Suchowolec (in Goebel, 2014) reported a case of Koenig & Bauer AG, a printing company that decided to improve the authoring process. As a result of focusing on consistency, reusability, and content structuring, a style guide emerged at the company that could potentially improve the end user's content. Still, numerous companies struggle to apply working rules and standards to their content. They often rely on intuition alone as a good (writing) practice. Even those companies that recognize the benefits of trained technical writers often fail to produce consistent and audience-specific documentation.

In this project, we aim to analyze both the product and the process of rule-based technical writing. Concerning the process, we aim to answer the following questions: How do technical writers write from sources? How do they reuse the content? How do they optimize and revise the content in developing text? Do they use style guides, standards, editing FAQs? Which of the rules and standards are most challenging to apply? Which of the rules and standards are considered essential and not?

By analyzing the product in the companies, we want to know: What are the differences between theoretically valid and practically applied intra- and extra-textual rules, guidelines, and standards?

SECTION 1. DATA DESCRIPTION

What data will you collect or create? Fill out the table below and/or describe it.

Type of data	Format	Volume	Storage location	How created?
Primary				
Expert Interviews: Audio, transcription, consent form	.mp4	80 MB per interview (ca. 30)	<ul style="list-style-type: none"> Personal laptop Google Drive (Personal Account) One Drive (Accessible through KU Leuven) (Personal) External Hard Disk 	Interviews were/are being recorded with Skype and the in-built mobile phone Recorder.
	.docx	30 KB per transcription (ca. 30)		Transcriptions were/are being transcribed manually into a .docx document.
	.pdf	130 KB per consent form (ca. 30)		Consent Forms were/are created in a .docx and converted into a .pdf.
Writing Screen-Capturing	.camrec .wmv	280 MB per session (ca. 9)	<ul style="list-style-type: none"> Personal laptop Google Drive (Personal Account) One Drive (Accessible through KU Leuven) (Personal) External Hard Disk 	Writing sessions will be recorded with Camtasia software in .camrec and later rendered as .wmv.
Think-aloud Protocol	.mp4	80 MB per session (ca. 9)	<ul style="list-style-type: none"> Personal laptop Google Drive 	Think-aloud protocols were/are being recorded with

			(Personal Account) <ul style="list-style-type: none"> • One Drive (Accessible through KU Leuven) • (Personal) External Hard Disk 	Skype and the in-built mobile phone Recorder.
Usability Study with Eye-Tracking	.mp4 .tsv .xlsx	210 MB per session (ca. 80)	<ul style="list-style-type: none"> • Personal laptop • Google Drive (Personal Account) • One Drive (Accessible through KU Leuven) • (Personal) External Hard Disk 	Usability sessions will be recorded with Tobii X2-60 eye-tracker.
Bibliographical references	.pdf	1 MB	<ul style="list-style-type: none"> • Personal laptop • Google Drive (Personal Account) • One Drive (Accessible through KU Leuven) • (Personal) External Hard Disk 	The accumulated literature library is managed in Mendeley, resulting in (automatic) bibliographical references.
*Notes and Annotations	.xlsx .pdf	-	<ul style="list-style-type: none"> • Personal laptop • Google Drive 	Notes and annotations will be stored as comments in the files/programs

			(Personal Account) <ul style="list-style-type: none"> • One Drive (Accessible through KU Leuven) • (Personal) External Hard Disk 	mentioned above.
--	--	--	--	------------------

Do you intend to reuse existing data?

No, I do not intend to reuse existing data.

Do you use personal data (i.e., all data possibly identifying an individual)?

Yes.

SECTION 2. METADATA AND DOCUMENTATION

Describe the documentation that will be created for the data. This section deals with the way in which you will document how the dataset was created and subsequently processed.

For each dataset – interviews (expert, think-aloud), recordings (writing sessions, usability studies) – documentation of how the data was collected will be stored as a README file. It will involve background information about the individuals (participants, companies) and locations.

Describe the metadata for the data. This section deals with metadata: information contained in your dataset that adds structure to the research data.

The metadata for this project will involve:

- *Data about the participants:*
 - *Individual: age, sex, educational and professional background,*
 - *Companies: Industry,*
- *Data about the recording sessions: how, with whom.*

SECTION 3. ETHICAL, LEGAL, AND PRIVACY ISSUES

Are there any ethical issues concerning the creation and/or use of the data?

The personal data collected in the context of this project will be pseudonymized. Each participant will be given a participant number during the data collection. All data collected are linked to this number and not to the person. The researcher has a key that keeps track of the participant number for each person. If

participants request the erasure of their data, the researcher can use the key to retrieve the data. The key file will be kept on OneDrive in a password-protected folder.

Did you consider all issues about copyrights and IPR?

Yes.

Are the collected data considered to be "data containing personal information," and are all the requirements about the collection of these data met?

Yes. For that, a Privacy and Ethical application has been submitted and approved.

SECTION 4. DATA STORAGE AND BACKUP DURING THE RESEARCH

How and where will the data be stored during research?

The data will be stored on KU Leuven OneDrive for Business, with an enabled Multi-Factor Authentication.

Which backup procedures are in place?

The backups of the laptop are made automatically on Google Drive and One Drive. The external hard disk will be updated once a week. These will be encrypted using Bitlocker. The bibliographical references are stored in Mendeley, which also has an online backup that one can log onto through an account.

Describe the data security procedures and who has access to the data.

Only the researchers and the supervisors of the above-described project will access the data. A log-in is necessary to access the gathered data on Google Drive and One Drive. The researcher will keep the laptop and the external hard disk in a safe place to protect it against theft and will be very attentive to these materials to rule out the loss.

SECTION 5. DATA SELECTION AND PRESERVATION AFTER THE RESEARCH

What is the long-term preservation plan for these dataset(s)?

The data will be preserved in a research data repository of KU Leuven (KU Leuven's Research Data Repository), which is a safe, secure, and sustainable environment. After the end of the research, relevant data will be stored there.

Which data will have long time value for the research and will be preserved?

These data include:

- *Expert interview guide,*
- *Expert interview transcriptions,*
- *Style Guide Survey,*
- *Consent form,*
- *The writing sessions,*

- *Think-aloud protocols,*
- *Usability study data,*
- *Bibliographical references.*

SECTION 6. DATA SHARING

Are there any restrictions for sharing the data?

Yes, the recordings and interviews might contain sensitive information.

If there are no restrictions, which mechanisms will be in place to assure that the data are discoverable, accessible, and intelligible?

To make the data findable, it will be described with rich metadata. The metadata will be assigned a globally unique and persistent identifier created by most repositories. This identifier will help to retrieve the data. Furthermore, only well-known and open formats and software will be used to account for the interoperability of the data. Finally, it will be ensured that the data is well-documented to support proper data identification.

How will you share the data?

The data will be shared only in an anonymized form that prevents the identification of the participating individuals, companies.

With whom will the data be shared?

The data will be shared in a research data repository.

SECTION 7. RESPONSIBILITIES AND RESOURCES

Who is responsible for Data Management during the project? This will be the person who might receive questions on the data management aspects of the research project.

Justyna Dlociok, Birgitta Meex and Bert Oben will be responsible throughout the project.

Justyna Dlociok: justyna.dlociok@kuleuven.be

Birgitta Meex: birgitta.meex@kuleuven.be

Bert Oben: bert.oben@kuleuven.be

Which additional resources are needed for the execution of the Data Management Plan?

The data will be stored on KU Leuven OneDrive for Business, with an enabled Multi-Factor Authentication. The backup data storage space will be accounted for on Google Drive, One Drive, and an external hard disk.

Did you read the KU Leuven Research Data Management Policy? (find the link to the policy in the guidance).

Yes, I did.

SECTION 8. PUBLISHING AND OPEN ACCESS

Do you intend to publish your research result in Open Access? If not, why? If yes, would you opt for Green or Gold OA?

Yes, I do intend to publish my research in Open Access.

I would opt for Gold OA because it offers numerous benefits such as increased citation and usage, easy compliance with institutional mandates, retention of copyrights by the author, faster impact, and greater public engagement. For that, a certain amount of money is secured from my project application. Also, I will attempt to share my publications in journals that do not require the author fee (DOAJ).

Do you have any other plans to openly disseminate your research results (e.g., blogs, platforms, etc.)?

Yes, I do have plans to disseminate my research openly. The data will be shared with companies and used by those companies for marketing purposes. Also, the research outcomes will be used during writing workshops.

Bibliography

- Byrne, J. (2004). Textual Cognetics and the Role of Iconic Linkage in Software User Guides.
- Goebel, R. (2014). LNAI 8625 - Controlled Natural Language.
- O'Brien, S. (2003). Controlling Controlled English. An analysis of several controlled language rule sets. Proceedings of EAMT-CLAW, 3, 105–114.
- Shubert, S., Holmback, H., Spyridakis, J., & Coney, M. (1995). Testing the comprehensibility of simplified English: an analysis of airplane procedure documents. IEEE International Professional Communication Conference, 206, 171–173. <https://doi.org/10.1109/ipcc.1995.554896>
- Steensland, H., & Dervisevic, D. (2005). Controlled Languages in Software User Documentation. 224. <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-4637>