FWO DMP - Flemish Standard Data Management Plan

1. General Project Information	
Name Grant Holder & ORCID	Nina Arisci (0000-0002-1190-1897)
Contributor name(s) (+ ORCID) & roles	Supervisors: ● Ellen Rombouts (0000-0003-2700-8910), Assistant Professor at ExpORL, KU Leuven ● Laura Rosseel (0000-0002-1005-7452), Assistant Professor at BCLS, Vrije Universiteit Brussel ● Eline Zenner (0000-0002-8114-5425), Associate Professor at QLVL, KU Leuven
Project number & title	3H220762, The development of socially meaningful language variation in (pre)adolescents with Down Syndrome
Funder(s) GrantID	1105623N
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR identifier when possible: https://ror.org/05f950310
Please provide a short project description	This project studies the development of socially meaningful language variation in (pre)adolescents with Down Syndrome (hence 'DS'). 'Socially meaningful language variation' can be observed when two or more variants denote the same concept, but each of these variants conveys different social meaning about the speaker, the hearer, and the setting in which the variant occurs (Labov, 1972). To illustrate, a Belgian Dutch (hence 'BD') speaker might prefer to address their doctor with the pronominal variant 'u' rather than the variants 'je' or 'ge' because they consider it the most appropriate in the formal context of a doctor's visit. Such BD pronouns of address (including 'u', 'je', and 'ge') are employed as a case study in this project, since they show socially meaningful language variation on two dimensions: On the first dimension, they range from more to less formal, and on the second dimension, they range from standard to colloquial (Plevoets et al., 2008). To obtain the best view into how the development of socially

meaningful language variation is related to social functioning (e.g., socially engaging with others) and language processing (e.g., using certain grammatical structures while talking), the participants of this project are (pre)adolescents with DS. These individuals' have a unique developmental profile: They have strengths in social functioning on the one hand (Næss et al., 2017), and challenges with language processing on the other (Abbeduto et al., 2020). Overall, this project aims to recruit 32 dyads, each consisting of a (pre)adolescent with DS and their caregiver. It will investigate (a) how (pre)adolescents with DS vary in their productions of Belgian Dutch pronouns of address (via spontaneous language samples and a "discourse completion" experiment where they complete utterances), (b) how they vary in their evaluations of Belgian Dutch pronouns of address (via a "speaker evaluation" experiment where they express their attitude towards 'u'/'je'/ 'gij' use in specific contexts and utterances), (c) the role of the caregiver in their productions and evaluations, and (d) how the participants reflect upon the results specific to their dyad, via a semi-structured interview. In line with the United Nation's slogan "Nothing about us without us", the project employs a participatory research framework by including a focus group of (pre)adolescents with DS and their caregiver, a consultation board of persons professionally or privately involved with individuals with DS outside academia, and by involving the participants in the co-analysis of their results. The results are correlated with several measures of language processing and social functioning in the (pre)adolescents with DS to map out the development of socially meaningful language variation. This project contributes to our understanding of the social meaning in language and advances current knowledge in developmental sociolinguistics by adding a unique perspective to the discipline.

Literature

Abbeduto, L., Arias-Trejo, N., Thurman, A. J., Ramos-Sanchez, J., & del Hoyo Soriano, L. (2020). Language Development in Down Syndrome. In J. A. Burack, J. O. Edgin, & L. Abbeduto (Eds.), The Oxford Handbook of Down Syndrome and Development. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190645441.013.18

Fidler, D. J., Most, D. E., Booth-LaForce, C., & Kelly, J. F. (2008). Emerging Social Strengths in Young Children With Down Syndrome. Infants & Young Children, 21(3), 207–220. https://doi.org/10.1097/01.IYC.0000324550.39446.1f

Labov, W. (1972). Sociolinguistic patterns (11. print). Blackwell.

	Næss, K.-A. B., Nygaard, E., Ostad, J., Dolva, A.-S., & Lyster, S.-A. H. (2017). The profile of social functioning in children with Down syndrome. Disability and Rehabilitation, 39(13), 1320–1331. https://doi.org/10.1080/09638288.2016.1194901
	Plevoets, K., Speelman, D., & Geeraerts, D. (2008). The distribution of T/V pronouns in Netherlandic and Belgian Dutch. In K. P. Schneider & A. Barron (Eds.), Variational pragmatics: A focus on regional varieties in pluricentric languages (pp. 181–210). John Benjamins.

2. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

An asterisk (*) indicates that the dataset contains personal data (see further specifications in Sections 2 and 6)		*Generate new data OR Reuse existing data*	*Digital OR physical*	*Only for digital data. Observational OR Experimental OR Compiled / aggregated OR Simulation OR Software OR Other OR NA*	*Only for digital data. .por OR .xml OR .tab OR .csv OR .pdf OR .txt OR .rtf OR .dwg OR .tab OR .gml OR Other OR NA*	*Only for digital data. < 100 MB OR < 1 GB OR < 100 GB OR < 1 TB OR < 5 TB OR < 10 TB OR < 50 TB OR > 50 TB OR NA*	*Only for physical data.*
Dataset name	**Description**	**New or reused**	**Digital or physical**	**Digital Data Type**	**Digital Data Format**	**Digital Data Volume (MB, GB, TB)**	**Physical Volume**
Literature sources	Files with articles from previous research	Reused	Digital and physical	For digital sources: Other	For digital sources: .pdf. Otherwise print versions	For digital sources: TBD	For physical sources: TBD

Bibliographic references	Metadata related to individual literature sources	Reused	Digital	Other	.bib	< 100 MB	NA
Literature folders	Semantically meaningful categorization system for the literature sources	New	Digital	Other	NA	< 100 MB	NA
Highlights	Color annotations in the digitally available literature sources	New	Digital	Other	NA	< 100 MB	NA
Notes	Written annotations in the literature sources	New	Digital	Other	.md	< 100 MB	NA
Tags	Keywords assigned to the literature sources	New	Digital	Other	NA	< 100 MB	NA
Personal notes	Various notes taken throughout the project	New	Digital and physical	Other	.txt	< 100 MB	For physical sources: TBD
Ethics Documents	I.e., protocol, Informed Consent Form, Informed Assent Form...	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Data Management Documents	I.e., Digital Scholarship Plan, Data Management Plan	New	Digital	Other	.doc, .pdf, .xlsx	< 100 MB	NA
Pronouns of address survey	Survey conducted to determine how adults with L1 Belgian Dutch judge the appropriateness of pronouns in a given context	New	Digital	Other	.doc, .pdf, .csv	< 100 MB	NA
Responses to Pronouns of address survey*	(cf. previous item)	New	Digital	Other	.csv	< 100 MB	NA
Recruitment materials	I.e., brochure, flyer, website, logo	New	Digital and physical	Other	.jpg, .png, .pdf	< 100 MB	For physical sources: TBD

Certificate of participation*	Document certifying participants' involvement in the project	New	Digital and physical	Other	.pdf	< 100 MB	For physical sources: TBD
Focus group questions	Questions asked to the focus group regarding the appropriateness of materials and procedures	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Focus group input*	Feedback and ideas provided by the focus group (cf. previous item)	New	Digital	Aggregated data	.doc, .pdf	< 100 MB	NA
Consultation board questions	Questions asked to the consultation board regarding the appropriateness of materials and procedures	New	Digital	Other	.doc, .pdf, .csv	< 100 MB	NA
Consultation board input*	Feedback and ideas provided by the consultation board (cf. previous item)	New	Digital	Aggregated data	.doc, .pdf, .csv	< 100 MB	NA
Participant details*	Participants' identification and contact information	New	Digital	Other	.csv	< 100 MB	NA
Caregiver questionnaire	Document containing questions about participant background variables such as gender, previous exposure to Belgian Dutch dialects etc.	New	Digital	Other	.doc, .pdf, .csv	< 100 MB	NA
Answers to the caregiver questionnaire*	(cf. previous item)	New	Digital	Other	.csv	< 100 MB	NA
Speech-language therapist and/or teacher questionnaire	Document containing questions about participant background	New	Digital	Other	.doc, .pdf, .csv	< 100 MB	NA

	variables such as speech dysfluencies etc.						
Answers to the speech-language therapist and/or teacher questionnaire*	(cf. previous item)	New	Digital	Other	.csv	< 100 MB	NA
Diagnostic Assessments	Tests to assess general language production, social functioning, and input characteristics	Reused. Only exception: adapted auditory discrimination test (= new)	Digital and physical (according to the assessor's preference)	Other	.pdf	If digital: < 100 MB	If physical: According to the size of the original Diagnostic Assessments.
Results obtained in the Diagnostic Assessments*	(cf. previous item)	Reused (if collected as secondary data) or new (if collected as primary data). Secondary data is prioritized if available	Digital (any physically collected data – cf. previous item – is digitized and the physical data subsequently destroyed)	Other	.pdf, .csv	< 100 MB	NA
Spontaneous production audio samples*	Audio-recordings obtained by participants in their everyday lives	New	Digital	Observational	.mp3	< 100 GB	NA
Annotations of the spontaneous production audio samples*	Including transcriptions, codes etc. (cf. previous item)	New	Digital	Observational	.eaf, .csv	< 100 MB	NA
Experimental Stimuli	Stimuli developed for the discourse completion task and for the speaker evaluation experiment	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Responses to the Stimuli*	(cf. previous item)	New	Digital	Experimental	.csv	< 100 MB	NA

Video-recordings*	Made of the experimental sessions and of the interviews	New	Digital	Observational	.mp4	< 1 TB	NA
Annotations of the video-recordings	Including transcriptions, codes etc. (cf. previous item)	New	Digital	Observational	.eaf, .csv	< 100 MB	NA
Interview form*	Document protocoling the interview procedure	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Answers to the Interview questions*	I.e., video recordings, video transcript, handwritten notes (cf. previous item)	New	Digital and physical	Other	.mp4, .eaf, .csv, .doc, .pdf	< 100 MB	For physical sources: TBD
Recruiting materials	I.e., brochure, flyer, website, logo	New	Digital	Other	.csv, .pdf, .jpg	< 100 MB	NA
Journal Articles	Articles for an academic audience	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Conference outputs	Presentations and/or proceedings for an academic audience	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Presentations	At events other than conferences, for an academic or non-academic audience	New	Digital	Other	.pptx, .pdf	< 100 MB	NA
Publications on further platforms	I.e. podcasts, blogs, magazines...	New	Digital	Other	TBD	TBD	NA
Easy-to-read booklet	Summary of project results for (pre)adolescents with Down Syndrome	New	Digital and physical	Other	.doc, .pdf	< 100 MB	For physical sources: TBD
Project summary	Summary of project results for adults	New	Digital and physical	Other	.doc, .pdf	< 100 MB	TBD

Dissertation (digital version)	Written document elaborating and synthesizing PhD studies	New	Digital	Other	.doc, .pdf	< 100 MB	NA
Dissertation (printed version)	(cf. previous item)	New	Physical	NA	NA	NA	TBD

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	Reuse of existing data concerns literature sources and diagnostic assessments. Literature sources are stored in Zotero and will be cited in the materials where appropriate. The sources of the Diagnostic Assessments are: • de Jong, J. (2012). *CELF Preschool-2-NL: Clinical evaluation of language fundamentals preschool—2 Nederlandstalige versie*. Pearson. • Geurts, H. M. (2007). *CCC-2-NL: Children's Communication Checklist-2*. Pearson Assessment and Information B.V. • Hendriks, M. P. H., Ruiter, S., Schittekatte, M., & Bos, A. (2018). *WISC-V-NL Wechsler Intelligence Scale for Children-V*. Pearson. • Kort, W., Schittekatte, M., & Compaan, E. (2010). *CELF-4-NL: Clinical Evaluation of Language Fundamentals Nederlandse versie*. Pearson. • Roeyers, H., Thys, M., Druart, C., De Schryver, M., & Schittekatte, M. (2015). *SRS-2 Screeningslijst voor autismespectrumstoornissen*. Hogrefe. • Wiig, E., Semel, E., & Secord, W. E. (2019). *CELF-5-NL: Clinical Evaluation of Language Fundamentals – Fifth edition – Nederlandstalige versie*. Pearson.
Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, please describe these issues further and refer to specific datasets or data types when appropriate.	☒ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☐ No <u>If yes, please describe:</u> The project involves collection of personal data from the participants through several non-invasive methods (datasets which contain personal data are indicated in the table above with an asterisk). The personal data collected from participating minors (with Down Syndrome) is particularly sensitive, as they are not legally able to give informed consent (though they will be asked to give informed assent).

Will you process personal data? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☒ Yes Short description of the kind of personal data that will be used: Datasets which contain personal data are indicated in the table above with an asterisk. Privacy Registry Reference: G-2022-5917 (KU Leuven PRET-number), S67580 (Ethics Committee Research UZ / KU Leuven S-number) ☐ No
Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.	☐ Yes ☒ No If yes, please comment: NA
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☒ Yes Explanation: Literature sources and Diagnostic Assessments are subject to copyright and shall be reused only with appropriate reference to the original source. Results from Diagnostic Assessments may have been previously collected by (pre)adolescents with DS' speech-language-therapist within the scope of diagnostics and therapy. These materials shall only be reused upon caregivers' consent obtained in the Informed Consent Form. ☐ No
Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No

3. Documentation and Metadata	
Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).	Documentation and metadata will be systematically developed and indicated throughout the project. At project end, all metadata and related documentation will be re-evaluated in terms of understandability and usability. The specific measures taken include: • Organization and storage of sources relevant for literature management (articles, tags, annotations) on Zotero. A readme-file describing the principles of organization (e.g., folder name system or color annotations) will be added to the main library. • Organization of all further project documents in a folder system. The file and folder names will be schematically consistent and semantically meaningful. Folders will contain readme-files wherever deemed helpful for a full understanding of their contents, and individual files will contain links to other files if appropriate (via indication of file names). Versioning will be implemented with all the files created throughout the project and older versions will be archived.
Will a metadata standard be used to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse.	☒ Yes Specification: The CHAT transcription conventions will be relied on for transcription of audio and video files. These are well-established in linguistics and allow transcribing multimodal communication (e.g., gestures). Further conventions are TBD. ☐ No

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	Pseudonymized data will be stored on SharePoint Online. Non-pseudonymized data (i.e., the audio recordings, the video recordings, the contact list containing personal and identification data) will be stored on an external hard disk.

How will the data be backed up?	SharePoint Online provides an automatic backup system, which performs backups every few hours of the day where the 7 versions are saved, as well as every day where the last 6 versions are saved, and every week where the last 2 versions are saved. The external hard disk will be backed up with a second external hard disk. These backups will be performed manually whenever changes are made on the principal external hard disk. Backup events will be documented. Physical data other than literature sources available at the library for which no digital pendant exists will be immediately digitized and stored on one of the two platforms described above (as appropriate in the individual case).
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	☒ Yes Specification: SharePoint Online offers 20 GB net storage, which pseudonymized data is not expected to surpass throughout the project. External hard disks will be acquired in line with the expected storage space required (currently estimated to be < 1 TB in total). ☐ No
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	In consultation with two KU Leuven ICT experts, all data will be encrypted via the tool AES Crypt before being stored. SharePoint Online is secured via Multi-Factor Authentication. The external hard disk will be stored at the PhD student's office space at KU Leuven Brussels campus, where it will be locked with a key that is only available to the PhD student. Outside office hours, the building is only accessible for KU Leuven members. The office itself is only accessible for KU Leuven staff who have a fixed desk in that office. All data will be processed on company-owned laptops protected by Bitlocker Drive Encryption, which integrates with the operating system and protects the data from theft or exposure from lost, stolen, or inappropriately decommissioned computers. Finally, as tools and regulations for protecting sensitive data evolve rapidly, IT and privacy experts at KU Leuven will be consulted regularly to re-evaluate the most current security measures to be taken and to ensure that the latest laws and regulations are followed. A PRET application to describe the processing of the participants' personal data has also been submitted and necessary updates or improvements included in the feedback will be taken into consideration.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Free storage of pseudonymized data on SharePoint Online, storage of non-pseudonymized data on external hard drive (likely between 100 and 300 euros per exemplar). These costs will be covered through the FWO bench fee.

5. Data Preservation after the end of the Research Project	
Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).	The pseudonymized data will be retained for 10 years after the end of the project. The non-pseudonymized data will be deleted at project end, given that it is highly sensitive data.
Where will these data be archived (stored and curated for the long-term)?	On SharePoint Online.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	No costs are expected.

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available.	☐ Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☒ Other Specification: Pseudonymized data will be made available under restricted access, non-pseudonymized data will be presented in terms of metadata under closed access. Excerpts of the data can be made available to other reviewers on request, and after signing a confidentiality agreement.
If access is restricted, please specify who will be able to access the data and under what conditions.	Access to pseudonymized data will be given to other reviewers on request, and after signing a confidentiality agreement.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain per dataset or data type where appropriate.	☒ Yes, privacy aspects ☒ Yes, intellectual property rights ☒ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☐ No Specification: All datasets containing personal data (see asterisks in Table above) are particularly sensitive in terms of privacy and related ethical considerations (e.g., given that minors are able to give informed assent but not informed consent). They are thus only shared after pseudonymization. Literature sources are subject to intellectual property rights and are accordingly referred to wherever appropriate.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Data will be made available on the data repository of the Meertens Institute (see https://www.re3data.org/repository/r3d100011334).
When will the data be made available?	Upon publication of research results.
Which data usage licenses are you going to provide? If none, please explain why.	Data that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators.
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.	☒ Yes ☐ No
What are the expected costs for data sharing? How will these costs be covered?	No costs are expected.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The PhD researcher (Nina Arisci).
Who will manage data storage and backup during the research project?	The PhD researcher (Nina Arisci).
Who will manage data preservation and sharing?	The PhD researcher (Nina Arisci) until the end of the project, supervisor prof. Zenner after the end of the project.
Who will update and implement this DMP?	The PhD researcher (Nina Arisci).