Sign Language Translation Mobile Application and Open Communications Framework

Deliverable 9.5: First Annual Ethical Report on Internal Affairs

Project Information
Project Number: 101017255
Project Title: SignON: Sign Language Translation Mobile Application and Open Communications Framework
Funding Scheme: H2020 ICT-57-2020
Project Start Date: January 1st 2021

Deliverable Information
Title: First Annual Ethical Report on Internal Affairs
Work Package: WP 9- Ethics
Lead beneficiary: Trinity College Dublin
Due Date: 31/12/2021
Revision Number: V1.0
Authors: Lorraine Leeson, Dimitar Shterionov, Vincent Vandeghinste, Henk van den Heuvel, Josep Blat, Jorn Rijkaert, Myriam Vermeerbergen, Catia Cucchiarini, and Aoife Brady
Dissemination Level: PUBLIC
Deliverable Type: Report

Overview: This report offers an overview of activities of the SignON Ethics Committee over year 1 of the project.

Revision History

Version #	Implemented by	Revision Date	Description of changes
V1.0	Lorraine Leeson	17/12/2021	Final version

The SignON project has received funding from the European Union's Horizon 2020 Programme under Grant Agreement No. 101017255. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the SignON project or the European Commission. The European Commission is not liable for any use that may be made of the information contained therein.

The Members of the SignON Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the SignON Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Approval Procedure

Version #	Deliverable Name	Approved by	Institution	Approval Date
V1.0	D9.5	Aoife Brady Shaun O'Boyle	DCU	02/12/2021 10/12/2021
V1.0	D9.5	Andrea Cavallaro	FINCONS	07/12/2021
V1.0	D9.5	Vincent Vandeghinste	INT	29/11/2021
V1.0	D9.5	Olatz Perez de Viñaspre Adrián Núñez-Marcos	UPV/EHU	03/12/2021

V1.0	D9.5	John J O'Flaherty	MAC	02/12/2021
V1.0	D9.5	Santiago Egea Gómez	UPF	07/12/2021
V1.0	D9.5	Irene Murtagh	TU Dublin	10/12/21
V1.0	D9.5	Lorraine LEESON	TCD	28/11/2021
V1.0	D9.5	Karim Dahdah	VRT	10/12/2021
V1.0	D9.5	Mathieu De Coster	UGent	06/12/2021
V1.0	D9.5	Caro Brosens	VGTC	08/12/2021
V1.0	D9.5	Anthony Ventresque	NUID UCD	10/12/2021
V1.0	D9.5	Henk van den Heuvel	RU	04/12/2021
V1.0	D9.5	Catia Cucchiarini	TaalUnie (NTU)	10/12/2021
V1.0	D9.5.	Myriam Vermeerbergen	KU Leuven	04/12/2021
V1.0	D9.5	Davy Van Landuyt	EUD	08/12/2021
V1.0	D9.5	Mirella De Sisto	TiU	06/12/2021

Acronyms

The following table provides definitions for acronyms and terms relevant to this document.

Acronym	Definition
MTSL	Machine Translation for Sign Languages
REC	Research Ethics Committee
SL	Sign Language
DMP	Data Management Plan

Table of Contents

Note to reader:

Please note that this report and the content of our report in D9.2 is highly similar. This is because the information we wish to share here, in our internal report, is very similar to the content that we include in our report to our stakeholders (D9.2). Both contain information about the membership of our SignON REC, our meeting schedule, the work we have done to map our ethical procedures to our co-construction methodology. We also include in both reports, information relating to where we get our data from, what kinds of data we use at what parts of the research pipeline, and our concern (shared with deaf communities) around the fact that there is so little data that is 'born' as sign language material - instead much (most) content in the public sphere is interpreted data.

1. Overview

This report offers an overview of the work that the SignON Research Ethics Committee (REC) has completed over the first year of the project.

2. SignON Research Ethics Committee Membership and Meetings across 2021

The SignON REC comprises the following members:

	Name	Institution	Country
1	Prof. Lorraine Leeson	TCD	IE
2	Dr. Dimitar Shterionov	TiU	NL
3	Dr. Vincent Vandeghinste	INT	NL
4	Dr.. Bob Boelhouwer	INT	NL
5	Dr. Henk van den Heuvel	RU	NL
6	Prof. Josep Blat	UPF	ES
7	Caro Brosens	VGTC	BE
8	Prof. Myriam Vermeerbergen	KUL	BE
9	Dr. Catia Cucchiarini	DLU	NL
10	Aoife Brady	DCU	IE

The Committee has formally met online 6 times across the year to date, on the following occasions:

- 3 March 2021
- 29 April 2021
- 20 May 2021
- 9 June 2021
- 29 October 2021

● 1 December 2021

Minutes of these meetings are available.

3. Key Issues Addressed

3.1 European Code of Conduct for Research Integrity

We are committed to the principles of the European Code of Conduct for Research Integrity (ALLEA (All European Academies), 2017) and are working to uphold the four key tenets of (i) reliability, (ii) honesty, (iii) respect and (iv) accountability.

3.2 Mapping Ethics to a Co-construction Philosophy

Key issues addressed by the SignON REC included preparation of our deliverables D9.1, D9.2 and D9.5.

In developing D9.1, our "Ethical Guidelines and Protocols" document, we explored how our ethical philosophy would map to and support our project's commitment to co-construction. We committed to the Sign Language Communities' Terms of Reference Principles (Harris et al., 2009) which state that:

1. Investigators should acknowledge that Sign Language community members have the right to have those things that they value to be fully considered in all interactions.

2. Investigators should take into account the worldviews of the Sign Language community in all negotiations or dealings that impact on the community's members.

3. In the application of Sign Language communities' terms of reference, investigators should recognize the diverse experiences, understandings, and way of life (in sign language societies) that reflect their contemporary cultures.

4. Investigators should ensure that the views and perceptions of the critical reference group (the sign language group) is reflected in any process of validating and evaluating the extent to which Sign Language communities' terms of reference have been taken into account.

5. Investigators should negotiate within and among sign language groups to establish appropriate processes to consider and determine the criteria for deciding how to meet cultural imperatives, social needs, and priorities.

The team also adopted the Sign Language and Linguistics Society (SLLS)'s Ethics Statement and we are committed to embedding these principles in our approach to ethical engagement with Deaf communities [1].

3.3 Developing the SignON REC Process

The SignON REC also spent time developing the protocol to support any partner who seeks to undertake research that involves human subjects. We developed a protocol for reviewing applications for local (institutional) research ethics approval, with a SignON REC preview of all applications that are submitted. This allows us to ensure that the principles we uphold as a consortium are consistently implemented in our documentation.

Our goal was not to supersede the institutional research ethics committees to whom university based partners must apply ahead of commencing on any data collection involving human subjects. Rather, on a practical level, this means that the SignON Ethics Committee has sight of participant information that is shared with all stakeholder organisations who engage in our project related research activities (e.g. this allows us to confirm that Participant Information Leaflets (PILs) and Consent forms are available in accessible languages).
This is important for a number of reasons:
- Ensures access to information about the project in a language that is understandable to the participants.
- Ensures that the implications of participation are clearly articulated in an accessible language.
- Ensures that rights of withdrawal without consequence are clearly articulated in an accessible language.

To date, two applications have been reviewed and commented on, with additional applications due in early 2022. Both have successfully been approved at institutional level and data collection has taken

---

[1] https://slls.eu/slls-ethics-statement/

place, facilitating the work of WP1. The SignON REC process therefore facilitates a harmonisation of approach to our data collection and offers an additional layer of comfort with regard to ensuring that we are embedding the Sign Language Communities Terms of Reference and our co-construction principles in the work that most directly draws on deaf and hard of hearing communities input.

3.4 Data

An issue that we have spent quite some time discussing throughout 2021 is data provenance and usage, as this raises ethical questions that map to the issues of power (who gets to select data, who determines what level of quality of the sign language data input and output is satisfactory), authority (who determines what the quality of data input to the project should be) and future consequences (whose voice is heard by the European Commission and national funders – e.g. deaf communities call for investment in human resources and language documentation and natural language data set building projects yet large scale funding for such initiatives is extremely difficult to source. Deaf communities see investment in large scale 'tech solutions' that may be very imperfect and do not address the pressing current needs of their members. Deaf communities feel that this funding goes to institutions and organisations from the technical sphere. These are considered powerful (and are mostly 'hearing') voices, and are felt to be more likely to secure funding than institutions/organisations working on less technically biased or non-technical issues identified as important to deaf communities) .

It is clear that there is a concern in thedeaf communities we engage with around the kinds of sign language data leveraged in machine translation for sign languages (MTSL). From the perspective of deaf communities, the use of data containing sign language materials that originates from within the community of deaf signers is the gold standard, but there is insufficient source sign language data in existence to work effectively with just these sets. Instead, interpreted data sets are also drawn on e.g. online/broadcast data where (typically, hearing) sign language interpreters work simultaneously to present content in a target sign language. There are challenges here from a language quality perspective, which include (but are not limited to) the following issues:

(i)   Interpreted data is not spontaneous language use and may be heavily influenced by the source language resulting in so-called *translationese*;

(ii)   The majority of sign language interpreters have learned to sign as adults, with associated late learner "accents", and reduced syntactic/semantic/pragmatic/lexical range vis-à-vis those who have a sign language as a first or preferred language;

(iii)   Working simultaneously places significant cognitive load on an interpreter, with trade-offs for how the target language information is packaged, influenced by the source language structure.

Additionally, datasets that do exist are not – or not appropriately marked up for use in projects as no standard annotation process currently exists for use with sign language data sets and no automatic tagging system has yet been developed to assist in the process of data tagging. Thus, we note the need for investment of resources in deaf community originated sign language content (e.g. expanded corpora) and annotation projects to support the development of more robust datasets for a wider range of European sign languages to help us to ensure that we can build MTSL products that are informed by natural sign language content as articulated by signers who use the language as their preferred or first language. At the same time, it is clear that the reasons why certain kinds of data suffice for different points in the MTSL pipeline have been unpacked and discussed. In regular MT research data usually comes in the form of parallel training data, consisting of source text with sentence-aligned translations in the target language. As there is no standardized written form of sign languages (SLs), and no unified annotation approach in different sign language corpora for different languages, this is a challenge. Apart from the fact that many of the SL corpora are partially annotated, only certain features of the SL can be automatically extracted by computer vision techniques. It remains to be seen, through experimentation, which are the most effective features to be used. Datasets that have an SL as source are already scarce, having a reliable translation in a spoken language is even rarer. Discussions with deaf communities about which datasets are suitable are ongoing. While the AI researchers usually adopt a *there is no data like more data* approach, deaf communities have a key concern that the data are of high quality (i.e. approved or generated by a deaf community). This discussion is still ongoing. In light of this, our Belgian colleagues, including AI researchers, interpreters and members of the Flemish deaf community, will meet in December 2021. Their goal is to exchange ideas and concerns regarding the data in order to improve cooperation and enhance understanding.

We have also committed to having a consortium-wide discussion about data at our February 2022 meeting to ensure that this point – one that is extremely important to the deaf communities we engage with – is fully understood by everyone working on the SignON team, and that we (within the consortium) understand why different kinds of data are appropriate (or not) for use at different phases of the

project's work. This, in turn, will situate us all to be better able to engage in an open and clear discussion on this matter with our local deaf communities.

Finally, the SignON REC has an important role in the Research Data Management of the project. A generic Data Management Plan (DMP) was set up and submitted as deliverable D7.8. This deliverable outlines the framework and the principles which individual DMPs in the project should meet in line with the contract obligations stipulated in Article 29.3 of the Grant Agreement of the project. D7.8 is complemented by the specific DMPs for each partner. These will be delivered in D7.10 which is due in M36. In order to accommodate for the dynamic nature of DMPs intermediate versions of D7.10 are scheduled for M12 and M24. These specific DMPs are reviewed by SignON's Research Ethics Committee (REC). The first versions of the DMPs of individual partners have now been written and reviewed by the REC. Last, but not least a template for a Data Transfer Agreement is in preparation which will allow SignON partners to share (personal) data amongst each other.

3.5 Authorship Guidelines

Another issue that emerged was that of authorship. Our SignON team has adopted the CREDiT matrix[2] approach to authorship allocation and we have issued reminders that we wish to reflect on and acknowledge the various contributions that our team and collaborators make in the work that we do. We acknowledge that different disciplines have traditionally had a variety of approaches to determining authorship and position of authors on publications and we have offered a framework for determining authorship on our consortium outputs. We have also emphasized that 'gift authorship' is not acceptable.

4. Key Deliverables

The key deliverables that the SignON Research Ethics Committee were responsible for developing in 2021 were threefold:

(i)   Ethical Guidelines and Protocols (D9.1) (Completed June 2021)

(ii)   Annual Ethical Report on Practices towards the Stakeholder Communities (D9.2) (Due December 2021)

---

[2] https://casrai.org/credit/

(iii)   Annual Ethical Report on Internal Affairs (D9.5) (This deliverable, due December 2021).

5. Summary

Across 2021, the SignON REC has worked collaboratively and effectively to support the work of the project consortium. While mainstream ethical considerations which address all research, are addressed by our REC, the particular considerations of what it means to be a multi-disciplinary team comprising mostly hearing members working with deaf signing communities that have been marginalised has been a key focus. Discussions that have emerged here have informed work in WP1 and WP3 and fed back into wider conversations at the Project Management Team Meetings. Our discussions will inform further consortium wide reflection and feed into ongoing conversations with stakeholders in our target communities.

References

ALLEA (All European Academies). (2017). The European Code of Conduct for Research Integrity (Revised Edition). In. Berlin: ALLEA (All European Academies).

Harris, R., Holmes, H. M., & Mertens, D. M. (2009). Research ethics in sign language communities. *Sign Language Studies*, *9*(2), 104-131.