

## Report on the outcomes of a Short-Term Scientific Mission<sup>1</sup>

**Action number:** CA19102

**Grantee name:** Joanna Dolińska

### **Details of the STSM**

Title: How can users of the endangered, selected Inner Asian and Southeast Asian languages benefit from voice technologies?

Start and end date: 11/09/2023 to 24/09/2023

### **Description of the work carried out during the STSM**

During Dr. Joanna Dolińska's STSM stay at the Faculty Campus Fryslân in Leeuwarden, she successfully established with Dr. Matt Coler, the Director of MSc Voice Technology Program, a cooperation network between the Faculty Campus Fryslân (University of Groningen) and the Faculty of "Artes Liberales" (University of Warsaw) dedicated to collaboratively exploring voice technology opportunities for the lesser resourced languages, with the focus on the minority languages in Thailand. During the meetings with the team members of the Voice Technology Program, Dr. Dolińska identified ethical implications for the speakers of the language minorities in Thailand which result from voice technologies. Soon after her arrival to Campus Fryslân Dr. Dolińska delivered a public 45-minutes long seminar "The Social and Ecological Environment Influence on Multilingualism in the Akha, Lahu and Karen communities of Northern Thailand" representing her field work research results, a short lecture "The historical beginnings of speech processing" and a workshop "ASR for diverse genres, language varieties and small languages" for around 45 people, including the staff members and students of the Voice Technology program. Furthermore, she has received inquiries from interested Students concerning the potential co-supervision of their master theses. In addition, Dr. Dolińska has been invited by Dr. Jelske Dijkstra to visit the Fryske Academy, where she also met Drs. Cor van der Meer and the Team of the The Mercator European Research Centre on Multilingualism and Language Learning, hosted by Fryske Akademy. During this visit J. Dolińska learnt about the Frisian language policy and revitalization strategies, as well as speech technology for Frisian language and she shared her academic expertise on the minority languages in Thailand and Poland. Moreover, Dr. Joanna Dolińska participated in Dr. Laurent Besacier's lecture "Automatic Speech Recognition: Introduction, Current Trends and Problems". In addition, Dr. Dolińska received a thorough introduction by Dr.

---

<sup>1</sup> This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

Shekhar Nayak on audio data collection in field work research. In so doing, Dr. Dolińska fulfilled her goals specified in three work packages:

**WP1 Knowledge Exchange** – Dr. Dolińska shared her expertise in endangered languages, as well as in Automatic Speech Recognition with the students and Researchers from Campus Fryslân and the Researchers from the Fryske Akademy. She learnt about the scope of research of the Researchers from Voice Technology Program, while conceptualizing further cooperation with them and participating in the lecture by Dr. Laurent Besacier.

**WP2 Ethics and Data Management** – Dr. Joanna Dolińska participated in several discussions with Dr. Matt Coler and Dr. Shekhar Nayak on the ethics and data management related to collecting, storing and processing voice data.

**WP3 Workshop and dissemination** – Dr. Joanna Dolińska delivered a workshop and two lectures to the students and Researchers from the Voice Technology Program, Campus Fryslân and invited her network from the Faculty of “Artes Liberales” to participate online.

*(max. 500 words)*

**Grantee enters max 500 word summary here.**

### **Description of the STSM main achievements and planned follow-up activities**

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

The main STSM achievements included a public **workshop and lecture** at the Faculty Campus Fryslân, University of Groningen delivered by Dr. Joanna Dolińska during a seminar organized by Dr. Matt Coler for the students and researchers from Campus Fryslân. Another important milestone was a development plan for a joint **paper** addressing the aspects of voice technology application in relation to endangered languages of Inner Asia and Southeast Asia. Furthermore, Dr. Dolińska presented Dr. Coler with first draft ideas for a future **grant application** for Marie Skłodowska-Curie Postdoctoral-Fellowship 2024. In addition, Dr Dolińska has been invited to the future development of the idea of “**Center of Excellence**” that might be created at the Campus Fryslân in the future. Furthermore, Dr. Dolińska has been invited by Dr. Coler to deliver a public workshop and lecture devoted to Automatic Speech Recognition and endangered languages during the **Summer School** at Campus Fryslân in early July 2024. Further cooperation plans include Dr. Coler’s and Dr. Dolińska’s **joint supervision of master theses**. The students of Voice Technology Program expressed their initial interest in carrying out research on the voice technologies for endangered languages and received the possibility from Dr. Coler to sign up for a co-supervised master thesis focused on the above-mentioned topic. This STSM allowed Dr. Dolińska to study the methodology of voice technologies applied to lesser resourced languages and learn how to develop ASR tools for the selected languages. The knowledge exchange during this STSM encompassed the topics from **computational linguistics, language rights, language diversity, vitality and endangerment, ideologies, beliefs and attitudes**, as well as **language variation**.

This STSM advanced Dr. Dolińska’s understanding of emerging technologies and their potential influence on the endangered languages of Thailand. Furthermore, thanks to the acquired knowledge Dr. Dolińska will be able to share informed advice concerning voice technology with the interested language communities while carrying out her fieldwork research in November and December 2023. Moreover, Dr. Dolińska and Dr. Coler started a dialogue between Faculty of “Artes Liberales” and Campus Fryslân on the accessibility standards for linguistics and technology developers working on

endangered languages. In addition, this STSM certainly enabled Dr. Dolińska to develop methods to study Southeast Asian languages in the human-machine era. Dr. Dolińska received valuable advice on the ethics, technicalities and best practices concerning collecting audio data and creating data sets for endangered languages. Last but not least, this STSM established the preliminary framework for dialogue among linguistics and voice technology. This framework includes potential series of regular meetings (in the form of master thesis supervision and Dr. Dolińska's participation in the Summer School 2024) devoted to the voice technology and linguistics. As a conclusion, this STSM undoubtedly contributed to the creation of a working group with critical mass to drive scientific progress through the application of new development technologies to the endangered languages with the simultaneous focus on establishing and sharing ethical standards among the engaged Researchers and Students from the Faculty Campus Fryslân and Faculty of "Artes Liberales".

*(max. 500 words)*

**Grantee enters max 500 word summary here.**