# Report on the outcomes of a Short-Term Scientific Mission[1]

**Action number: CA21129**

**Grantee name: Manjola Lumani (Zaçellari)**

## Details of the STSM

Title: Internship in digital humanities

Start and end date: 31/07/2023 to 31/08/2023

## Description of the work carried out during the STSM

During my time at the ACDH-CH (The Austrian Center for Digital Humanities and Cultural Heritage), Austrian Academy of Sciences I had the opportunity to experience the full spectrum of research and activities carried out at the institute.

First of all, I was part of 'Meet The Researchers' programme, which allowed to meet with ACDH-CH digital humanities experts, gain insight into various projects and activities, and experience the scope of Digital Humanities research today.

The 'Meet The Researchers' programme is to learn more about the inner workings of the ACDH-CH and its broad range of projects and a great opportunity to get to know more of the ACDH-CH staff and other interns.

The working plan at the Academy of Sciences in Vienna:

01.08.2023 Welcome to the ACDH-CH
02.08.2023 – 07.08.2023 Working with **Dr. KARLHEINZ MÖRTH**, DARIAH (Digital Research Infrastructures for the Arts and Humanities) and CLARIN- Common Language Resources and Technology Infrastructure (opinion mining)
09.08.2023 – 10.08.2023 Working with **Dr. KIKI CZEITSCHNER**, Text Encoding Initiative (TEI) and the OXygen XML Editor, Semantic annotation
14.08.2023 – 15.08.2023 working with **PETER KÖNIGSHOFER**, HAPA project, Digital Lexicography
16.08.2023- 18.08.2023 working with **Dr. PHILIPP STÖCKLE**, WBÖ project, Digital Lexicography
23.08.2023 – 28-08.2023 working with **Dr. TANJA WISSIK**, CLARIN project (Common Language Resources and Technology Infrastructure), Opinion mining, Sentiment analysis
31.08.2023 09:00 Working with **Dr. MARKUS KUNZMANN**, WBÖ and LIÖ project, Digital Lexicography

These meetings have provided an insight into the range of projects carried out at the ACDH-CH. Furthermore, tools and methods used in the Digital Humanities were also presented.

[1] 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.

The number of hours for my internship covered 120 (30h/week).

Furthermore, I was also able to learn and use the CLARIN tools and platforms (with the assistance of Dr. Karlheinz Mörth and Dr. Tanja Wissik), which provides access to tools dedicated to sentiment analysis or opinion mining, therefore having a practical approach to text analysis methods that identify and extract people's opinions, attitudes and sentiments within a text. I have also mastered software applications for opinionated communication. (details in the following section).

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

I participated in several ACDH-CH's research projects (such as DARIAH-EU, DEMOS, DSA, DITAH, The ABC of dialects: exploring historical notes digitally, SEMANTICKRAUS, etc.) and used tools from various infrastructures (such as CLARIN) that engage in humanities scholarship with the help of digital tools related to linguistics.

I also dealt with insights into developing and improving software for DH research, data analysis, user interface design, in relation to HAPA ("Linguistic History of Albanian Place Names" project, directed by Dr. Joachim Matzinger).

I was also able to support ACDH-CH networks, knowledge transfer, and outreach teams in their task of internal and public communication, event management, and other outreach activities.

My internship was twofold:

- Conduct research related to the topics these projects cover
  (the publication of a paper is also considered in collaboration with Dr. Tanja Wissik and Dr. Besim Kabashi related to sentiment analysis of various Twitter content extracted from the Albanian language)

- Being part of the team of several projects at the Academy of Sciences in Vienna for a month has contributed to my capacity building as a young researcher by progressing in the field of digital humanities and computational linguistics.

After completing this internship, I have therefore strengthened my capacity to use computational methods for studying digital texts, which is one of the objectives of OPINION COST action.

Specifically, my internship was focused on using CLARIN (Common Language Resources and Technology Infrastructure) infrastructure.

Together with Dr. Karlheinz Mörth and Dr. Tanja Wissik we explored the potential of CLARIN tools and resources in sentiment analysis and opinion mining research.

First of all, I got familiarized with CLARIN Infrastructure, its tools, and resources. I gained an understanding of how CLARIN facilitates access to language resources and provides computational tools for various natural language processing tasks. This included exploring the CLARIN website, reading relevant documentation, and interacting with the CLARIN community.

To gain a comprehensive understanding of opinion mining, I conducted an extensive literature review. I explored various research papers, surveys, and articles related to sentiment analysis and opinion mining. This allowed me to grasp the theoretical foundations of the field and identify potential research gaps to address during the internship.

I then focused on data collection and pre-processing. I leveraged the CLARIN infrastructure to access relevant language resources and datasets for opinion mining. I learned how to retrieve, filter, and pre-process textual data using CLARIN tools, ensuring data quality and consistency.

Building upon the acquired knowledge, I performed sentiment analysis experiments using CLARIN tools and resources. I explored different approaches, such as rule-based, machine learning-based, and

hybrid models, to extract sentiment from text. This experimentation allowed me to gain hands-on experience with CLARIN-supported sentiment analysis techniques and refine my understanding of their strengths and limitations.

After conducting sentiment analysis experiments, I evaluated the performance of the implemented models using appropriate evaluation metrics. I analyzed the results, identified areas of improvement, and explored strategies to enhance the accuracy and reliability of opinion-mining techniques. This analysis helped me gain insights into the challenges of sentiment analysis and the potential of CLARIN infrastructure in addressing them.

On the final day, I prepared a presentation summarizing my internship experience and findings.