



---

## Deliverable 4.3 Final Report on ROXANNE Data

---

<b>Grant Agreement:</b>	833635
<b>Project Acronym:</b>	ROXANNE
<b>Project Title:</b>	Real time network, text, and speaker analytics for combating organised crime
<b>Call ID:</b> <b>Call name:</b>	H2020-SU-SEC-2018-2019-2020, Technologies to enhance the fight against crime and terrorism
<b>Revision:</b>	V1.0
<b>Date:</b>	23 December 2022
<b>Due date:</b>	30 November 2022
<b>Deliverable lead:</b>	BUT
<b>Work package:</b>	WP4
<b>Type of action:</b>	RIA

## Disclaimer

The information, documentation and figures available in this deliverable are written by the “ROXANNE - Real time network, text, and speaker analytics for combating organised crime” project’s consortium under EC grant agreement 833635 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

## Copyright notice

© 2019 - 2022 ROXANNE Consortium

Project co-funded by the European Commission within the H2020 Programme (2014-2020)		
Nature of deliverable:		R
<b>Dissemination Level</b>		
<b>PU</b>	Public	<input checked="" type="checkbox"/>
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	<input type="checkbox"/>
<b>EU-RES</b>	Classified Information: RESTREINT UE (Commission Decision 2015/444/EC)	<input type="checkbox"/>
* R: Document, report (excluding the periodic and final reports) DEM: Demonstrator, pilot, prototype, plan designs DEC: Websites, patents filing, press & media actions, videos, etc. OTHER: Software, technical diagram, etc.		

## Revision history

Revision	Edition date	Author	Modified Sections/Pages	Comments
V0.1	01.11.2022	Erinc Dikici (HENS)	All	Template
	09.11.2022	Marek Kovac		minor changes in headers
	10.11.2022	Zahra Ahmadi	3.5	Write on burglary dataset
	17-20.11.2022	Erinc Dikici (HENS)	2.1, 2.3.2	Edited document format, added first draft
	18-21.11.2022	Johan Rohdin	Executive summary, 1, 3, 5, 3.6	Various additions in these sections.
	23-25.11.2022	Marek Kovac	2, 2.6.2, 2.6.3, Annex 2	Edited audio recording, updated text
	25.11.2022	Nikos Nikolaou (ITML)	2.5	Writing, editing on ROXHOOD dataset
	27.11.2022	Johan Rohdin (BUT)	1,5	Editing these sections
	25-30.11.2022	Erinc Dikici (HENS)	1, 2.3, Annex 3	Added metadata descriptions, figures and tables. Edited Section 1. Inserted Annex 3
	02.12.2022	Dawei Zhu (USAAR)	3	Provided comments for Section 3
	02.12.2022	Nikos Nikolaou (ITML)	2.3	Review. Comments
	04.12.2022	Erinc Dikici (HENS)	1, 2.1, 2.2, 2.3	Proofreading
	06.12.2022	NFI	2.6	Review
	15-18.12.2022	Erinc Dikici (HENS)	2, Annexes	Minor edits
	19-12.2022	Johan Rohdin (BUT)	1,3	Addressing comments.
V1.0	23-12.2022	Petr Motlicek (IDIAP)	All	Review



## Executive summary

This deliverable, D4.3, *Final report on ROXANNE data describes* describes all data sets that has been used to evaluate the final set of technologies developed in the ROXANNE project. Compared to previous deliverables on data sets, D4.1<sup>1</sup> and D4.2<sup>2</sup>, some data sets have been extended, added or discarded as the requirements of the evaluation data has evolved along with the development of the ROXANNE platform. The data sets described in this deliverable are those that have considered interesting for evaluating the final capabilities of the ROXANNE platform. Several data sets have been provided by LEA partners. However, for legal and ethical reasons, these content of these data sets are constrained and can only be used to evaluate specific aspects of the ROXANNE platform. Therefore, large efforts have been made to collect a data set from a simulated criminal investigation named ROXANNE simulated data (ROXSD). This data set contains all types of information and modalities that can be processed by the ROXANNE platform and it is the main focus of this deliverable.

---

<sup>1</sup> D4.1 Overview and analysis of lawfully intercepted and publicly available data, Deliverable of H2020 project No. 833635 - ROXANNE, December 2019.

<sup>2</sup> D4.2, Simulated data for development and demonstration, Deliverable of H2020 project No. 833635 - ROXANNE, December 2020.



## Table of contents

Disclaimer.....	2
Copyright notice.....	2
Revision history.....	3
Executive summary.....	4
Table of contents.....	5
1 Introduction.....	7
2 ROXANNE Simulated Dataset (ROXSD).....	8
2.1 Motivation and Plan.....	8
2.2 ROXSD Scenario.....	10
2.2.1 ROXSD version 1.0.....	10
2.2.2 ROXSD version 2.0.....	11
2.2.3 ROXSD version 3.0.....	13
2.3 ROXSD Calls Subset.....	16
2.3.1 Data Collection.....	16
2.3.2 Data Size and Statistics.....	18
2.3.3 Metadata.....	22
2.3.3.1 Speaker and Device Metadata.....	22
2.3.3.2 Call Metadata.....	22
2.3.3.3 Transcripts.....	24
2.3.3.4 NLP Annotations.....	25
2.4 ROXSD Video Subset.....	26
2.4.1 Data Collection.....	26
2.4.2 Data Size and Statistics.....	28
2.4.3 Metadata.....	29
2.5 ROXSD Text Subset: ROXHOOD.....	29
2.5.1 Data Collection.....	29
The ROXANNE Forum - ROXHOOD.....	29
The ROXANNE Crawler - Rocraw.....	32
2.5.2 Data and Metadata; Info, Size and Statistics.....	36
2.5.3 Fusion of ROXHOOD with ROXSD.....	36
2.6 ROXSD Release.....	37
2.6.1 Content and structure.....	37
2.6.2 Accessibility and availability.....	38
3 Other Datasets Employed in the Project.....	39
3.1 Real Case Data Corpus.....	39
3.2 FRIDA - Forensically relevant inter-device audio database.....	41
3.3 Case 1 Data.....	41
3.4 Case 2 (Burglary) Data.....	42



3.5	Case 3 data .....	43
4	Ethics in Data Collection and Processing.....	43
5	Summary and Conclusion.....	44
	Annex 1 - ROXSD Transcription Guidelines, v2.2.1 .....	44
	Introduction: What to transcribe.....	44
	Transcription Process.....	45
	Step 1: Segmentation.....	45
	Step 2: Transcription (Left-to-Right Scripts).....	45
	Annex 2 - ROXSD Annotation Guidelines .....	47
	Annex 3 - Informed consent .....	50
	Annex 4 - License.....	56

# 1 Introduction

ROXANNE aims to enhance criminal investigations through the extraction of information from multiple modalities of data obtained in the criminal investigation. In order to evaluate the technologies developed to this end, realistic data is indispensable. In addition, the core technologies of the ROXANNE solution such as speaker recognition, speech recognition, natural language processing, video processing and network analysis rely heavily on machine learning approaches in which large amounts of data are required to accurately train the associated models. Standard public data sets have been used to train the individual models of the ROXANNE technologies<sup>3</sup>. These are, however, not rich enough for evaluating the multimodal nature and the for fulfilling the practical objectives of the ROXANNE solution.

The ROXANNE project was originally planned to operate on data from the following sources:

- lawfully intercepted communications (wire-tap recordings) and videos of criminal network members, accessed by the ROXANNE end-users (LEAs),
- data collected from open sources (i.e., web and social media platforms such as Facebook, Twitter, Instagram and YouTube) which are relevant to criminal cases,
- data collected and/or produced by other research projects.

An initial list of sources which were considered to be relevant for ROXANNE at the beginning of the project was compiled and presented in D4.1. This list contained publicly available datasets as well as some lawfully intercepted data which could be used for scientific and evaluation purposes. Three of such datasets (CSI, NIST and ENRON) were then selected for common use to evaluate and demonstrate the individual technologies. The Consortium also had a mutual agreement with the sister project LOCARD to exploit their online grooming dataset. Key properties of these datasets<sup>4</sup> together with the modalities they contain and the technologies they can support are summarized in Table 1.

Dataset		CSI	NIST	ENRON	LOCARD
Domain		TV	Public	Crime	Crime
Language		EN	EN	EN	Multi
Modalities	Audio	x	x		
	Transcripts	x			x
	Text			x	
	Image/Video	x			
	Spkr/Author Labels		x	x	x
Technologies	Speaker Ver/Cluster	x	x		
	Speech Recognition	x			
	Named Entity Recognition	x			x
	Topic Detection				x
	Network Analysis	x	x		
	Image Similarity	x			

Table 1. Key properties of the initially selected datasets

<sup>3</sup> These data sets were well described in the related previous deliverables with references to the literature, therefore they are not described in this deliverable.

<sup>4</sup> Detailed information about these common datasets were presented in deliverables D1.3 Data Management Plan (Month 6), D1.4 Internal Progress and Quality Planning Report and D5.1 Initial Speech/Text/Video Technologies, as well as in Ethics Deliverables D10.8 and D10.10.



Although these datasets served as a good starting point for adaptation and evaluation of models in the early phase of the project, they were unsuitable for the greater goals of the project, i.e., the interplay of network and speech analysis, therefore could only be used to demonstrate a limited number of technologies involved. Besides, the data collected from open sources brought with some legal and ethical issues, and were relatively far from the LEAs' use-cases of interest. Lawfully intercepted data were impossible to be transferred from the Project's LEA partners to the technical partners, due to legal, privacy and operational reasons. The LEAs were only able to provide some real data to the consortium partners in limited form.

The Consortium has therefore decided to prioritize the work defined in Task 4.6, which deals with the creation of a simulated dataset of communication in organized crime for development and demonstration activities. This dataset was called the ROXANNE Simulated Dataset (ROXSD, in short). Early design phases and the first round of data collection efforts were explained in the deliverable D4.2. After the submission of D4.2, two special workshops were held to collect more data emphasizing on multimodality and multilinguality, and special attention was given to provide the ground-truth annotations of all modalities involved. The ROXSD has become the main evaluation and demonstration dataset in the project by covering all modalities and technologies listed in Table 1, and were used in the three ROXANNE field tests conducted between 2020-2022. ROXSD is one of the two major tangible outcomes of the ROXANNE project, besides the ROXANNE (Autocrime) platform.

In Section 2 of this document we explain the ROXSD in detail, from the idea and scenario creation to data collection to ground-truth annotation. Then, in Section 3, we talk about the other datasets which were provided by the LEA partners for their use-cases. All of the data collection and processing in the project were handled within legal and ethical norms and under supervision, which we briefly explain in Section 4. Section 5 concludes this deliverable with a summary and comments. Notice that this deliverable focus on datasets used for evaluation of the ROXANNE platform. Standard public datasets used for training core technologies are not described since they are well described in the literature. The relevant references for these technologies are provided in their descriptions in WP5 and WP6 deliverables. A full list of datasets used in the project is available to consortium members in D1.3, D1.7 and D1.8<sup>5</sup>.

## 2 ROXANNE Simulated Dataset (ROXSD)

### 2.1 Motivation and Plan

As introduced in Section 1, the need for collecting a dedicated dataset came from the necessity to test and evaluate all of the individual ROXANNE technologies as well as the ROXANNE platform, and to be able to demonstrate the results and outcomes in a criminal investigation scenario without having to deal with legal and ethical issues. The Consortium has considered it essential for the dataset to have the following three features:

- **Realistic:** The data must resemble what is collected in a real investigation as much as possible, both qualitatively and contextually.
- **Multimodal:** The data must contain multiple data types (modalities) including audio, images/video and text.
- **Multilingual:** The communication between individuals must be carried out in multiple languages<sup>6</sup>.

The work towards preparing a screenplay for the ROXSD began in Spring 2020. One of the LEA partners, the National Drug Headquarters of the Police of the Czech Republic (PCR) came up with an original scenario draft, based on their professional experience and without uncovering real cases. This draft scenario did not exactly match

---

<sup>5</sup> D1.3, D1.7, D1.8, [ROXANNE's data management plan](#), Deliverable of H2020 project No. 833635 - ROXANNE, February 2020, February, 2021, December 2022.

<sup>6</sup> The minimum required set of languages were defined to be English, German, Russian and Czech.





any of the real cases, but was inspired by them<sup>7</sup>. The main characters and roles in the story, their motivation and relationships had to be defined in more details by a working group composed of some of the WP4 technical partners.

The responsible partners then asked the Consortium members and their colleagues to volunteer in the data collection process. The participants were selected with the goal in mind to create groups (cliques) who would speak a particular language among themselves. Real personal data were collected only for administrative reasons (consent forms and data for speakers who were financially compensated). For the data recordings, only **simulated (fake) personal data** were used. The coordinating partner (Phonexia) asked the participants to suggest fake personal data for the characters which they role-played, roughly matching their own actual country, region and native language.

50 volunteers took part in the first round of data collection. The participants were given a summary of the conversation and a number of keywords to talk about, but they were free to structure the conversation. Some partners chose to prepare the conversation script in advance and stick to it in the recording, while others chose to improvise.

In the meantime, the recording hardware and software were set up. For the first round of the recordings, the Twilio software was used as the recording platform, hosted on Phonexia's servers. The platform allows telephony connection between participants and recording of stereo audio file of each call. The setting up of the acquisition environment and its testing were conducted between December 2019 and February 2020. The first phase of recordings were split into two sections: target and non-target calls. The acquisition of both call types took place between June-December 2020<sup>8</sup>. A total of 236 calls were collected. This version of the dataset was named the ROXSD v1.0.

With the inclusion of ZITiS in ROXANNE, the Consortium had the chance to use the real interception equipment, and exploited this opportunity to extend the dataset. A new volunteering process began, and based on the availability of speakers, the scenario was extended with a suitable follow-up story. The main focus of the preparation and recording was on metadata acquisition. A two-day workshop at ZITiS was held in November 2021 to collect more data. Despite the challenge of recording all audio and CDR data in a short amount of time, and several technical issues and organizational obstacles which had to be dealt with during the data acquisition process, the two-days of participants' intense work brought a valuable extension to ROXSD. In the months following the ZITiS workshop it became obvious that cleaning and further processing of metadata to prepare the ground truth information is a very difficult and time consuming task, especially that all metadata, which were based on a fictional story, had to be coherent and consistent. At this stage the ROXSD was extended to 345 audio recordings collected from 83 speakers. This version of ROXSD was named the v2.0.

Finally in May 2022, the partners gathered again at ZITiS for a second, three-day workshop to extend the story even further. This time, video, text and geo-location data were also collected. The number of recordings and speakers in the dataset increased to 481 and 103, respectively. The final version was named the ROXSD v3.0.

Besides the scenario building and data collection, a great amount of effort was also spent towards preparing the metadata information and the ground-truth annotations. This included, for instance, determining the speaker labels, manually diarizing the audio, manually transcribing the conversations, annotating the named entities and topics, etc.. A particular attention was paid to the ethical and legal issues concerning the data collection. A description of the process and informed consent were elaborated in cooperation with the project partner Trilateral (UK), who is handling the ethical and legal issues in the project (see Annex 3 - Informed consent ).

As the dataset progressed and the structure of the technological components as well as the user's needs and goals shaped, the partners had to change the structure and formatting of these metadata labels and annotations a couple of times, finally reaching to a common understanding in M39. The latest version of the ROXSD Metadata set is versioned as v3.2.0.

In the next subsection, 2.2, we first explain the ROXSD scenario and its evolution during the three phases of data collection. This is followed by details about the three main components of ROXSD, the calls subset (2.3), the video

---

<sup>7</sup> To get an insight about the range of criminal activities dealt with by NPC, an interested reader is encouraged to read their public annual reports, available in Czech and English: <https://www.policie.cz/clanek/vyrocnizpravy-annual-reports-jahresbericht.aspx>

<sup>8</sup> Some of the target calls had to be re-made by replacing the characters with new voice actors, as the terms of data use had to be changed after the data collection, and some volunteers did not accept the new terms. Their data were deleted from the collection.



subset (2.4) and the text subset (2.5). Finally in 2.6, we talk about ROXSD as an outcome of the project, touching upon its structure and the distribution plans.

## 2.2 ROXSD Scenario

The ROXSD story is built on a drug dealing case, in which a group of criminals communicate with each other over the telephone. Their communications are intercepted (wire-tapped) by several (fictional) police organisations. The wire-tapped data includes also a number of “innocent” people communicating with the criminals and with each other.

The offenders speak in Czech internally (planning local criminal activities) and (possibly bad and heavily accented) English when planning trans-national activities. The scenario also includes other international criminal groups which speak their native language such as German and Russian.

In the following subsections, we present the original ROXSD story and its evolution through the two additional workshops.

### 2.2.1 ROXSD version 1.0

The original scenario evolved into three fictional interconnected cases:

- Drug Distribution A (Case DDA)
- Drug Lab A (Case DLA)
- Drug Distribution B (Case DDB)

#### *Case DDA*

In Case DDA, Kryštof, a university student in Prague, is suspected of selling drugs. The police have wire-tapped two of his mobile telephones. Kryštof’s movements are mostly within Prague with some travels to Brno. The wiretaps have shown that Kryštof is in contact with other individuals who are either users or distributors of drugs. Communication is in Czech or Slovak. Most of the communication occurs at the point where the drugs change hands.

The police also start a wiretap on the mobile phones of his contacts Kristýna and Horký (Kuba), who are regular drug takers, and also occasional distributors working for Kryštof.

Kryštof is often in touch with Sergej (in English), his schoolmate and main drug dealer, with whom they have also agreed on some larger transactions. As a consequence, Sergej’s telephone is also wiretapped. Most of Sergej’s conversations are in Russian. During one call between Sergej and Oleg, also Kryštof’s schoolmate, the tap caught information about contact between Oleg and Kryštof. That is the reason why the police wire-tap also Oleg’s phone. One conversation mentions a delivery of drugs to London.

In the meantime, Kristýna buys a new SIM card. Being notified about this, the police immediately start intercepting also this new number.

#### *Case DLA*

In Case DLA, the police suspect that two Vietnamese, Tuán and Hoàng are dealing in large quantities of drugs and that Hoàng may have a production site. Their telephones are wire-tapped. They mostly speak Vietnamese and call each other several times. They are also often in touch with another two Vietnamese contacts.

In the meantime, Tuán starts communicating with an unknown person in English, planning a large delivery of drugs.

#### *Case DDB*

In Case DDB, the police investigate Max, an Austrian student at a university in Prague suspected of the distribution of drugs in Prague city centre. Max’s telephone is wire-tapped, and he is in contact with several unknown individuals. He speaks German and English.

The investigation of these three cases uncovers a connection among the criminals and shows that all the cases are connected. Kryštof appears to be at the center of the whole network. He resells drugs to foreign dealers and his source of drugs are Tuấn and Hoàng.

Figure 1 depicts the structure of the criminal network based on the communication between individuals<sup>9</sup>. The red lines represent the (“target”) calls made within the criminal group and the green lines represent the (“non-target”) calls between speakers who are either not suspects or calls which do not contain valuable information for an investigator in the simulated case. The direction of the arrow shows who called whom: from the caller (initiator) to the receiver. The number on a line shows how many calls were made by that pair of speakers. The nodes of the graph are speakers with attached speaker label in the investigation case and the name used in the fictional case.

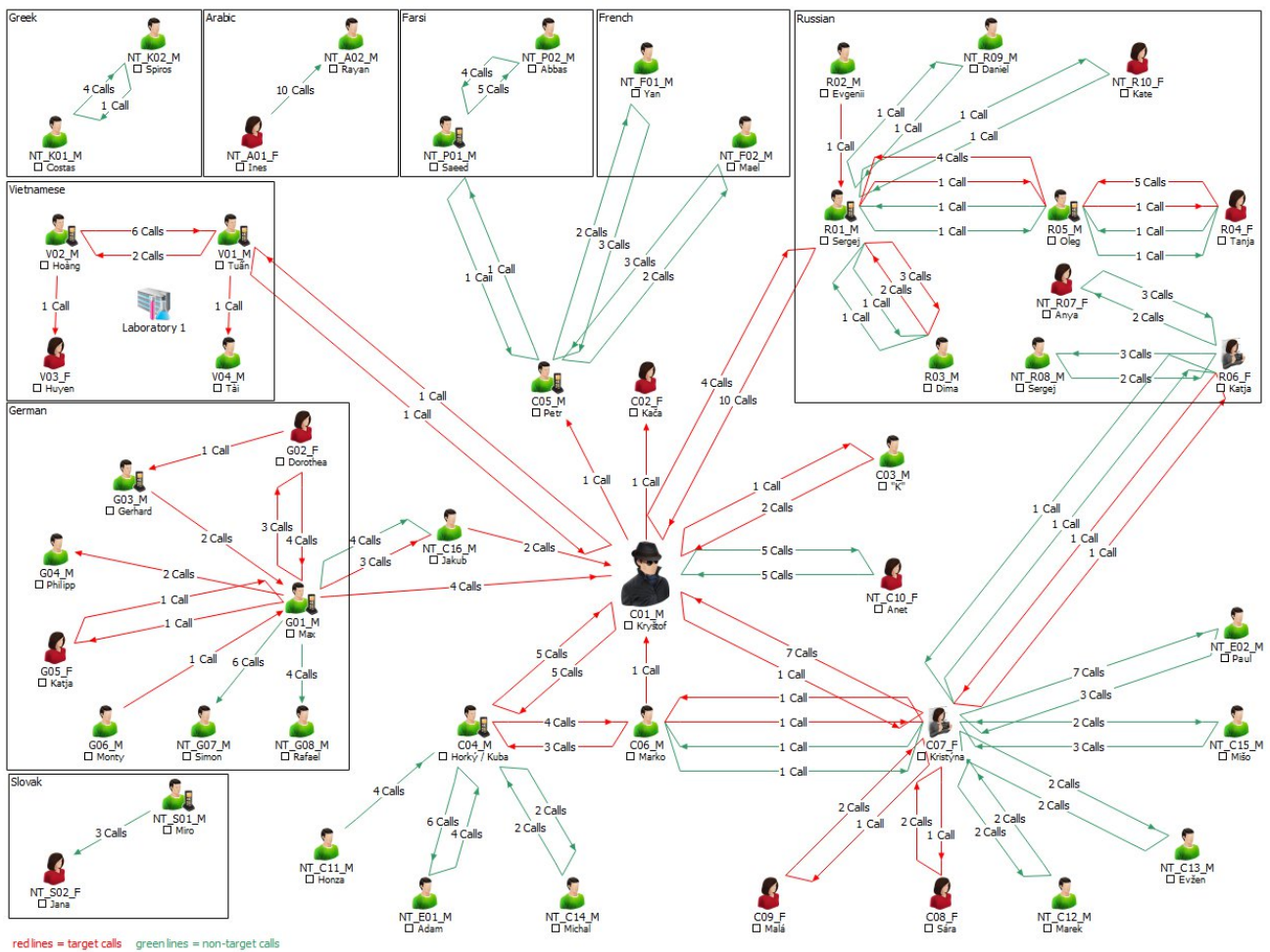


Figure 1. Structure of target and non-target calls in the ROXSD v1.0 network

## 2.2.2 ROXSD version 2.0

ROXSDv2 story evolves from and contains the ROXSDv1.

### Case DDG

<sup>9</sup> This network depiction was manually created at Brno University of Technology with the i2 Analyst’s Notebook tool.



The links between the three cases are uncovered during the individual investigations. Then the police continues the investigation by fusing the three cases, and realize that over time, the criminal activities spread across borders to Austria and Germany.

Kryštof becomes addicted to methamphetamine and after a few months of investigation, he dies from an overdose.

Kristina used to be Krystof's close friend and has many connections who were used by Kryštof for drug distribution and production. But some of Kristina's contacts are lost, for example, Horky (Kuba), who is in anti-drug treatment.

Kristýna was hiding some of the narcotics for Kryštof in a safe place. When Kryštof diseased, she decides to make some money by selling drugs. She decides to organize a drug business with Marko, who should have additional contacts and is often travelling (in Czechia and abroad, too). Marko becomes a closely monitored person based on a previous investigation and his telephone activities.

Kristýna starts thinking about selling narcotics abroad because it seems to her like a more profitable option than selling in Czechia. The Russian group and their connections to London seem to be lost because Sergej was not in touch with Kristýna, and Marko and Kristýna are not successful in reaching another colleague of theirs, Katja.

Marko and Kristýna recall that Jakub had some connection with people (Max) in Germany and ask him to give these contacts to them. Kristýna exchanges text messages with Max. Max had left Prague after finishing his studies at the university and he is returning to Vienna. Kristýna gets from Max the contact information to Gerhard. Gerhard seems to be attracted by the possibility of making money by selling narcotics and getting back to business after Kryštof died. Kristýna decides to try to enter the business of drug distribution at parties in bigger German cities. Marko and Gerhard prepare a plan for a meeting in Munich.

Marko joins Slava on the way to Germany. They are friends from studies. Marko calls Kristýna (on her new intercepted number) from his new telephone number and lets her know that he will mostly use that number when making calls from Germany. Consequently, the Czech Police ask the German Police to intercept all calls from Marko's new German telephone number.

Marko has multiple intercepted calls with Gerhard. The German Police therefore start intercepting Gerhard's number.

Slava participates in a two-day conference/workshop at a hotel in Munich. He takes a small amount of narcotics from Marko and gives them to his friend at the workshop.

The telephone number of the hotel in Munich where the conference is taking place is added to the list of intercepted numbers when Marko calls Slava on this number and speak about suspicious stuff. Otherwise, the calls from and to this number are mostly non-target calls.

Gerhard meets Marko in Munich and they visit a big party where Gerhard buys all drugs from Marko. Their meeting point is a beer garden/brewery ("Bräuhaus"), where another of Gerhard's contacts known as Dorothea has a part-time job.

The number of the Bräuhaus is immediately added to the intercepted numbers. Recordings from this number are mostly target calls. Gerhard calls Dorothea on this number.

Everything works well for them, and they decide to proceed in the same manner. Marko buys drugs from the Vietnamese lab and then has Andi to transport them from Czechia to Germany. Gerhard takes over drugs from Andi and then passes them on to his dealers.

By the end of Marko's stay in Munich, he loses the telephone which he was using in Germany.

A suspiciously behaving young man is arrested two days after the big party. He is under the influence of narcotics. The German Police confiscate a cell phone from the young man and the content is extracted (pictures, text messages, videos). The pictures are from Marko. The phone had been stolen from him.

Kristýna and Marko have to argue with the Vietnamese producers of drugs in Czechia. Because of that, they start looking for other channels of obtaining drugs. They register themselves and start messaging with people in the darknet "drug" forum called "ROXHOOD".

### 2.2.3 ROXSD version 3.0

*ROXSDv3 story evolves from and contains the ROXSDv1 and ROXSDv2.*

Narcotics business and transfer to foreign countries (Austria, Germany, France) are still doing well for Kristýna and Marko, but due to the war in Ukraine, there may be possible complications such as border controls. Kristýna, Marko, and Samuel (an contact of Kristýna's from France) are trying to solve possible problems.

Marko and Slava travel together to meet Gerhard again in Munich to check the status of the business. Kristýna experiences a car failure on her way to Munich. She comes to Munich a bit later than others.

Marko, Kristýna and Gerhard want to push drugs to the planned dance party (Rave party in Munich 10 May 2022 at 19:00, Bräuhaus) which is going to happen during Kristýna and Marko's stay. The dance party is organized in the bar where the employees are also involved in the business with narcotics. A bar employee, Christiana, is calling their friends and drug users with an invitation to the dance party. Christiana is connected to the German and Czech criminal groups through Dorothea, who is also employed at the bar. Information about the party is also shared via the ROXHOOD forum with potential drug users.

Once they get together, Samuel opens the topic of drug transport to France. Samuel wants to involve the “white horse” in the transport and sale of drugs but he is concerned about his safety due to the intense controls at the borders and the general situation in Europe.

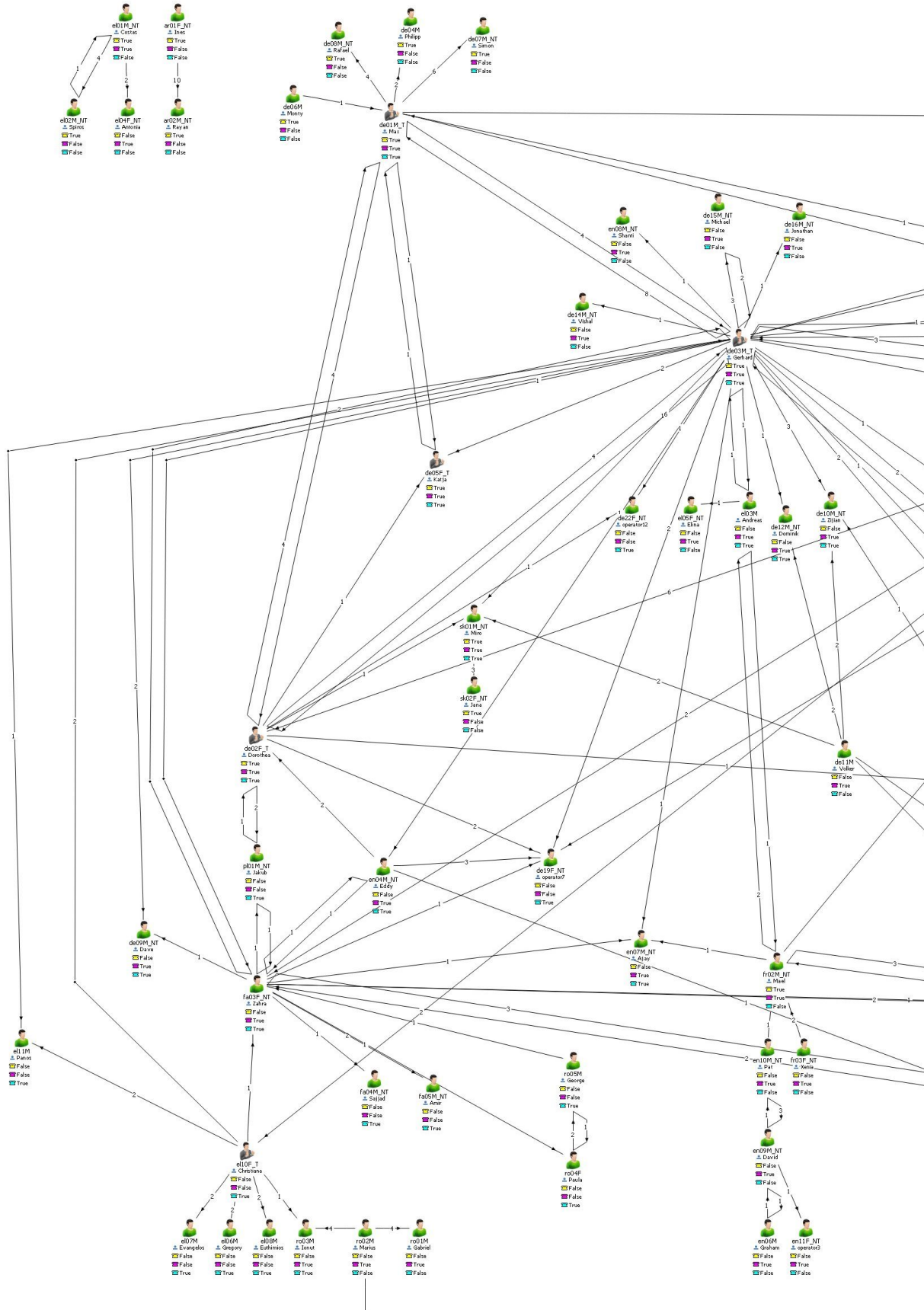
Samuel brings up another problem, caused by the damaged packaging of the drugs delivered by Andi. At the same time, Andi wants more money for his work because of the higher prices of petrol.

The dance party in Munich is a success and all the narcotics are sold to users at the party.

An overview of the ROXSD network as per v3.0 is given in Figure 2 on the next two pages.



# ROXANNE | D4.3 Final Report on ROXANNE Data



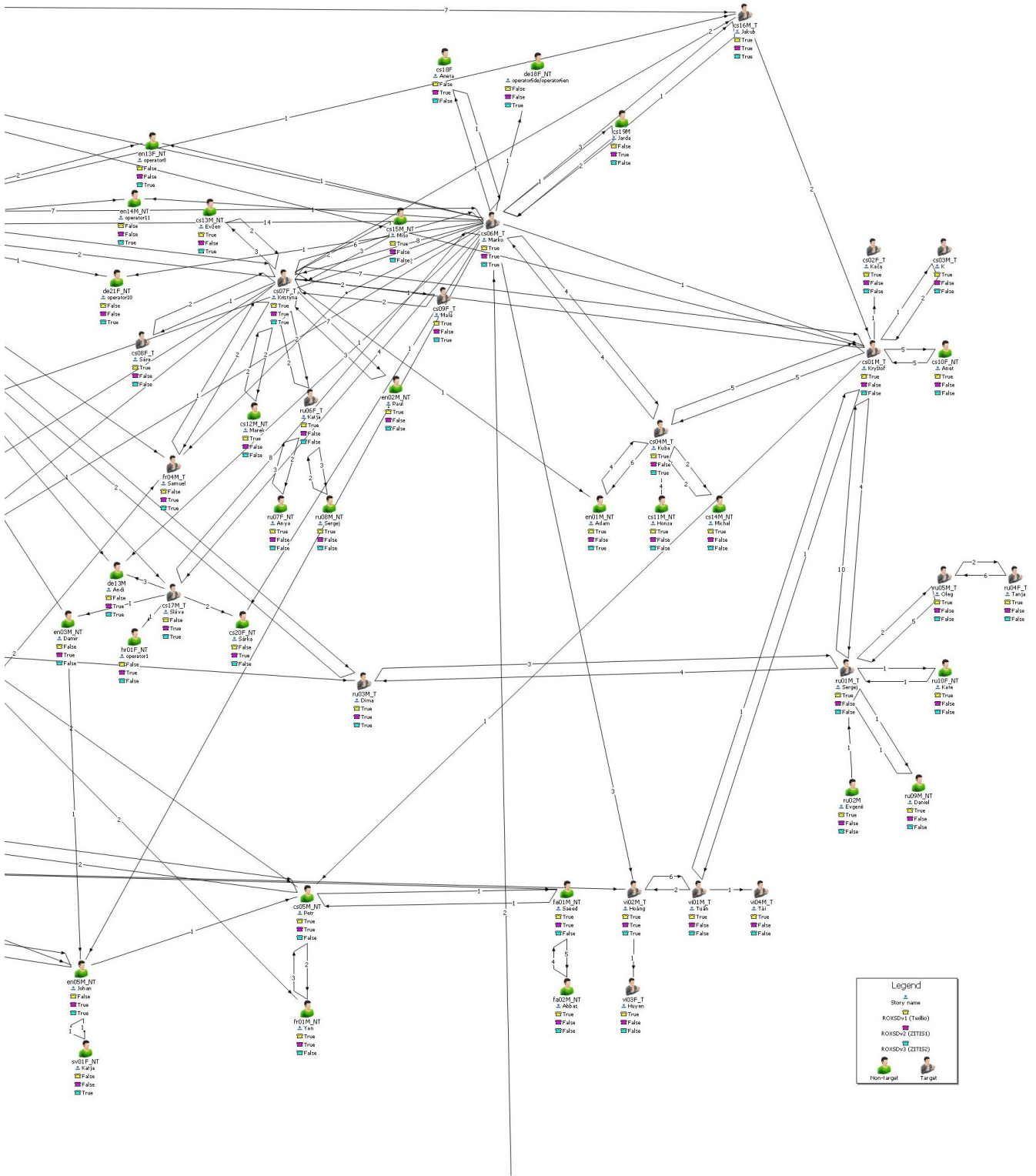


Figure 2. Structure of the ROXSD v3.0 network

## 2.3 ROXSD Calls Subset

The ROXSD calls subset is one of three essential data parts of the dataset. It contains the audio recordings of the telephone conversations made between the members of the fictional criminal network, and between the criminals and some innocent (*non-target*) persons. The topic of each call has been manually scripted and the calls were realized by voice actors (volunteers) who role-played these characters.

The structure of the ROXSD calls subset and its metadata information allows for the processing and evaluation of many audio and text related technologies in ROXANNE. Speaker identification applied on the audio files is used for detecting persons of interest in the intercepted calls. This information is then used by network analysis and link prediction. Language identification applied on speech utterances to detect the language of the conversation, which helps selecting the right language model for automatic speech recognition. The transcriptions of calls are subsequently used by several technologies like named entity detection, co-reference resolution and topic detection.

### 2.3.1 Data Collection

The acquisition of the simulated data and metadata was initially planned to be realized by using the production equipment of one of the LEA partners. However, the LEAs were not able to provide such support due to their security restrictions. Thus, it was decided by the consortium to have Phonexia to provide *their* recording equipment which resembled the profile of a LEA equipment. In the second year of the project, a new partner, ZITiS, entered the project and offered their interception equipment for use in audio recording and metadata acquisition. This was a big step forward in the authenticity of formatting and quality of audio recordings and metadata content. The call data records (CDR) which were provided together with audio files significantly enriched the content of metadata by new parameters.

#### *Twilio data collection*

The calls were recorded as closely as possible to actual data acquisition from real criminal cases. All participants used their phones (mobile or landline) that they use for normal communication. Real phone numbers are recorded as original meta-data of the calls, but pseudonymized, so that real telephone numbers are, under no circumstances, part of ROXSD.

The Twilio Application was used as the interface for the initiation of the calls. Calls were made via the Twilio call centre platform. The platform connects the specified telephone numbers and stores call recordings and metadata for one month. In order to initiate a call, a representative enters the two phone numbers (caller and receiver) on the Twilio portal and clicks on the call button. The caller then receives a call from the Twilio server. As soon as the caller picks up, the recording begins and the receiver's phone is called.

The calls and metadata were downloaded by the dedicated staff at Phonexia with authenticated access to the Twilio Application and the Twilio Server. The downloaded data were temporarily stored at Phonexia in a secured data storage and then transferred to BUT for further data cleaning and pseudonymization, which were also conducted by dedicated project personnel. ROXSD is kept at BUT FIT secured data storage and access is provided to ROXANNE partners upon request. Figure 3 shows the setup graphically.



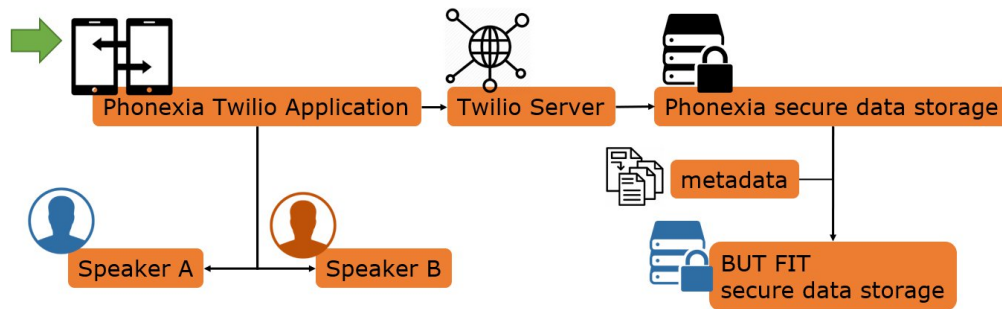


Figure 3. Recording and data-transfer setup

### ZITiS workshop data collection

The consortium partner ZITiS offered their setup for further audio file collection. Initially planned as a single data collection event, the efforts were then extended with a second workshop, based on the success of the first one and the ongoing request for a larger and richer ROXSD dataset. The two workshops were organized on 3-4 November 2021 and on 10-12 May 2022, respectively. These dates were dedicated to data creation.

Both workshops were preceded by organizational preparations. It was necessary to select volunteers who will represent the characters appearing in the scenario. It was also necessary to prepare the topics of the recorded phone calls and, due to the limited time intended for recording, define the time slots when the individual calls will take place. After the recording, the handover of the recordings and CDR data to the consortium was ensured. The consortium could then start processing the audio files and converting the CDR data to metadata in a format that was suitable for use by the developed technologies.

The organization was mainly ensured by three partners ZITiS, HENSOLDT Analytics and Phonexia. The workshops took place at the ZITiS premises, which was a necessary condition for the realization of the recording, connected with the provision of telephone devices, an interception system and the possibility of immediate control of the ongoing recording. Suitable spaces were prepared for the recording, where it was possible to carry out the preparations as well as the phone calls themselves. Evidence of the phone calls made was kept, thus ensuring the necessary ground truth for further use of the data. The participants (the majority of which were members of the consortium partners) were given instructions which included the time to start the call, the content of the call and the parties who would make the call. Most of the recording participants were physically present to enable efficient organization. However, some roles in the story did not require people to be physically involved in the recording, so it was ensured that these people could connect remotely.

Two cell phones and one SIP phone were put under interception. The wire-tapped phones were used by two volunteers who were the main suspects in the story. They called other characters who are suspected of criminal activity and also innocent people. The SIP phone was used as a shared device by several people, some of whom were also investigated as the target persons. The rest of the participants instead used their own telephone devices. Their telephone numbers were then anonymized.

In the framework of the two workshops, the content of the recorded phone calls was predetermined. Each of the participants in the phone call was informed in advance of the intended content of the phone call, but the detailed scenario was not presented to the callers, with some exceptions. The goal was to record phone calls that would be as close as possible to a natural conversation between two or more people.

The last workshop focused more on phone calls of an exceptional nature, such as multiple callers from one mobile device, acoustic noise in the background of the speaker, changing the spoken language and others. The course of the phone call, the keywords used, the reference to other people, the start of the call and the end of the call were recorded in detail so that they could be used later in the creation of metadata.

### 2.3.2 Data Size and Statistics

In its latest version v3.0, the ROXSD calls subset contains 432 intercepted telephone conversations recorded into 481 audio files, encoded in 8kHz, 16-bit, stereo<sup>10</sup> wave format.

The dataset is composed of different types of calls: standard phone calls in which the caller calls the receiver’s telephone number, teleconference calls in which the caller calls a third person while already talking to the receiver, and calls that are made to a web conferencing service (Zoom, Webex) where the callers dial a common telephone number (the service’s dial-in number) in order to talk to each other.

The difference between the number of calls (432) and the number of recordings (481) is due to the fact that some of the calls were intercepted multiple times by different sides of the conversation, which is a consequence of the variety in call types: 270 calls are intercepted only on the caller’s side, 111 are intercepted only on the receiver’s side, and 45 calls on both sides. There is an additional teleconference call which was intercepted a total of 10 times.

This results in some of the recordings being very similar in content. However, they are not an exact copy of each other, because of the following reasons: (i) The interception begins on the caller’s side as soon as the caller finishes dialing the receiver’s telephone number. Hence, the ringing dial tone as well as any sounds/speech which the caller’s phone picks up before the connection is established are captured by the intercepted recording coming from the caller’s side. For the same reason, the receiver’s intercepted recording is a few seconds shorter than that of the caller’s. There are also cases where, although both sides are intercepted, the receiver’s phone is not reachable, therefore there is no recording from the receiver (in such cases, either the receiver’s voice box message or the operator’s out-of-reach message can be heard in the caller’s recording). (ii) For teleconference calls involving three (or more) parties, a new interception is initiated when the caller calls a third (fourth, ...) person in order to connect them into the existing conversation. (iii) For web conferencing where multiple parties call the same (operator) telephone number, each party’s interception begins when they join the conference room. (iv) The audibility of speech in both recordings can be different than each other due to the background or microphone noise introduced by one of the parties, or issues with their interception equipment. These inexact copies of the same phone conversation are intentionally left in the dataset, as these artefacts closely reflect the nature of interception in the real world.

The calls in the dataset are also varied in terms of their connection status: A call is considered to be *successful* if the intercepted recording contains the full conversation between the caller and the receiver. In addition, there are some *failed* calls in which a connection cannot be established between the parties, due to either the caller dialing a wrong number, or the receiver’s phone being unreachable. The audio recordings of failed calls contain little useful speech, mostly an operator or a voicebox message. Finally, a *dropped* call happens when the call is picked up but the connection or recording is cut off before the parties finish their conversation. The latter is mainly caused by technical problems which were encountered during the data collection and interception process, and most of the time only one side of the conversation is audible. The partners decided to keep the failed and dropped calls in the dataset, as such artefacts also represented the real-life scenario of the data acquired in a typical investigation. The distribution of audio files with respect to the call type and the connection status are given in Table 2.

	Standard	Teleconference	Web conference	Total
Successful	418	3	6	427
Failed	17	-	6	23
Dropped	27	3	1	31
Total	462	6	13	481

Table 2. Distribution of recordings across different call types and connection statuses.

<sup>10</sup> 3 files out of the 481 contain only a single channel, due to a temporary issue with the interception equipment at the time of recordings.

The total duration of the 481 audio files is 18h 28min. The shortest call is 0 sec long (failed call) and the longest call is 12min 15sec long. Figure 4 shows the histogram of recordings with respect to their durations. An automatic voice activity detector (VAD) system was used to segment the speech-only parts of the audio. These boundaries were then manually adjusted by human annotators. Based on the outcomes of this analysis, the total duration of speech activity is 19h 34min. This difference comes from the fact that the speakers often interrupt each other in a telephone/teleconference conversation, causing overlapped speech. Part of this difference has also been observed to originate from the speech boundaries not being adjusted precisely (there is a buffer of silence of up to 1 second for each speech segment).

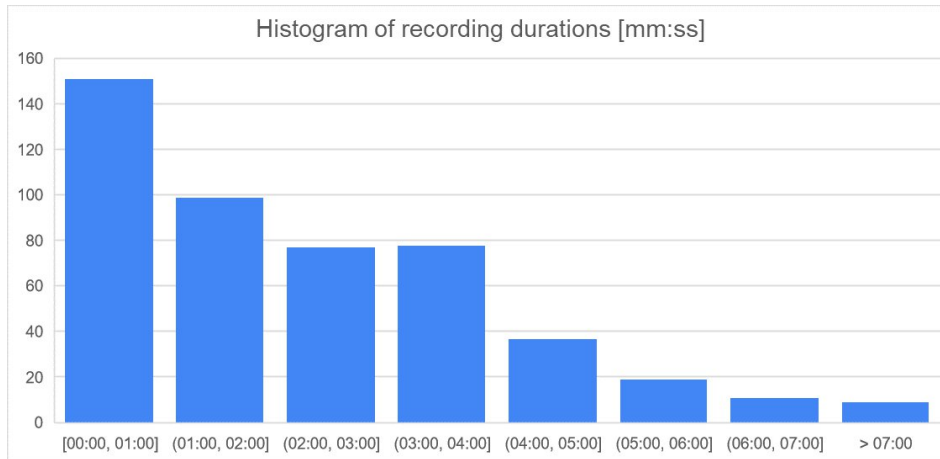


Figure 4. Histogram of recording durations

As already outlined in Section 2.1, there are 103 speakers (characters) in the ROXSD calls subset. 25 of them are target persons, 60 are non-target persons and 18 unknown characters (including operators, etc.). Figure 5 shows the speech activity of the top-15 speakers in the dataset. Apart from the teleconference and web conference calls which contains three or more speakers by nature, there are also calls in the dataset in which the caller or the receiver hands over the phone to another speakers. In some other calls, the entire conversation is carried out by a speaker who is not a regular owner of that particular phone number. Hence, that the speakers in a particular call may not be the regular speakers / owners of the telephone numbers making that call. This is one of the aspects which were introduced to yield a challenging, thus realistic environment.

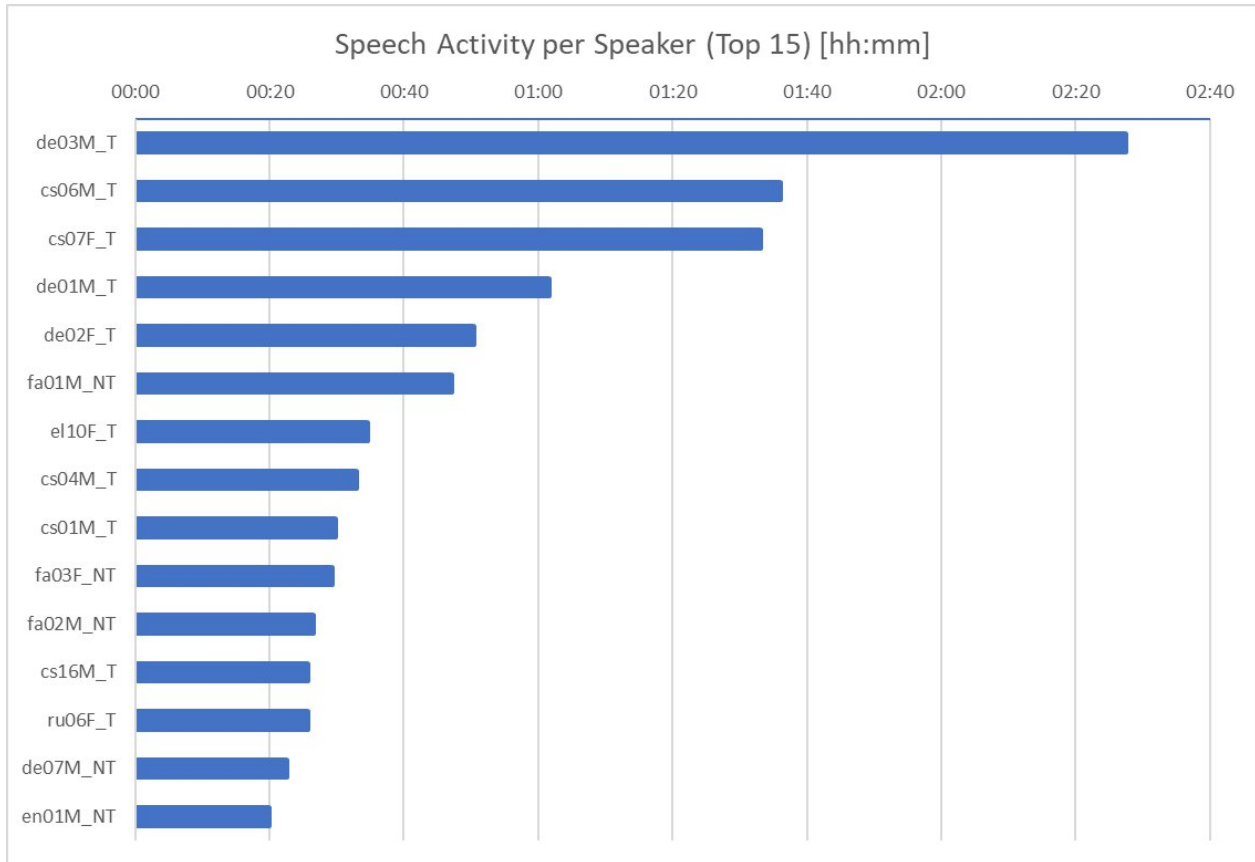


Figure 5. Speech activity per speaker

A total of 15 languages are spoken in the ROXSD calls subset. 13 of them are real conversations, whereas 2 of them are only single phrases or brief sentences. Figure 6 shows the distribution of speech activity across languages in the dataset.

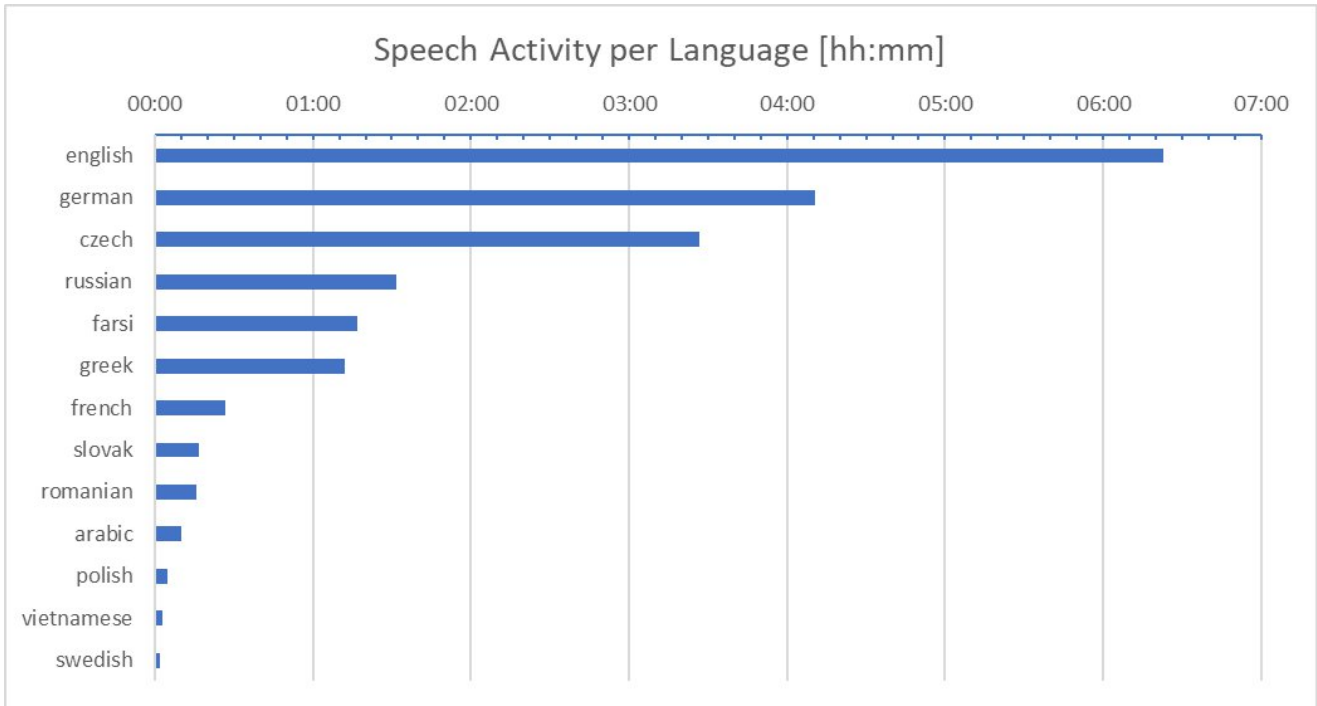


Figure 6. Speech activity per language

ROXSD is multilingual not only on the call level, but also within the calls. A conversation may start in a language and then continue in a different language. The language of conversation may change when the phone is handed over to another person. There are also cases where the speakers switch the language all of a sudden for a brief moment (a few sentences, which mostly is about “business”), and then back again. Whereas Figure 6 presents the total (aggregated) durations on a segment level (which, by definition, consists of only a single language), Table 3 lists the total durations of audio in each language and language combination on the call level.

Language	Duration [hh:mm:ss]	Language	Duration [hh:mm:ss]
english	04:59:44	vietnamese	00:04:26
german	03:31:02	czech;english;french;spanish	00:04:26
czech	03:02:15	english;romanian	00:04:21
russian	01:24:19	english;french	00:03:12
greek	01:06:52	english;german;romanian	00:03:00
farsi	01:04:57	english;russian	00:02:49
english;german	01:00:35	swedish	00:02:39
french	00:29:05	english;polish	00:02:30
czech;english	00:21:10	czech;spanish	00:02:04
slovak	00:14:18	croatian;english	00:01:44
romanian	00:13:25	czech;english;spanish	00:01:41
arabic	00:10:18	czech;german;spanish	00:01:30
czech;english;german	00:08:20	na	00:01:14
english;german;polish	00:05:01	polish	00:01:09

Table 3. Total audio length per language(s)

### 2.3.3 Metadata

The ROXSD calls subset is complemented with an extensive set of ground truth information, which we call the “metadata”. The metadata set can be semantically grouped into several categories such as speaker data, call data, call transcripts, NLP annotations, and other information. The information in the metadata set is anonymized / pseudonymized wherever possible, and in other cases, manually prepared / determined based on the circumstances of the fictional story. In the following subsections we present the categories of metadata, and the application of anonymization.

#### 2.3.3.1 Speaker and Device Metadata

The speaker metadata consists of speaker-related and device-related attributes of the characters (voice actors) in ROXSD. For each speaker, the following metadata are provided:

*Character ID:* A unique integer assigned to each character (speaker)

*Character name:* The character (story) name of the caller / receiver. All participants of the dataset were given the chance to pick a name/nickname for their characters. The character names are not unique; some characters share the same name (as in real world).

*Age range:* The age interval of the speaker in ranges of 5 years

*Gender:* The gender of the speaker/character

*Native language / Second language / Third language:* The native and foreign languages spoken by the speaker. This affects the articulation and determines the accent, when the speaker is speaking a foreign language.

*Character label:* The speaker label of the caller / receiver of the telephone conversation, delimited by a colon (;) in case there are multiple speakers in a single side (channel). Each speaker in the dataset is assigned a unique speaker label, which consists of a language code, a speaker number, gender, and character status. For instance, the label “cs07F\_T” is assigned to the speaker #7 of the Czech language group (“cs”), who is a “F”emale “T”arget speaker. The language code is given mainly based on the language which was spoken by that speaker in his/her first intercepted call, which may be different than that person’s native or most frequently used language.

*Drama number:* The anonymized telephone number of the speaker in full and structured format (including the county and area code). The anonymization is carried out either by changing some digits, or by deleting some digits, or by using a fictitious, reserved telephone number<sup>11</sup>. Please note that a speaker may be using more than one telephone number.

*Drama device type:* The type of device of the Drama\_Number (“mobile” or “landline”).

*Case name:* The name (abbreviation) of the case which this telephone number associated to in the investigation.

*Intercepted number ID:* A fictional codename which is used to assign the intercepted recording to an investigation, which contains the case name, intercepted device number and the location of interception. For instance, the intercepted number ID “DDA\_01\_PG” means that this audio recording was intercepted as part of the investigation case “DDA”, by the device “01” of that case, in “PraGue”.

#### 2.3.3.2 Call Metadata

The call metadata consists of information which could be obtained by the interception system through the telecommunication operator via the Call Data Records (CDR) and the Base Transceiver System (BTS) records. For each audio file, the following metadata are provided:

---

<sup>11</sup> [https://en.wikipedia.org/wiki/Fictitious\\_telephone\\_number](https://en.wikipedia.org/wiki/Fictitious_telephone_number)



*Call ID:* An integer which follows the order in which the audio recordings were collected

*Cluster ID:* An integer that is assigned to audio recordings to indicate that they belong to the same call. This number helps identify the near-duplicates in the audio recordings. Two (or more) recordings (with different Call IDs) may have the same Cluster ID if these belong to the opposite interceptions of the same telephone conversation, or if they are repeated calls with exact or very similar content. The repeated calls which have different content are not given a different Cluster ID.

*Intercepted number ID:* A fictional codename which is used to assign the intercepted recording to an investigation, same as in Section 2.3.3.1.

*Audio file:* The name of the audio (.wav) file with the extension.

*Filename:* The name of the audio file without the extension.

*Caller number / Receiver number:* The telephone number of the caller / receiver of the telephone conversation in full and structured format (including the county and area codes). All telephone numbers used in the dataset are anonymized.

*Caller number Police / Receiver number Police:* The telephone number of the caller / receiver of the telephone conversation as recorded by the interception equipment (hence, as seen by the police)<sup>12</sup>.

*Caller IMEI / Receiver IMEI:* The IMEI number of the device used by caller / receiver to make the call.

*Caller label / Receiver label:* The speaker label of the caller / receiver, same as in Section 2.3.3.1.

*Caller / Receiver:* The character (story) name of the caller / receiver, same as in Section 2.3.3.1.

*Caller channel:* The label (“L”/”R”) which denotes on which side of the conversation (i.e., on which channel of the stereo audio file) can the caller’s voice be heard. The standard structure of an audio telephony data has the caller on the “L”eft channel and the receiver on the “R”ight channel. Due to temporary issues with the interception mechanism, in some of the audio recordings, the caller was assigned to the right channel. These files were intentionally not corrected, as this presented a real situation in which the investigators encounter on a daily basis.

*Story date / Story time:* The fictional date (YYYY-MM-DD) and time (hh:mm:ss) at which the call (interception) started.

*Audio duration:* The duration (hh:mm:ss) of the audio file.

*Languages:* The language(s) spoken in the telephone conversation. If there are multiple languages involved, these are alphabetically sorted and separated by a comma (;).

*Connected:* The attribute which shows the status of the call in the audio recording, with the following values:

- Yes: The caller is successfully connected to the caller.
- Yes but not speaking: The connection is successfully set up, but the receiver’s voice is not heard in the recording.
- Yes but only to conference operator: In a conference call, the connection is set up with the conference room’s common telephone number, however only the conference operator’s voice is heard on the receiver side (i.e., the intended receivers cannot be reached).
- No: A proper connection with the receiver cannot be set up, due to line errors, wrong number being dialled, noise on the line, or receiver not picking up (in which case, the operator’s automatic message or the voicebox message of the receiver can be heard).

*Standard call:* The attribute which shows the details of the call status, with the following values:

---

<sup>12</sup> We are not able to present the structure of such data as this is a public deliverable.

- **Yes:** A standard call in which the caller successfully connects to the receiver, and both sides hang up the phone at the end of their conversation.
- **conf:** A teleconference call in which the caller calls a second person (receiver) in the middle of a conversation with the first one, joining him/her into the call.
- **Interrupted:** The conversation gets interrupted for some reason, and the recording is finished.
- **No:** Not a standard call, due to one side not being heard, noise, echoing, etc.
- **No, no speech from receiver:** Not a standard call as the receiver's voice is not heard in the recording.
- **Voice message:** An automatic voicebox message can be heard on the receiver's side.
- **webex:** A conference call (connected or not connected) made on the Webex platform.
- **zoom:** A conference call made on the Zoom platform.

*Include in demo:* An attribute with “Yes” and various “No” values, which is used to filter only the “healthy”/“ideal” calls.

*Caller Intercepted / Receiver Intercepted:* An attribute with “yes”/“no” values which shows whether the caller and/or the receiver is intercepted in that call.

*Enroll\_LEFT / Enroll\_RIGHT:* An attribute with “yes”/“no” values which shows whether the speaker's voice in the left / right channels of this audio file will be used a sample to train the speaker's voiceprints.

*CALLER\_START\_\* / RECEIVER\_START\_\*:* A series of values which contains fictional BTS (Base Transceiver Station) information for the intercepted sides of the conversation, with the following extensions:

- **ADDR:** Approximate location including (wherever possible) the street, district, city, country names
- **BTS\_LAT:** Latitude written in the decimal degree system
- **BTS\_LON:** Longitude written in the decimal degree system
- **BTS\_AZIMUTH:** Azimuth value
- **BTS\_MMC:** Mobile Country Code
- **BTS\_MNC:** Mobile Network Code
- **BTS\_REGION:** Location Area Code
- **BTS\_CELLID:** Cell ID
- **BTS\_ENBID:** eNodeB-IDentifier

### 2.3.3.3 Transcripts

This category contains the manual transcriptions of the conversations in each call (audio recording), and for the non-English conversations, their translations into English. As of the writing of this deliverable, 94% of all calls in ROXSD are transcribed.

The transcription process was planned from the very beginning of ROXSD, in which the aim was to prepare a written, searchable source of the spoken conversation. The first version of ROXSD (ROXSDv1) contained a simple, text-based transcription format which contained only the speaker labels. Due to the multilingual nature of the dataset the audio files were transcribed by different annotators, and there were inconsistencies in the markup across languages and even across the same language processed by different transcribers.

As the project progressed, the need for a more structured and unified set of transcriptions emerged, also to help the technical partners evaluate their speech and language related technologies. This was achieved by the project partners through a diligent work, through a series of steps and standardization discussions. The flow of data has been the following:

- The stereo audio recordings are split into single (mono) channels
- Each channel is sent through an automatic voice activity detection (VAD) system, which marks the segments containing speech, with its begin-end timestamps (in milliseconds)



- This initial segmentation information is initially processed by the partner coordinating the transcription efforts (HENS), who checks the validity, and adds speaker and language labels to each detected segment.
- The audio files are assigned to volunteered transcribers, based on the language which is spoken in those conversations. The audio files together with the segmentation information, are sent to the transcriber. The transcriber is asked to check and correct the segment boundaries, speaker label and the language name, and to transcribe the speech parts. If the audio contains speech in more than one language, the remaining segments are then forwarded to a second (third, ...) annotator with the relevant language skills.
- The transcription is made using a standard software (Audacity<sup>13</sup>) and by applying a set of pre-defined rules, selected by the responsible project partners after a series of discussions. The transcription rules and how to use the transcription software were defined in a report and shared with the transcribers through some tutorial sessions. This report is provided in Annex 1 - ROXSD Transcription Guidelines. The outputs of transcription are saved in a text file (with a file extension \*.lab — hence called the “lab” file) and sent back to the coordinator.
- The coordinator then redistributes these transcriptions to a smaller set of “experts” who then do the validity and quality checks. Any corrections required are either handled by the experts or sent back to the transcriber.
- Once the transcriptions are finished, the two channels are combined together to obtain a single lab file per call, as shown in Figure 7.

0.000000	0.471126	B:ru01M:english:yes?
1.643194	2.642899	A:cs01M:english:where are you?
3.677077	5.435180	B:ru01M:english:I've already gone home.
6.147613	9.135239	A:cs01M:english:what? we have to work. where are you going?
9.411019	15.753978	B:ru01M:english:I'm going to [Kaserak] but I'm not alone. he leaves and gives me weed. I was next to [Hlavák]. two minutes ago.
16.236594	16.937537	A:cs01M:english:okay.

Figure 7. Lab file structure for an example conversation

Once the transcripts are ready, the non-English text is passed through an automatic translation engine. The outcome of the automatic translation is manually corrected, whenever required. A similar lab file is then generated with the English content.

### 2.3.3.4 NLP Annotations

A similar process was carried out to obtain the annotations with respect to natural language processing (NLP) tasks, namely, the named entity recognition (NER), co-reference resolution (“mentions”), and topic detection (TD). These ground truth information are available for the English and German calls.

*Named-Entity Recognition:* All person names, location names and times that are mentioned in the conversation are annotated.

*Co-reference Resolution:* This task involves differentiating the annotated names by assigning them to one of the parties (caller or receiver), or a third person which is mentioned in the call. This information helps the investigator understand the nicknames of the suspects being intercepted, and their connections to other unknown persons / suspects.

*Topic Detection:* Each conversation is assigned to one of the following topics: Drugs, Work Conversations, Family-Friend Conversations, Money, Meeting, Other.

<sup>13</sup> <https://www.audacityteam.org>. Some of the transcribers preferred to use different software for transcription (Transcriber, Praat, ...), whose outputs were then converted into the format of Audacity.

The NLP annotations together with the transcription output are represented in a \*.yaml file, as shown in Figure 8. The annotation rules were defined in a report which was shared with the annotators. This report is provided in Annex 2 - ROXSD Annotation Guidelines.

```
transcript: "  
0.250000    0.635000    B:fr01M_NT:english:yes?  
1.730000    4.365000    A:cs05M_NT:english:hey {Yan|PARTY} this is {Petr|PARTY}. is  
it you?  
3.820000    7.685000    B:fr01M_NT:english:ah! hey hey hey {Petr|PARTY}. yeah yeah  
sure sure sure it's me.  
7.850000    26.656508   A:cs05M_NT:english:hey excuse me. well I hope it's not too  
late for you that I make this call but I heard from actually from {James|THIRD_PARTY}  
that you were checking the latest bestsellers at the book store in {Brno|LOCATION}  
{today|TIME} and I thought maybe you could give me some advice on what to read.  
...  
"
```

Figure 8. Excerpt from a yaml file for an example conversation

## 2.4 ROXSD Video Subset

In order to illustrate the interest of exploiting the image modality, ROXSD was complemented with images and videos representative of files which may be found on a seized smartphone, seized computer, or grabbed from the internet. This corresponds mainly to selfie images or videos where various people are heard and/or seen while observing certain objects or locations. The captured images and videos enable the evaluation of face and scene matching technologies used in the Autocrime platform to enrich the speaker network with additional nodes and edges (for instance an edge is added between two speaker's nodes when both persons are found - either through their voice or face - in a same video).

### 2.4.1 Data Collection

The image and video extension to ROXSD was built from three different sources:

- Voices and Scenes videos captured during the ZITIS2 workshop (154 videos , 8 locations, 15 voices)
- Pseudonymized selfie videos from the video stabilization dataset<sup>14</sup> (90 videos, 47 identities)
- Additional selfie images or videos captured to match ROXSD scenario (14 images or videos, 4 identities, 4 locations)

The videos captured during the ZITIS2 workshop aimed at evaluating the scene matching performances and the ability to match voices across devices (in particular match voices captured from tapped calls on one side and from videos captured from a smartphone on the other side). For this, 15 different people (most of them, included in the ROXSD tapped calls) were asked to capture videos in 8 different outdoor locations while speaking and mentioning known names or locations from the ROXSD story. Videos were captured with the participant's own smartphone resulting in a diversity of capturing devices.

<sup>14</sup> <https://github.com/jiy173/selfievideostabilization>

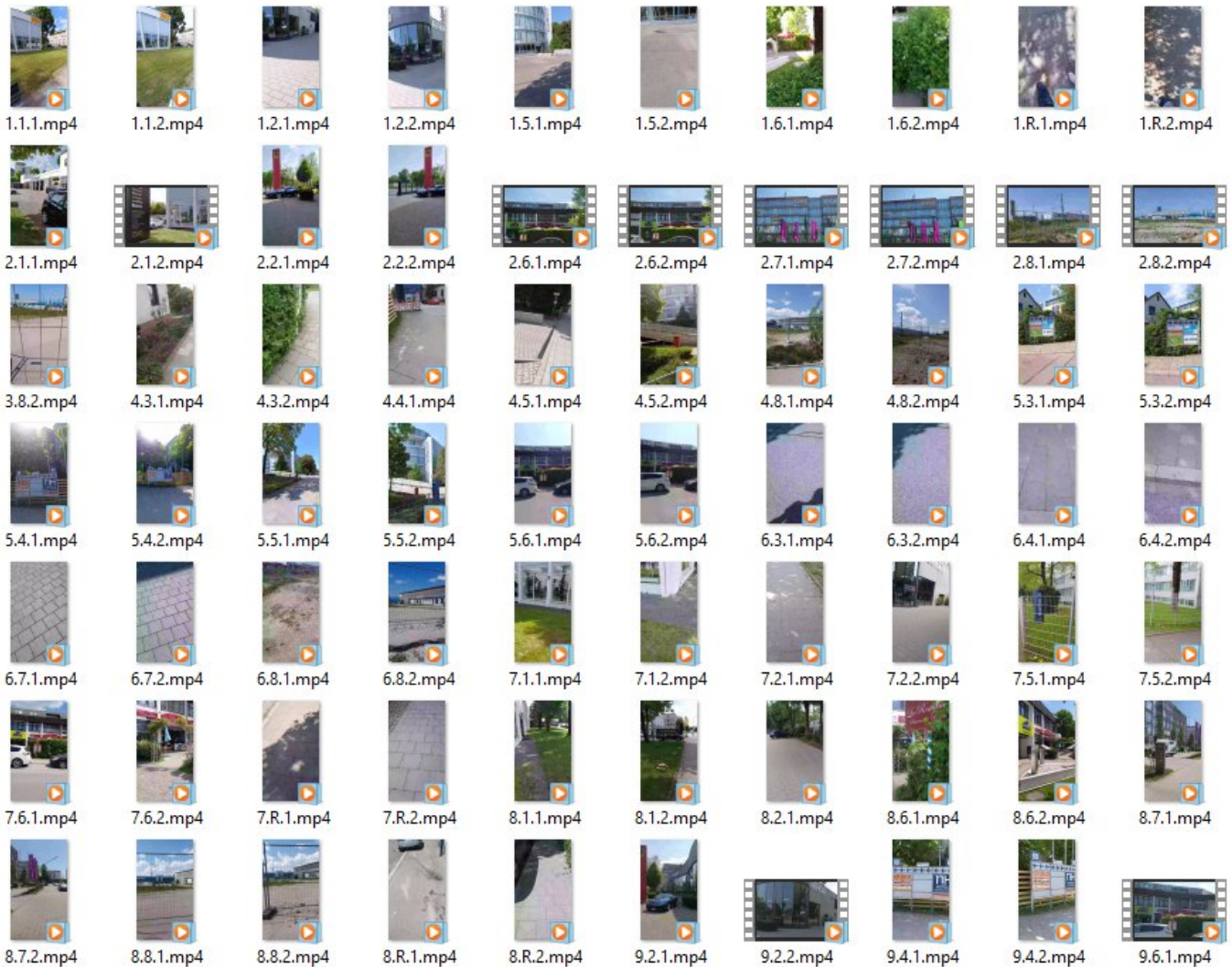


Figure 9. Snapshots of a subset of scene videos (naming convention: <speakerId\_locationId\_videoId>.mp4)

The naming convention of these files is the following: x.y.z.mp4 or x.y.z.mov with:

- x : person identifier number (from 1 to 15)
- y : location identifier number (from 1 to 8) or R for random location
- z : video identifier number

The second set of videos was added to enable the evaluation of face matching technology, while respecting regulations on biometric personal data. In practice, we adapted the SimSwap face swapping framework<sup>15</sup> to consistently pseudonymize faces observed in the “Realtime Selfie Video Stabilization” dataset referenced above. The SimSwap framework takes as input a video and a target face. It then consistently replaces the faces observed in the video by the target face. Fake faces generated by the StyleGAN2 model<sup>16</sup> were used as target faces, thus resulting in pseudonymized videos. We selected in the original dataset 47 identities for which at least two different videos were available. For each identity, we selected one video as a source for the enrolled face picture and the other(s) as videos to be processed by the Autocrime platform (resulting in a total of 90 videos). To complexify the face matching process, the enrolled picture for each identity was selected as follows: first we ran our face clusterization and cluster

<sup>15</sup> <https://github.com/neuralchen/SimSwap>

<sup>16</sup> <https://github.com/NVlabs/stylegan2>

summarization process on the video, resulting in the selection of 5 representative face observations of the person, the first observation corresponding to the cluster’s face centroid, the second one to the face observation the furthest from the centroid (but still in the cluster). We select this second picture as enrolment picture for each identity.

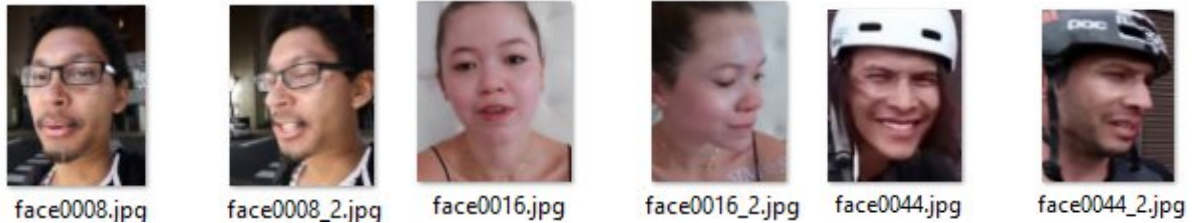


Figure 10. Examples of face observations selected for enrolment: for each person, the first picture corresponds to the cluster centroid, the second picture to the face observation, in the same cluster, the furthest away from the centroid. This is the one selected for enrolment.

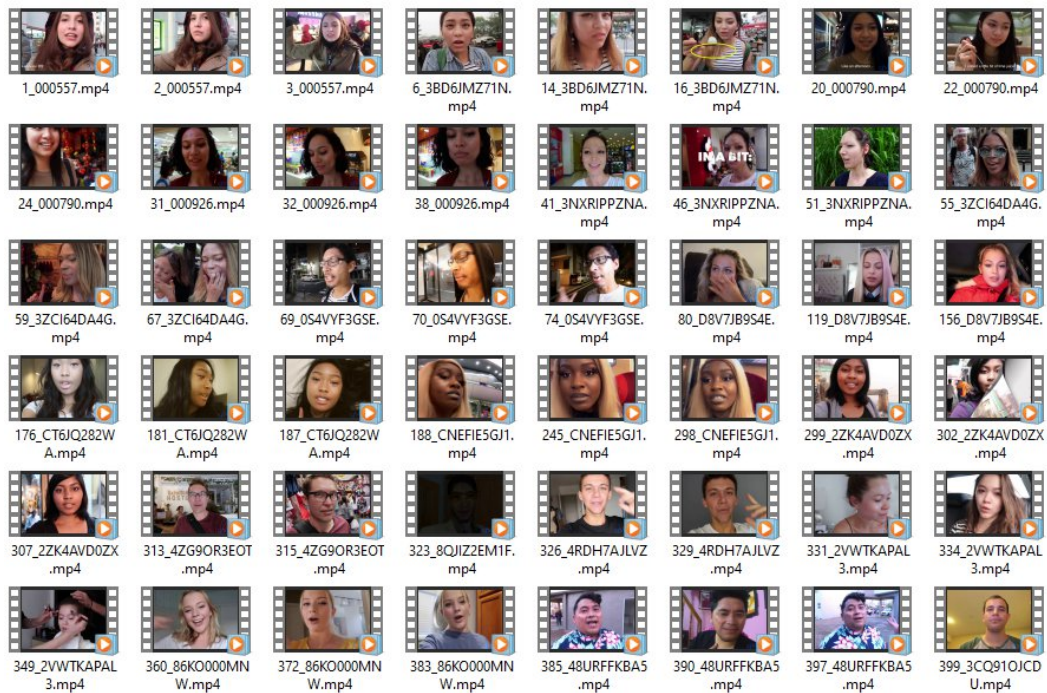


Figure 11. Snapshots of a subset of pseudonymized videos (naming convention: <nb-of-original-video>\_<fake-face-id>.mp4)

Finally, additional images and videos involving 4 characters of the ROXSD story were additionally captured to better match the ROXSD story. Faces involved in these videos were also pseudonymized.

## 2.4.2 Data Size and Statistics

In total, the added image and videos consist of:

- 51 different face identities
- 12 different locations
- 15 different voices

Some additional statistics:

- Total duration of videos: 1h38mn
- Spoken languages in the videos:

- Czech: 7 m 1 s
- English: 1 h 8 m 18 s
- French: 1 m 6 s
- German: 7 m 24 s
- Greek: 3 m 33 s

### 2.4.3 Metadata

A selection of 21 of the German speaking videos, and 23 of the English speaking videos were transcribed and annotated for NER.

A ground truth file provides, for each image or video:

- the list of speakers heard in the video
- the list of faces observed in the image or video
- the list of scenes observed in the image or video

## 2.5 ROXSD Text Subset: ROXHOOD

ROXHOOD dataset extends ROXSD by adding social media communications.

### 2.5.1 Data Collection

One of the objectives of WP4 is to use data from social media, either publicly available (public profiles related to the use-case), or lawfully intercepted from private profiles. More specifically, one of the activities in T4.3 - Social media data ingestion was also to consider including the forum of the INTERPOL's International Child Sexual Exploitation (ICSE) database. However, the ROXANNE consortium, following the advice and instructions of Ethics experts working in the Consortium decided not to investigate the INTERPOL's database. In addition, the exploitation of publicly available social media sources (platforms, websites or forums) was a very challenging activity in terms of following the legal European and National standards, based on written law(s) and regulation(s) and the ethical standards which are linked to the human rights. The Legal framework is basically defined by the General Data Protection Regulation (GDPR), which replaced the EU Data Protection Directive. One of the key challenges for data protection and ethics in social media landscape is the usage of personal or sensitive data. In order to avoid any legal and ethical issues arising from accessing and working on publicly available social media datasets, the WP4 team decided to simulate the social media environment (platform, website, forum) and the social media dataset that was to be collected. To do so, a mock-up website forum was set up and a tool was created for collecting and indexing the available data on this environment. The result of this collection is the so called ROXHOOD subset.

#### *The ROXANNE Forum - ROXHOOD*

The website for the text data collection sub-task of ROXSD extension is available here: <https://forum.roxanne.itml.gr/>. The forum is named "Roxhood". ROXHOOD was envisioned as a forum where legal and illegal content coexist. In this environment, info about several topics is being exchanged. ROXHOOD is a fully featured internet forum solution developed in Misago<sup>17</sup>. Misago is a fully featured forum application developed to follow the modern practices and trends currently used in web software development. The members of the ROXANNE Consortium, not only the partners involved in the WP4 tasks, participated in the collection of ROXHOOD. This happened during the time period from Feb 2022 to May 2022. In particular, during the ZITIS Workshop #2 in Munich

---

<sup>17</sup> <https://misago-project.org/>

(10-12/5/2022) the majority of the partners created an account (user registration) and provided data (mostly text but also some images and videos) across the available public threads.

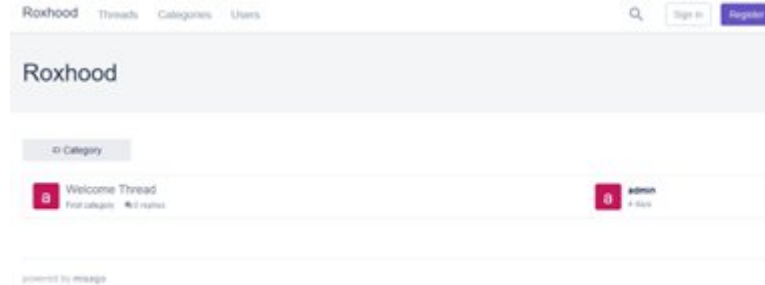


Figure 12. Starting page of the ROXANNE Forum - Roxhood

Visitors of the website (participants), only participants of the ROXANNE consortium, in the text data collection sub-task should register as users. Then, they should pick a Username & Password and provide a fake email account following the format of: YourFakeName@YourFakeDomain.com. This information must be saved for the next time that they will visit the website. So, users may register their accounts, set avatars, change options and edit their profiles. The option to reset the forgotten password is also offered. Moderator(s) from the /admincp panel, can also create categories together with unlimited number and depth of subcategories. Presence features let site members know when other users are online, offline, or banned. Individual users have setting to hide their activity from non-admins.

A toolset for complete moderation allowing admin-approved moderators to edit, move, hide, approve, delete, or close user posted content is also offered. This also includes option to delete or block user accounts or avatars. Ban system allows you to ban existing users as well as forbid certain usernames, e-mails or IP addresses from registering accounts. Permission system allows moderator(s) to control which features are available to users based on their rank, roles or category they are in.

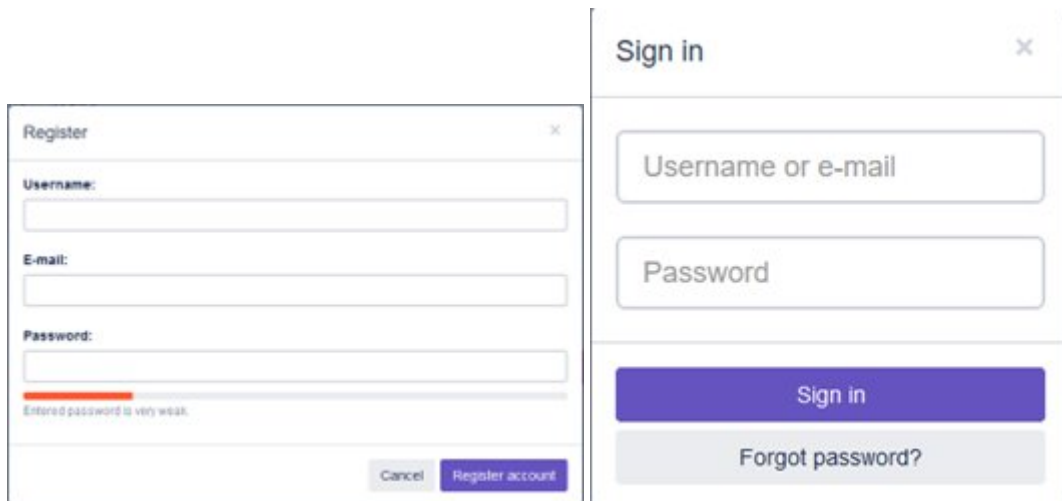


Figure 13. Registration on the ROXANNE Forum - Roxhood

### Registered User(s)

On the ROXANNE forum – ROXHOOD, registered users are allowed to see, set up and modify the options of their profile. In addition, they can change their avatars.

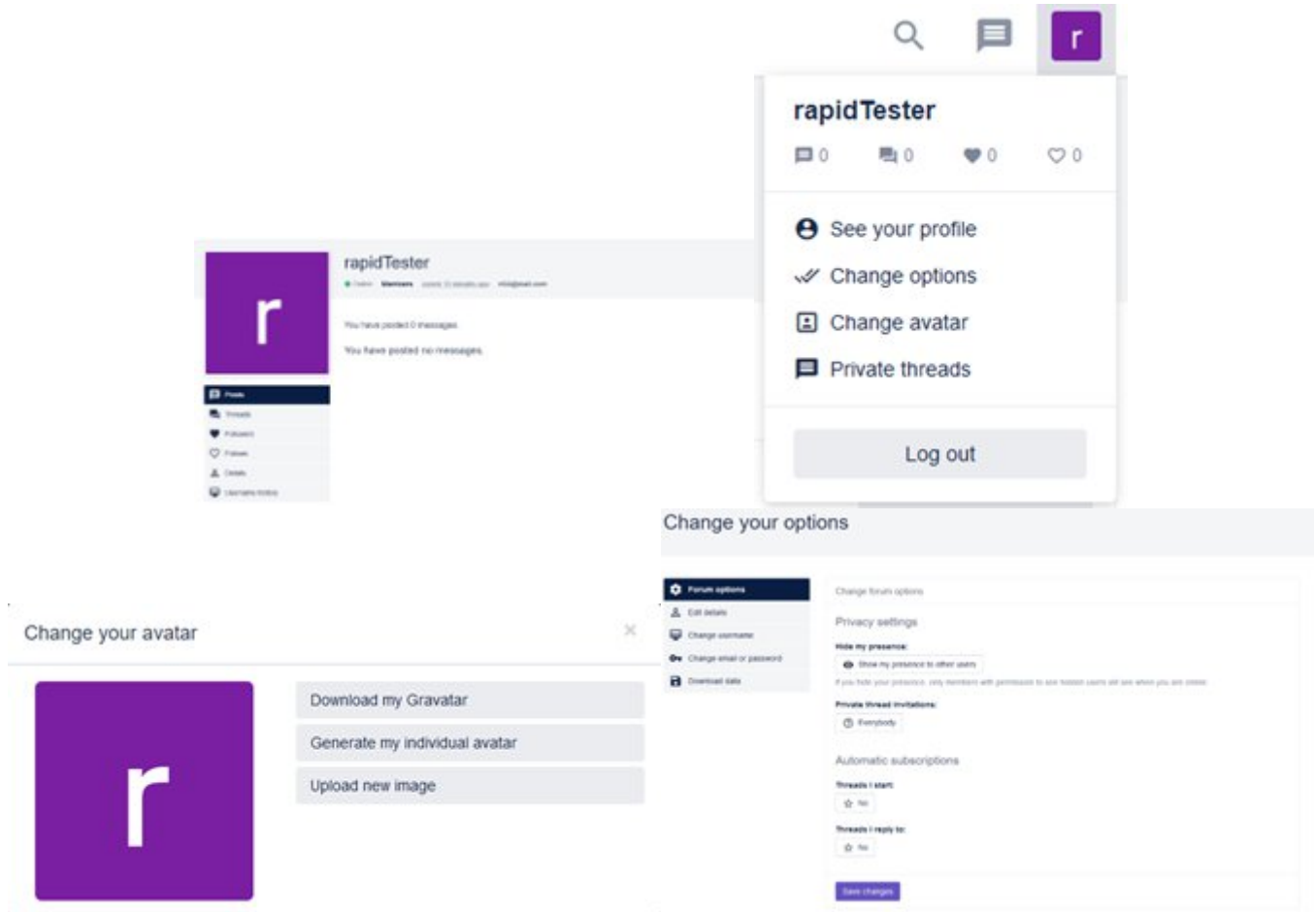


Figure 14. How to check, set up and modify options of a registered user

### Threads

In the public threads’ environment on the ROXANNE forum - ROXHOOD, registered users are allowed to start a new thread, comment, search also to set up a private thread. Private threads feature allowing users to create threads visible only to them and those they have invited.

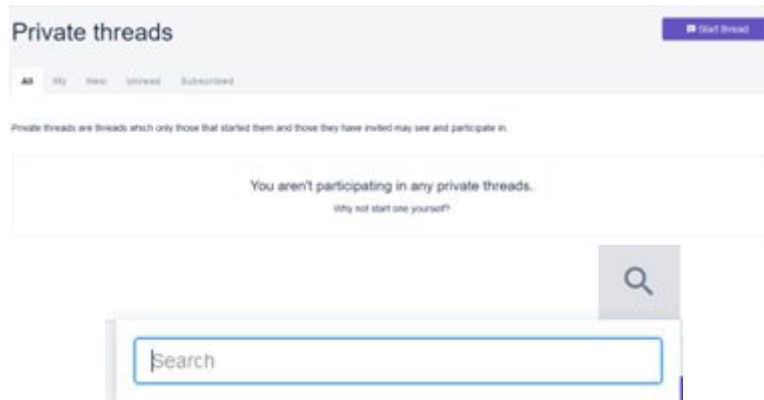


Figure 15. Private Threads; Available options

Registered users can write messages in text format, upload files of various formats (e.g. docx, jpeg, mpeg), include/provide links, comment on threads below the initial post, mention each other (via @usernameX). The moderator is handling the operation of the website.

### *Technologies*

The ROXANNE forum - Roxhood was implemented using the below technologies:

- **Django**, a high-level Python web framework that encourages rapid development and clean, pragmatic design.
- **Celery**, a task queue implementation for Python web applications used to asynchronously execute work outside the HTTP request-response cycle.

### *Additional features*

- Rich polls system, allowing polls with public and private voters, single and multiple choices as well as ones that allow vote change or limit voting to limited period of time.
- Post attachments complete thumbnailing and gif's animation removal.
- Mark post in question thread as best answer, bringing basic Q&A functionality.
- Posts edits log allowing you to see how user messages used to look in past as well as revert function protecting you from malignant users emptying their posts contents.
- Moderation queue for users and categories allowing you to moderate content before it becomes visible to other members of the community.



Figure 16. How to start or post on a new thread

### ***The ROXANNE Crawler - Rocraw***

#### *Login and execution of crawling*

The ROXANNE crawler - Rocraw is python solution to extract information from websites. Rocraw is implemented in order to be a site-specific(forum) solution targeting only the ROXANNE Forum – ROXHOOD (available here: <https://forum.roxanne.itml.gr/>). The main task of the crawler is to extract data/info about the threads created in the forum and store them at an Elasticsearch instance.

The ROXANNE crawler – Rocraw is available here: <https://forum.roxanne.itml.gr/crawler/>. A set of Username/Password is needed in order to Login.

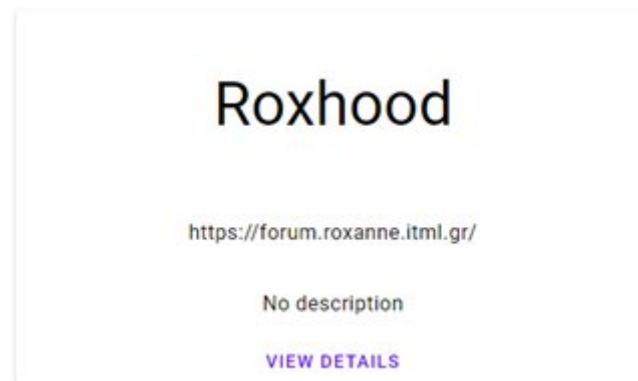




The screenshot shows a login form with the title "Log in" in a large, bold font. Below the title, there are two input fields: "Username:" followed by a text box, and "Password:" followed by a text box. At the bottom of the form, there is a purple button labeled "LOGIN".

Figure 17. The Log in interface

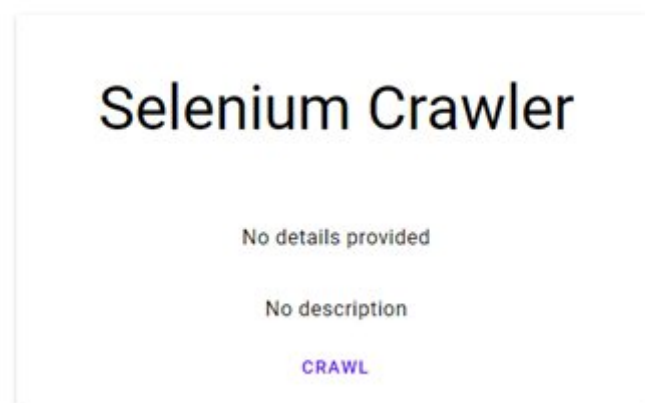
The next step is to select the site (“Roxhood”) that you want to crawl by hitting “VIEW DETAILS” button:



The screenshot shows the "Roxhood" site selection interface. The title "Roxhood" is displayed in a large, bold font. Below the title, the URL "https://forum.roxanne.itml.gr/" is shown. Underneath the URL, it says "No description". At the bottom of the interface, there is a purple button labeled "VIEW DETAILS".

Figure 18. The selection of site to execute the crawling task

One (1) crawler is available: the “Selenium Crawler”. By hitting the “CRAWL” button a new crawling task is initiated:



The screenshot shows the "Selenium Crawler" interface. The title "Selenium Crawler" is displayed in a large, bold font. Below the title, it says "No details provided". Underneath, it says "No description". At the bottom of the interface, there is a purple button labeled "CRAWL".

Figure 19. The “CRAWL” button to start a new crawling task

The task, with a “Unique id”, will be available in the “Tasks” section. “Results” is a json array<sup>18</sup>. An example can be found below. Each task when it’s successfully completed reaches the “Status” “finished”. Each task, has also information about when it occurred (“Created at”):

Unique id	Status	Crawler	Created at	Results
5de9eed-cdb7-48ea-905c-38c7c50fa93e	finished	Selenium Crawler	Nov 18, 2021, 11:52 a.m.	0 <a href="#">Download</a>
ca703a9-050c-477d-9165-e36ae56e6b0f	finished	Selenium Crawler	Nov 18, 2021, 11:54 a.m.	1 <a href="#">Download</a>
91248d38-5103-4870-ae31-1491f9e24868	finished	Selenium Crawler	Nov 19, 2021, 2 a.m.	1 <a href="#">Download</a>
3e5cd417-015d-4888-9e88-12a2989fa84a	finished	Selenium Crawler	Nov 22, 2021, 2 a.m.	1 <a href="#">Download</a>

Figure 20. The list of crawling tasks

### Access to the data

After the task has ended the “Results” column will have a number >0 and you can preview or download the results in a text file by hitting the “Download” button. The Results of the crawling tasks (data) are also available via the Elasticsearch infrastructure that is set up here: <https://forum.roxanne.itml.gr/elastic/> (a set of Username/Password is needed). Rocraw was fixed to automatically conduct a crawling task periodically every three (3) days. Each time a crawling task is requested (either on the rocraw webapp or via the Elasticsearch infrastructure) an index is created in the Elasticsearch. If the user(s) run several crawling tasks in the same day, then each one overwrites the previous one. At the end of the day only the latest version will be available in the Elasticsearch. In addition, in order to access data available in Elasticsearch directly from the from web browser a Kibana environment/tool is also available here: <https://forum.roxanne.itml.gr/kibana/> (a set of Username/Password is needed). The output (data) from one of the indices using Kibana to access Elasticsearch indices is:

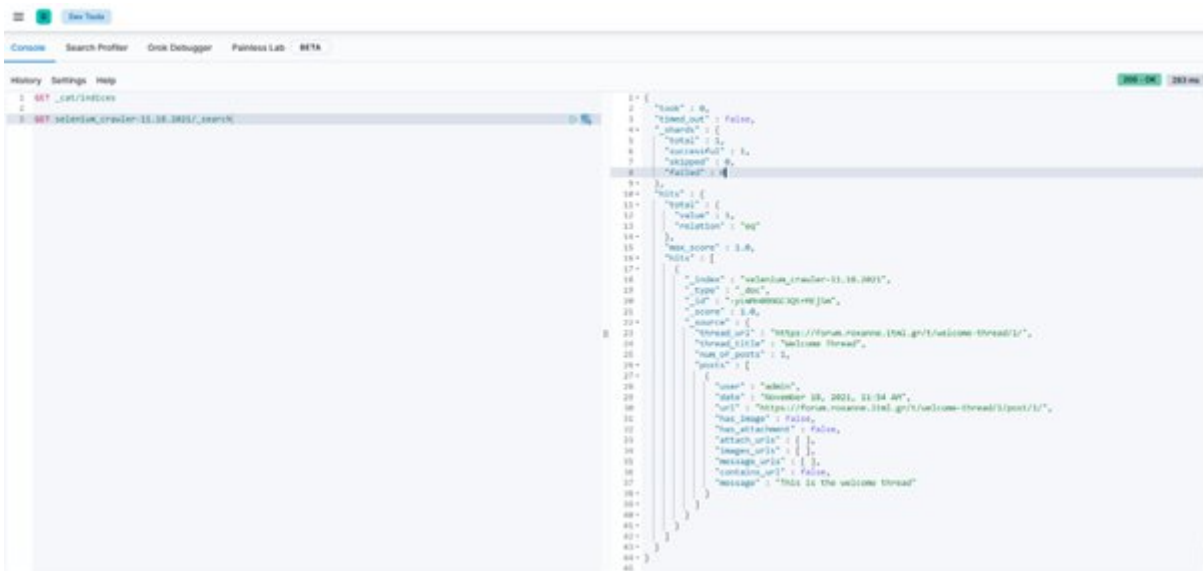


Figure 21. The output (data/metadata) of a crawling task using Kibana

while the output using Terminal (Powershell, CMD etc) is:

```

18 fields: {
  "num_of_posts", "thread_title", "thread_url",
  "posts": [{"date", "contains_url", "has_attachment", "has_image", "message", "attach_urls", "images_urls", "message_urls", "url", "user"}
}

```

<sup>19</sup> The body is provided in raw format and in UTF-8 encoding.

```

C:\Users\kdumpas>curl -u "crawler:RoCraw21" -XGET "https://forum.roxanne.itml.gr/elastic/selenium_crawler-11.18.2021/_search?pretty=true"
{
  "took": 0,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 1,
      "relation": "eq"
    },
    "max_score": 1.0,
    "hits": [
      {
        "_index": "selenium_crawler-11.18.2021",
        "_type": "_doc",
        "_id": "-yLmHn8BNGC3QtrMEJ5m",
        "_score": 1.0,
        "_source": {
          "thread_url": "https://forum.roxanne.itml.gr/t/welcome-thread/1/",
          "thread_title": "Welcome Thread",
          "num_of_posts": 1,
          "posts": [
            {
              "user": "admin",
              "date": "November 18, 2021, 11:54 AM",
              "url": "https://forum.roxanne.itml.gr/t/welcome-thread/1/post/1/",
              "has_image": false,
              "has_attachment": false,
              "attach_urls": [ ],
              "images_urls": [ ],
              "message_urls": [ ],
              "contains_url": false,
              "message": "This is the welcome thread"
            }
          ]
        }
      }
    ]
  }
}
C:\Users\kdumpas>

```

Figure 22. The output (data/metadata) of a crawling task using Terminal

### Technologies

The crawler is implemented using the below technologies:

- **Django**, a high-level Python web framework that encourages rapid development and clean, pragmatic design.
- **Scrapy**, an open source and collaborative framework for extracting data from websites.
- **Selenium**, a powerful tool for controlling web browsers through programs and performing browser automation.
- **Gunicorn**, is an application server for running python application instances.

The Crawling Task Flow has 4 (four) steps and is depicted below.

1. **Gunicorn** gets an http post request
2. A new crawling task is started by **Scrapy**
3. **Scrapy** using multithreading and **Selenium** extracts data from website
4. Once the crawling tasks is finished, extracted data are stored at Elasticsearch and PostgreSQL

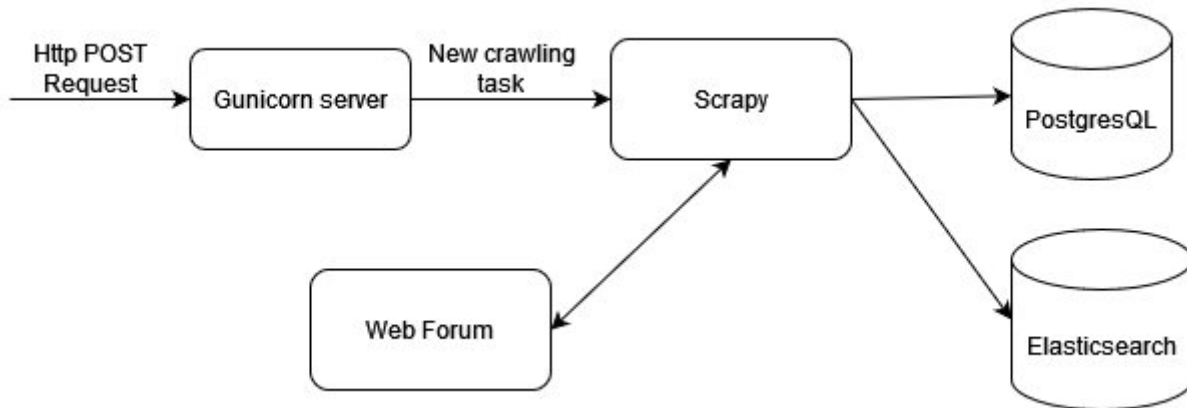


Figure 23. The crawling task flow.

## 2.5.2 Data and Metadata; Info, Size and Statistics

The ROXANNE Forum – ROXHOOD in numbers can be described based on the available fields. These are:

Available fields: "thread url", "thread title", "number\_of\_posts" (per thread), "posts", "thread\_creator", "mentions-replies", for each post ("posts" field), "message", "user", "date", "url", "has\_image" (true/false), "has\_attachment" (true/false), "contains\_url" (true/false), "image\_urls", "attach\_urls", "message\_urls".

The Number of public threads is 32 public; The Number of public posts is 333; The Number of registered users is 48. Following the analysis of Natural Language Processing (NLP), including Named-Entity Recognition (NER) on the ROXHOOD dataset metrics describe the content that was and the entities that were extracted. ROXHOOD included mostly Text Messages but also English Videos. Across 299 posts, 110 entities were included in the ROXHOOD dataset. The entities were about person(s) [21], location(s) [49] and time [40]. Finally, regarding the content of English Videos, across 23 videos, 107 utterances and 91 entities have been defined. Again, for the entities of person, location, time the numbers are 44, 26 and 21 respectively.

## 2.5.3 Fusion of ROXHOOD with ROXSD

ROXHOOD dataset, although collected separately of ROXSD, is finally fused with the ROXSD dataset so that the ROXANNE technologies can be fully evaluated.

Extension of ROXSD by ROXHOOD helps in many ways:

- Helping to analyse more complex cases, where not only intercepted telephone communications are available, but also the social media communications significantly enrich the analysis of the criminal cases. The idea, according to several use-cases collected during the project length, where voice/video communications from intercepted phones are often extended by social media was presented several times as part of WPs by LEAs.
- As there are few same target persons (suspects) appearing in both datasets, we can see the fusion (combination) of these datasets directly in the graph created from persons. This therefore significantly increases the size of the evaluation data especially for testing the network analysis technologies.



- Cross-network analysis: specifically, cross-network analysis, analysing similarities across two or more separated networks, can be tested.
- Testing natural language processing technologies can be done on both ROXSD and ROXHOOD datasets.

## 2.6 ROXSD Release

The ROXSD exists as a distributable package which is composed of several files structured in a folder hierarchy. In this subsection, we briefly talk about the structure of the dataset package as well as the tentative plans for its distribution to the research community after the end of the project.

### 2.6.1 Content and structure

The folder package containing the ROXANNE Simulated Dataset (ROXSD) consists the following main parts:

- 00\_ROXSDv3\_License - contains “License and Data Processing Agreement”
- 01\_ROXSDv3\_Audio\_Files - contains the wave files associated with the ROXSD calls
- 02\_ROXSDv3\_Metadata - contains excel file describing the content and connections between speakers
- 03\_ROXSDv3\_Transcriptions - contains transcriptions of calls
- 04\_ROXSDv3\_Translations - contains translations of calls transcriptions
- 05\_ROXSDv3\_ASR - contains ASR outputs
- 06\_ROXSDv3\_NLP - contains NLP outputs
- 07\_ROXSDv3\_Speaker\_Language\_Gender\_ID - contains mentioned technologies outputs
- 08\_ROXSDv3\_Network\_Analysis - contains network analysis details
- 09\_ROXSDv3\_Geolocation - contains geolocation information
- 10\_ROXSDv3\_ROXHOOD - contains social network communication subset related to relevant speakers
- 11\_ROXSDv3\_Videos - contains video subset related to relevant speakers
- 12\_ROXSDv3\_Presentation\_Case - contains relevant information to understand the case
- md5sum.txt - checksum to validate that the whole dataset was downloaded successfully
- README.md - “read me” file containing basic dataset description and technical information

The Excel spreadsheet inside the 02\_ROXSDv3\_Metadata folder is the main “metadata” file which contains the ground-truth information of the speakers, calls and videos in ROXSD. The structure of this spreadsheet had to be modified several times during the data preparation and collection process. As of the writing of this deliverable, this metadata spreadsheet (v3.3.1) is composed of the following sheets and columns:

- README: An introductory sheet containing basic information and statistics about the data in ROXSD
- Research\_numbers: Sheet containing ground-truth information about the telephone devices, numbers and users in ROXSD calls, with a wholistic (researcher’s) view. The columns are as described in Section 2.3.3.1.
- Research\_calls: Sheet containing ground-truth information about the calls and geolocation with a wholistic (researcher’s) view. The columns are as described in Section 2.3.3.2.
- Police\_numbers: Sheet containing information about the telephone devices, numbers and owners as viewed from the police investigator’s side (as shown by the interception mechanism/CDR, etc.)
- Police\_calls: Sheet containing information about the calls and geolocation as viewed from the police investigator’s side (as shown by the interception mechanism/CDR, etc.)
- IMEI: Sheet containing the telephone numbers, associated IMEI numbers, and the duration of use according to the story
- SMS\_IMEI: Sheet containing information about the SMS messages as sent from one character to another.



Excerpts from the Research\_numbers and Research\_calls sheets are shown in the following Figure 24 and Figure 25, respectively.

	A	B	C	D	E	F	G	H	I	J	K	L	M	
	Character_ID	Drama_number	Anonymiza	Drama_device	Intercepted_nu	Character_lab	Character	Ca	Age_ranj	Gend	Native_languag	Second_languag	Third	
2	17	17	+420 736 88221	x	mobile	DDA_07_PG	ar01F_NT	Ines	DDA	25-30	female	tunisian arabic	french	english;german
3	40	40	+420 784 57162	x	mobile	-	ar02M_NT	Rayan	-	36-40	male	tunisian arabic	german	french
4	25	25	+420 608 44763	x	mobile	DDA_01_PG	cs01M_T	Kryštof	DDA	41-45	male	czech	english	german
5	80	25	+420 724 26714	x	mobile	DDA_02_PG	cs01M_T	Kryštof	DDA	41-45	male	czech	english	german
6	81	25	+420 736 88221	x	mobile	DDA_07_PG	cs01M_T	Kryštof	DDA	41-45	male	czech	english	german
7	23	23	+420 736 98828	x	mobile	-	cs02F_T	Kača	-	25-30	female	czech	english	
8	1	1	+420 606 32873	x	mobile	-	cs03M_T	"K"	-	36-40	male	czech	english	
9	14	14	+420 728 47325	x	mobile	DDA_04_PG	cs04M_T	Kuba	DDA	31-35	male	czech	english	french;slovak

Figure 24. Screenshot from the Research\_numbers sheet

	A	B	D	F	H	I	K	L	M	N	O	P	Q	R	S	T	U
	Call_ID	Cluster_ID	Audio_file	Caller_number	Caller_IMEI	Receiver_number	Receiver_IMEI	Intercepted_nu mber_ID	Caller_label	Caller	Caller_cha nnel	Receiver_label	Receiver	Story_date	Story_time	Audio_dur ation	Language s
1	12	-	RE8ff8ec4c4e445c13a3d3a6258a7f4a61.wa v	+420 724 26714	00-000000-0000000	+420 736 98828	35-153504-4960978	DDA_02_PG	cs01M_T	Kryštof	L	cs02F_T	Kača	2019-12-04	14:51:36	0:00:45	czech
2	30	-	REf1097d1ff654a560b387e94ff5bc361f.wav	+420 702 90329	35-079890-4341112	+420 724 26714	00-000000-0000000	DDA_02_PG	ru01M_T	Sergej	L	cs01M_T	Kryštof	2019-12-05	21:05:39	0:01:08	english
3	217	-	RE369dd60dba400abac1067dd994882cda. wav	+420 726 14522	35-271304-4473501	+420 728 47325	12-345678-9012345	DDA_04_PG	en01M_NT	Adam	L	cs04M_T	Kuba	2019-12-05	17:07:00	0:03:05	english
4	218	-	RE2b493cdf4a29aa6262aa6d03800b2153.w av	+420 726 14522	35-271304-4473501	+420 728 47325	12-345678-9012345	DDA_04_PG	en01M_NT	Adam	L	cs04M_T	Kuba	2019-12-05	11:56:56	0:04:07	english
5	243	-	RE245e0716593e88ae1851d917654a7e2d.w av	+420 738 61659	00-000000-0000000	+420 606 20626	35-067371-5488508	DDA_03_PG	cs07F_T	Kristýna	L	en02M_NT	Paul	2019-12-05	12:40:28	0:03:03	english

Figure 25. Screenshot from the Research\_calls sheet

The Cluster\_ID, Call\_type and Connection\_status columns in the Research\_calls sheet allow for selecting a subset of the calls, or replacing device-speaker assignments to prepare a different version or interpretation of the ROXSD story. The users of Autocrime will receive some Python scripts together with their platform installation to make use of the ROXSD. One of these scripts convert this Excel spreadsheet into a CSV file which can be digested as input by the platform. Another script is capable of re-assigning telephone numbers to speakers to modify the ROXSD story in a way which is suitable for the desired use-case.

## 2.6.2 Accessibility and availability

During the project realisation, the dataset is available to ROXANNE Partners and End Users (LEAs in EU) for research, development and demonstration purposes. To maximize its value, care was taken to ensure that the dataset will be usable after the lifetime of the ROXANNE project, mainly for research and presentation activities.

A form of the license was discussed among the consortium partners, regarding GDPR regulations and other legal concerns. This document is also part of the exploitation plan.

The main aim is to distribute the dataset outside of the consortium even after the end of project, specifically to further support research in FCT types of projects, as supported by the EC. Beyond the members of the consortium, who already have access, the data \*must not leave the jurisdiction of the EU.

The tentative plan for distributing the ROXSD data after the project ends (after 12/2022), is as follows

- legal points
  - dataset accessible under “License and Data Processing Agreement” (see Annex 4 - License),



- license driven mainly by GDPR regulation - i.e., we will follow the rule that any speaker has right to be forgotten (it means he/she can withdraw his/her consent that his/her voice is part of dataset)
- distribution - to be managed by one party on behalf of authors
- usage for - limited for research and for free (obligations need to be fulfilled)
- usage in geography - defined as EEA (EU and Swiss is there; not Israel - we did not find better way)
- obligations of the receiving party - they will be obliged to delete certain parts on their local copy (i.e., if someone withdraw his/her consent) upon distributor request
- technically
  - availability - distribution will be done through a Zenodo portal, where similar research datasets are distributed.

The exact terms are available in the draft ROXSD License (see Annex 4 - License).

### 3 Other Datasets Employed in the Project

This section describes data provided by LEA partners.

#### 3.1 Real Case Data Corpus

One of the ROXANNE partners has provided data from a real criminal case for analysis within the project. The original data contains telephone communication between criminals involved in a drug dealing case. There is information about approximately 40K interactions. Around 30k of the interactions are phone calls, the rest are SMS. For the phone calls, the speakers have been identified manually. For both phone calls and SMS, information about the source and target number as well as the start and end time for the call is available. Among the 30k phone calls, around 200 were of special interest for the LEA and was transcribed manually by them.

Due to the sensitive nature of the data, audio, transcriptions, identity of speakers and phone numbers have not been transferred to the project partners. Instead, speaker embeddings (a.k.a. voiceprints) were extracted from the ~200 recordings of special interest by the LEA on their site. The speaker embeddings together with pseudonymized speaker labels and phone numbers were provided to the other project partners. This allows for experiments on speaker recognition and network analysis as well as their combination. On the other hand, experiments on ASR will have to be run on the LEA side.

The data has the following metadata:

- Locations mentioned in the call.
  - Obtained by named-entity recognition (NER) followed by a manual check.
  - When distributed to the project partners, the real location names were substituted with capital letters.
- People mentioned in the call.
  - Obtained by NER followed by a manual check.
  - The two people in the call are excluded, i.e, these are third party mentions.
  - When distributed to the project partners, the real location names were substituted with digits.
- Identity of the speakers in the call. Manual confirmation by listening.
  - Not ordered since the labels were obtained by listening to mono audio.
  - The labels are consistent with the labels used for people mentioned in the call described in the previous bullet
  - Available for all ~30K telephone calls.

- Anonymized telephone number of the receiver.
- Anonymized telephone number of the caller.
- Manual transcript.
  - Only available to the LEA owning the dataset.
  - Available for the ~200 relevant calls. Generally these are calls between criminals but it may not be always.
- End date and time of the call.
- Start date and time of the call.

The network calls using speakers and phone numbers as nodes are shown in *Figure 1* and *Figure 4*, respectively.

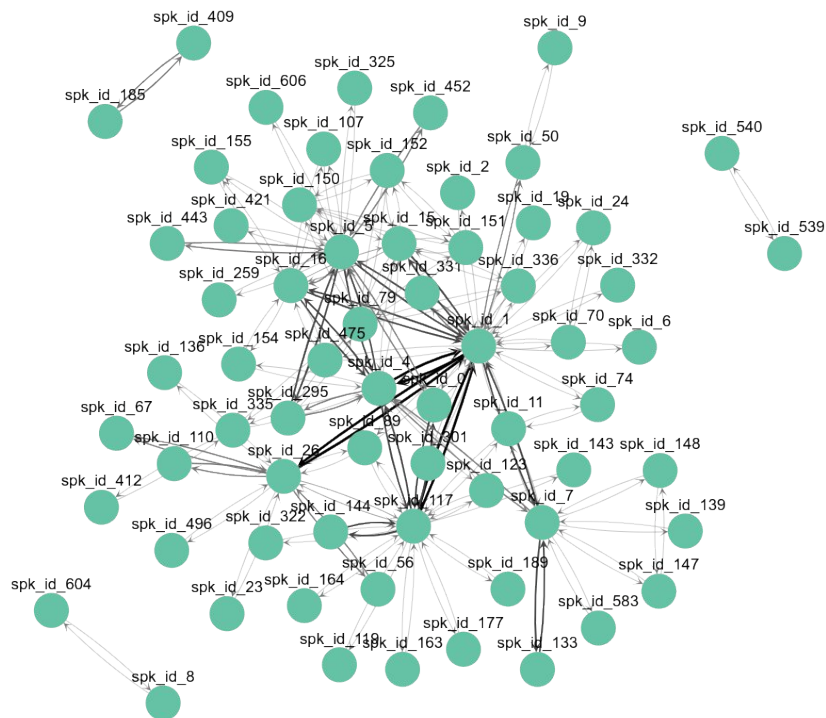


Figure 26. Groundtruth network. The nodes are speakers and the links are calls. There are 66 speakers in the network.



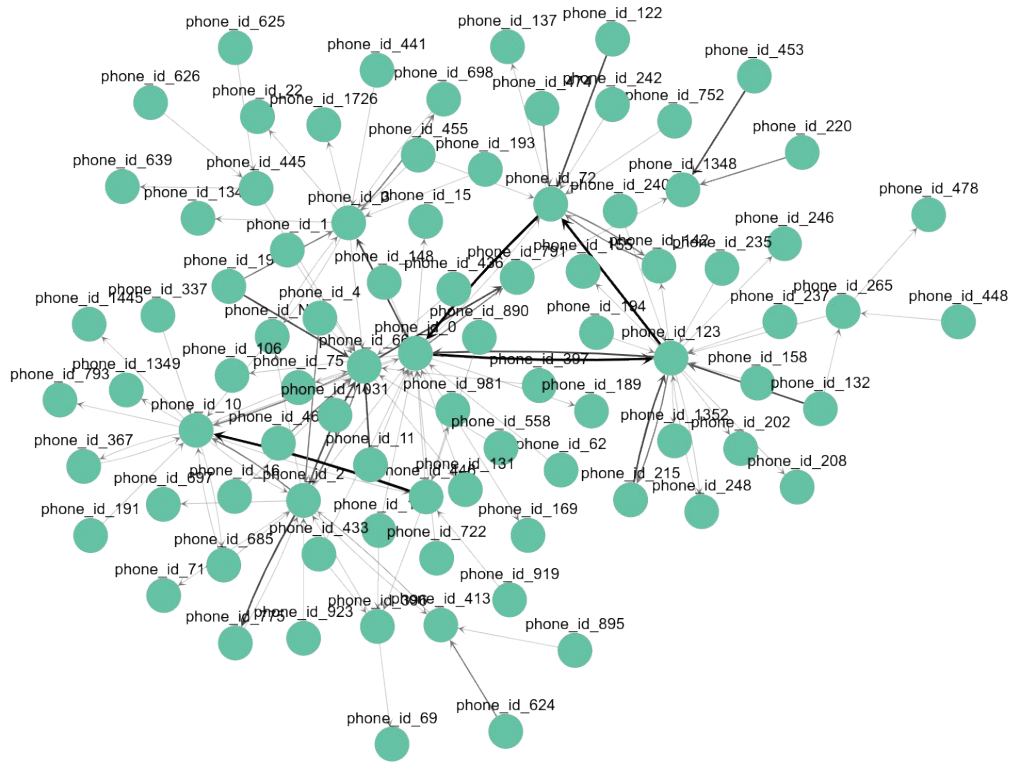


Figure 27. Phone number network. Each node is a phone number and the links are calls. There are 86 phone numbers in the network.

### 3.2 FRIDA - Forensically relevant inter-device audio database

FRIDA<sup>20</sup> is a database of forensically relevant speech recordings that were acquired simultaneously by multiple recording devices. The telephone conversations are in Dutch between 250 speakers. Particular care was taken to recruit speakers from different socio-economic backgrounds. Each speaker has 16 conversations with another participant for approximately five minutes. Approximately half of the conversations, i.e., around eight conversations per speaker, have been orthographically transcribed by three native Dutch speakers following a transcription protocol. This resulted in the transcribed data from 223 participants.

This dataset has been used on text-based speaker recognition, as described in D5.3<sup>21</sup>.

### 3.3 Case 1 Data

Case 1, is a simulated drug trafficking case. The database consists of 28 recorded conversations carried out according to the script by police officers of our services for the needs of the programme. The names, telephone numbers and metadata of the calls are not real. In addition, five calls from an old and closed case were included. This data was completely anonymized by a police consortium partner. The scenario was based on the assumption that we know 11 of the people involved in drug trafficking, some of whom in the the case changed phones and numbers and made calls to some of the known individuals. The purpose of the platform would be to be able to identify the new users of the phones if they are the same as those already known. Most of the calls are in Greek except two which are in English.

<sup>20</sup> D. van der Vloed, et al., “Exploring the Effects of Device Variability on Forensic Speaker Comparison Using VOCALISE and NFI-FRIDA, A Forensically Realistic Database.” In Odyssey 2020 The Speaker and Language Recognition Workshop, 402–7. ISCA: ISCA.

<sup>21</sup> D5.3 Final set of ROXANNE speech/NLP/video technologies for network analysis, Deliverable of H2020 project No. 833635 - ROXANNE, October 2022.

### 3.4 Case 2 (Burglary) Data

The dataset used in this study is provided by a police consortium partner and anonymized to a standard compliant with their national law and GDPR (and approved by a legal adviser at the police management), contained around 30,000 reports of solved burglary cases that occurred in a country (analysed by the police partner) between 2018 and 2021. The information contained in the reports included a crime identifier, the respective anonymized identifiers of the offenders, the min-max-scaled site coordinates of the crime (which prevent precise retrieval of the localization of the site), timestamps and a parameterized free-text description of the case in the form of an embedding vector, in particular, a Smooth Inverse Frequency (SIF) embedding<sup>22</sup>. The properties of the crime nodes in our graphs are transformed into embedding vectors that are built from the textual description of the incident. Although the spatio-temporal information of a crime is very important for applications like link detection and prediction, we have ignored these features, and considered the problem as a routine independent investigative filtering procedure. We parameterized the textual information of the crime using (SIF). This method can encode word sequences in a sentence or paragraph into a single vector, which mathematically represents the crime description. In short, this method smartly combines the embeddings of each word within a sequence as to identify the most relevant ones, a simple semantic text similarity task that has been shown to work well previously<sup>23</sup>. Although there are other more sophisticated alternatives for text embedding, such as those based on BERT transformers<sup>24</sup>, we finally opted to implement SIF embeddings, which was suggested by a preliminary experiment carried out using a development dataset of crime textual descriptions. In this experiment, we observed that SIF embeddings led to greater distinction of linked and unlinked cases compared to the BERT embeddings. One possible explanation is that SIF used a domain-specific trained word2vec, while BERT relied on a pre-trained general model. Three different networks were then generated based on the original burglary data and its text embeddings, including:

- A *crime-offender network* with 41,324 nodes and 34,156 edges, where nodes represent both crimes and offenders, and links indicate whether an offender participated in a crime. This network is a bipartite graph with two different types of nodes, which are separated next into two networks for further analysis.

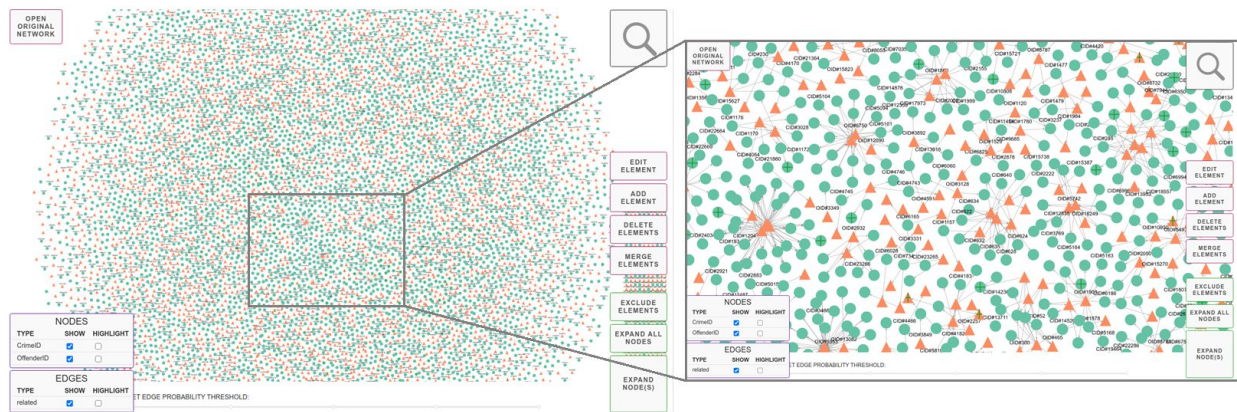


Figure 28.

<sup>22</sup> Arora, S., Liang, Y., & Ma, T. (2017). A simple but tough-to-beat baseline for sentence embeddings. In International conference on learning representations (ICLR).

<sup>23</sup> Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

<sup>24</sup> Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics.

- An *offender network*, where the nodes represent offenders and the links indicate whether two offenders are involved in one or more burglary cases. To generate the offender graph, we used the original undirected crime-offender network to connect two offender nodes each time they shared a crime, and then filtered the crime nodes. The offender network resulted in a total of 17,232 nodes and 21,302 edges. We used this dataset in the link prediction experiments as the network is large enough for this purpose.

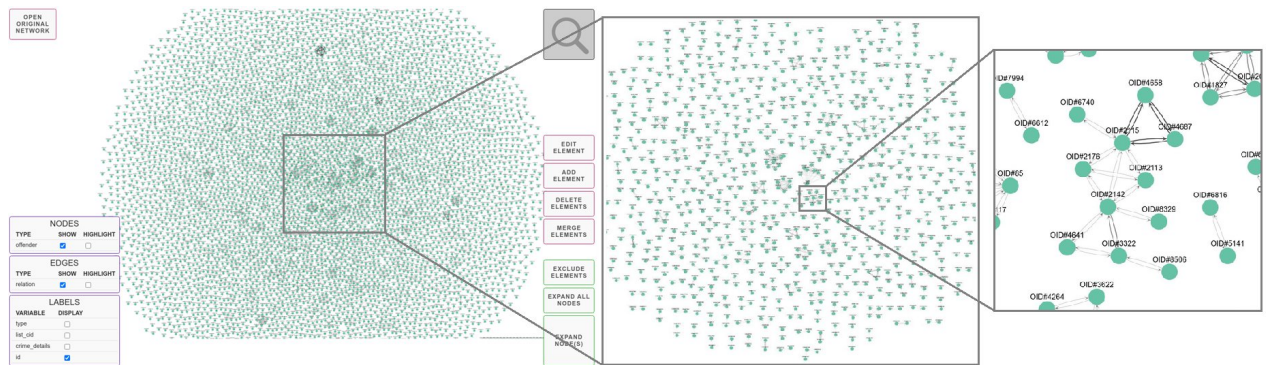


Figure 29.

- A crime network with nodes representing offenses and edges indicating the number of shared offenders between the crimes. This network included 23,380 nodes and 42,604 edges, which is also used for link prediction experiments, especially inductive link prediction.

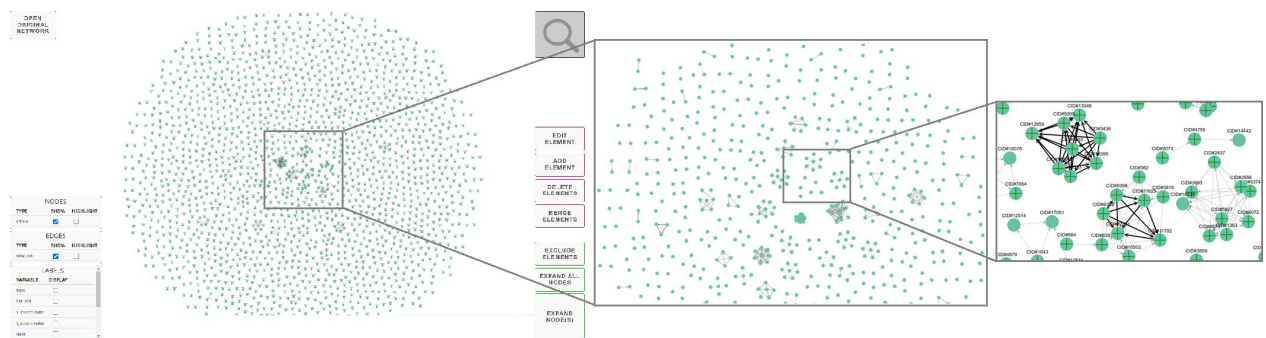


Figure 30.

### 3.5 Case 3 data

This data set contains ten intercepted GSM telephone conversations in Lithuanian language from various real cases (mostly swindling). The recordings are stored as mono, i.e. the two sides in the call are mixed. In addition, the voice sample of a known criminal is also available (comparative recording). The interest of the law enforcement agency is to detect whether the criminal speaks in any of the ten intercepted calls.

## 4 Ethics in Data Collection and Processing

During the proposal stage, there was an expectation that partners would be able to easily access and use data from LEA partner investigations. During the project, it became clear that this would not be as easy as anticipated. Primarily, because LEAs are very careful about the data they share, and many LEA officers are resistant to data sharing. Secondly, where LEA officers were willing to share data, there was a lack of legal support to complete the necessary documentation (i.e., data protection impact assessments, as required for the processing of criminal data under Art.35,



GDPR, and controller-processor contracts). As a consequence, the consortium had 2 main options: anonymising LEA data (due to the difficulties in successful anonymisation, this would be difficult and would not be able to provide enough data quality for all tasks that realistic data was required for); recording realistic, but fictitious, high-quality data that would meet the data needs of the ROXANNE research partners.

The consortium partners opted to record fictitious data using volunteers (whose participation was reviewed by the Ethics Review Board at USAAR). In ROXSDv.1, recordings were made using a Twillo application that allowed for recordings in a way that mimicked covert recordings conducted by LEAs. In ROXSDv.2 and v.3, which were recorded with ZITiS using real interception equipment (no data entered into investigation files).

During the project, the data were stored on a secure server at BUT, and were accessible by project partners. After the project, the tentative plan is that the dataset will be uploaded to Zenodo or a similar data hosting platform, and will be distributed by a ROXANNE partner for other researchers to use.

As noted in work conducted in WP3 and WP10, the recordings used volunteers that were fully aware of what their participation would involve. They were also informed that they were able to withdraw their consent to participate in the research activity at any time, and how to exercise their data-subject rights if they no longer wish to be part of the data set. Further, as noted above, the criteria for participation, and the procedures used for managing participants were approved with the USAAR Ethics Review Board. As such, the use of human participants in this instance was fully compliant with research ethics principles.

## 5 Summary and Conclusion

This deliverable describes the final data sets used to evaluate the ROXANNE technologies. The criteria and importance of realistic data as well as the difficulties in obtaining real data from criminal investigations led the project partners to collect data from a simulated criminal investigation named the ROXANNE simulated data set (ROXSD). The main focus of this deliverable is on that data set. In addition, several smaller data sets have been provided by law enforcement agencies. These are either simulated or anonymised and while being very valuable by providing great insight in the challenges and peculiarities of data from real cases, not all technologies in the ROXANNE platform can be evaluated on them.

ROXSD is a dataset of calls simulating a criminal case, based on a scenario provided by a LEA organization and covering drug-trafficking. It was created in order to have data suitable for R&D and demonstration in the ROXANNE project and beyond. It contains 432 telephone conversations recorded in 481 audio files, making up to more than 18 hours of audio data. The calls are made by 103 speakers in 15 different languages.

We believe that the ROXSD data-set is as close as possible to the real target data and it has been indispensable for R&D in the project. Its main benefit over existing “speech” data-sets is the availability of ground-truth on the case, story, and criminal network structure. It has successfully been used in the three field tests of the project. We also anticipate that the ROXSD data set will be a valuable resource to the research community in the future due to the increasing need and interest for security related research.

## Annex 1 - ROXSD Transcription Guidelines, v2.2.1

### Introduction: What to transcribe

This documentation is a guide for producing time-aligned audio transcriptions to be used in speech technology for machine learning and model building.

The transcription of a speech recording should contain a technically faithful, objective, and verbatim transcript of what has been said - supplemented by minimal markup.

You should transcribe exactly what was spoken, not what you expect to hear or what you consider would be correct.

In general, you should follow these guidelines:

- Transcribe exactly what is being said, not what you think is correct – do not improve, correct, or smooth out. Transcribe all errors and repetitions. Do not correct grammatical errors such as, e.g., 'she finded it' for 'she found it'. Do not correct non-standard usage or misused words, but otherwise follow the standard orthographic rules of the given language.
- Transcribe speech passages as long as the words are mostly intelligible, even when they are accompanied by constant background noise, contain scratching, cracking, clicking, or other sounds, or have bad quality audio with overmodulation or clipped amplitudes, and alike.
- Do not spend too much time on trying to understand difficult passages. If the words are not intelligible or there is too much noise, use the appropriate markers presented in the “Step 2: Transcription” Section.

## Transcription Process

### Step 1: Segmentation

---

Before starting to transcribe, the recorded audio needs to be divided into small time-aligned pieces. These units are called segments and usually correspond to speaker turns, sentences, or utterances, all starting and ending with a speech pause.

The ROXSD audio files which you will transcribe come pre-segmented, i.e., parts of the audio containing speech have already been marked. However, since the segments were obtained using an automatic system, there may be errors in the markup. If you find any errors or inconsistencies, please modify them (extend, shorten, delete or create new) as necessary. Follow these rules when modifying the segments:

- Ideally, a segment consists of a speaker taking a breath, uttering a sentence, and then stopping.
- Segment borders should always be placed in between words, not inside of words, and not at the very boundaries of words where speech sounds could be truncated.
- When a speaker pauses, typically longer than 0.5 seconds, begin a new segment.
- Also, when a new speaker starts to speak, begin a new segment.
- If in doubt begin a new segment.

### Step 2: Transcription (Left-to-Right Scripts)

---

#### Orthography and spelling

- Do NOT capitalize the first letter of each sentence.
- Capitalize only the first letter(s) of proper nouns: person names, locations, organizations, etc., and words which are typically capitalized in isolation (this may be language specific: the pronoun “I” in English, all nouns in German, etc.).
- Use underscores within names of well-known entities such as person names, locations, organizations, companies, etc. that consist of more than one word and usually occur in the same, known context:

*New\_York Joe\_Biden*

- Separate the component letters of abbreviations by underscores if it was spelled out:

*A\_B\_C news is talking about F\_B\_I and C\_I\_A.*

*we moved to a hotel called H\_N\_Messe.*

- Do not use underscores for acronyms! Acronyms are formed from the initial letters of its constituent words, but unlike abbreviations those letters are not spelled out separately but connected to form a new word:

*NATO UNESCO*

- Numerals (cardinal, ordinal, decimal numbers) are to be written as complete words, not as digits:

*I was born in nineteen seventy five. let's meet at two thirty.*

### Foreign words, dialect, mispronunciations

- Type foreign and dialectal words (the words which do not exist in the vocabulary of the spoken language) inside square brackets. If you are unsure whether that word exists in the vocabulary or not, opt for using square brackets:

*I am going to [Kaserak].*

### Punctuation

- Use only the following marks for punctuation:  
dot (.) question mark (?) exclamation mark (!) apostrophe (') hyphen (-)
- Use a dot, a question mark or an exclamation mark to mark the end of a sentence:

*it is great to hear from you! how are you? I think you have been busy lately.*

- Use a dot at the end of a segment or before the beginning of a new utterance, even if the sentence is not grammatically complete:

*but I. if you. because.*

- Use an apostrophe for writing possessives and for contractions:

*let's meet at Mark's house at two o'clock.*

- Decide whether to use contraction or not depending on the pronunciation and the intended meaning:

*I'm fine. how are you?*

- Use a hyphen between words for compounding:

*soon-to-be son-in-law*

- Do NOT use any other punctuation mark, for example:

comma (,) colon (:) semicolon (;) single quotes ( ' ' ) or double quotes ( " " ), etc.

*he said I bought apples oranges and pears.*

### Disfluent speech

Disfluent speech is particularly difficult to transcribe. Speakers may stumble over their words, repeat themselves, utter partial words, restart phrases or sentences, and use hesitation sounds (filled pauses) to maintain control of a

conversation while thinking of what to say next. Do not spend too much time trying to precisely capture difficult sections of disfluent speech. Make the best effort to transcribe what you hear after listening to the segment once or twice, and then move on.

- Type all repetitions as they are heard

*but it is but it is it is it is unacceptable.*

- Mark interjections like exclamations (e.g., Oh!), response particles (Huh?), hesitation markers and filled pauses (ah, er, uh, hmm) with an asterisk (\*), preceded and followed by a space:

*I was \* I was going to say something \* let me see.*

- Some of the most common interjections in spoken language are the words “okay” denoting approval, acceptance or indifference; “yeah” and “yep” denoting agreement; and “wow” denoting excitement or surprise. Type these words as they are shown below:

*OK O.K. O\_K Okay à okay yeah yep wow*

- Mark every individual unrecognizable, unknown, or not thoroughly pronounced, stuttered or fragmented part of a word with a plus (+), without spacing:

*we observe deteri+ det+ detert+ deterioration all around us today.*

*+ication is a word which is commonly used in medicine.*

*then I told him to give me + Euros.*

### Non-speech sounds, noises

- Mark any orally created sounds like breathing, lip-smacking, throat-clearing, laughter, coughing, burping, or a grunt, bump, sniff, scratch, shriek, whatever, with an asterisk (\*)
- Do NOT Mark any other external (background, environmental) sound or noise of short duration happening at the same time with speech, such as phone ringing, rustling, whistling, traffic noise, someone slamming a door or dropping something, etc.
- Sections which contain no or only little intelligible speech, such as several persons speaking at the same time, speech with long loud background noise, long-lasting sounds, telephone ringing sounds, or long sections in a foreign language or dialect, must be put into a separate segment and marked it with a single hyphen (-) without the parentheses.

## Annex 2 - ROXSD Annotation Guidelines

This section outlines the annotation procedure undertaken for the NLP annotations for ROXSD datasets. In particular, four types of annotations are made on each conversation: Named Entity Recognition (NER), Mention Disambiguation (MD), Relation Extraction (RE) and Topic Detection (TC).

The annotations are maintained in a .yaml format from corresponding .lab files (where transcriptions are stored). Each .yaml file contains the transcript of a conversation along with the entire set of NLP annotations made for it. The annotations for NER and Mention Disambiguation are maintained inline while RE and Topic annotations are maintained externally. An example is illustrated below.

XXX.lab

```
0.000000    0.471126    B:ru01M_T:english:yes?
1.643194    2.642899    A:cs01M_T:english:where are you?
3.677077    5.435180    B:ru01M_T:english:I've already gone home.
6.147613    9.135239    A:cs01M_T:english:what? we have to work. where are you going?
9.411019    15.753978   B:ru01M_T:english:I'm going to [Kaserak] but I'm not alone. he leaves and gives me weed. I was next to
[Hlavák]. two minutes ago.
16.236594   16.937537   A:cs01M_T:english:okay.
```

XXX.yaml

```
Transcript: "0.000000    0.471126    B:ru01M:english:yes?
1.643194    2.642899    A:cs01M:english:where are you?
3.677077    5.435180    B:ru01M:english:I've already gone home.
6.147613    9.135239    A:cs01M:english:what? we have to work. where are you going?
9.411019    15.753978   B:ru01M:english:I'm going to [Kaserak|LOC] but I'm not alone. he leaves and gives me weed. I was
next to (Hlavák|LOC). two minutes ago.
16.236594   16.937537   A:cs01M:english:okay"
"
Topic: Drug Conversation
Relations: ['<sub> I <loc> Kaserak <rel> go']
```

The following sections detail the guidelines kept in mind while annotating ROXSD conversations.

### Named Entity Recognition (LOC/TIME)

We annotate person names (PER), location names (LOC) and time (TIME) of a given text. This section introduces the criteria how we annotate LOC and TIME. As the person names are also useful for Mention Disambiguation module, we describe the annotation criteria in the Mention Disambiguation section below.

We define “LOCATION” as:

- Geopolitical entities: countries, cities, states.
- General locations: mountain ranges, bodies of water, street names.

#### Examples

- So I think our trip to **Germany** from **Czech Republic** might be even easier than we thought originally.
- let's go to **Mount Fuji**
- I will wait at **Eiffel Tower**
- let's meet at **Saarland University**
- Let's go to **Starbucks**

We define “TIME” as:

- Date: Absolute or relative dates or periods.
- and times smaller than a day.

#### Examples





- He will leave **tomorrow at 9PM**
- The show will start **tomorrow morning at 10:00**
- I already did this **last year**
- it will take **2 minutes** to break the door
- My flight is on **March 23**
- I sold them last **Sunday**
- I sold them **the day before yesterday**
- **Yesterday** was **Friday** and **the day after tomorrow** is **Monday**.
- I saw him **30 seconds ago**

### Mention Disambiguation

Mention disambiguation is annotated for Person entities appearing in the call. For each person entity, please make an annotation if the person is a

**PARTY** : Refers to one of the participants in the call

**THIRD\_PARTY** : Refers to people who are not speaking in the call but are mentioned in the duration of the call

For intuition, while annotating every PER entity, annotate it with one of the labels given above instead.

### Relation Extraction

Relation Extraction defines two kinds of annotations:

1. **Go** : A person is going to a location
2. **In** : A person is currently at a location

Each relation has a Person Token and a Location Token. The task is to annotate a pair of entities **and** the corresponding relation.

The annotation is made in a “<sub> subject <loc> location <rel> relation\_type” format

Example:

I'm in Munich -> “<sub> I <loc> Munich <rel> in”

Annotation Guidelines:

1. There needs to be a PER and a LOC entity pair for every RE annotation. Make sure all relations are triplets.
2. We don't care about objects moving to locations in the given context. While there is some leniency with respect to PER annotations, location entities are screened more cautiously.
3. The **in** tag should be given only if someone implies **explicitly** that he/she/they are at a location.



4. Include pronouns while annotating subjects.:
  - a. Example : I am going to Paris → <sub> I <loc> Paris <rel> go
  - b. we are in Stockholm → <sub> we <loc> Stockholm <rel> in
5. Avoid annotating anything more than the subject when annotating the entity, e.g.
  - a. I’m going to Paris → <sub> I <loc> Paris <rel> go
  - b. I am going to Paris → <sub> I <loc> Paris <rel> go
6. Locations strictly mean “Places”. i.e., schools, buildings etc. do not count. Please refer back to the NER annotation guidelines for a better understanding on what qualifies a Location entity. Please note that utterances like ‘I am here’ or ‘I will be there’ should be ignored. Locations should not be ambiguous.

## Topic Detection

The topics of a conversation are annotated. Currently, six pre-defined topics are supported. They are:

- Drug
- Work
- Family-Friend
- Meeting
- Money
- Others

Example (A conversation about drugs):

Speaker A: Hi, how are you doing? Let’s meet tomorrow. I will bring some good stuff for you.

Speaker B: Hi, I’m good, sure. If possible, bring 10 more than the previous time.

## Annex 3 - Informed consent

### PARTICIPANT INFORMATION SHEET

#### The ROXANNE Project

You are being asked to participate in a research activity for the ROXANNE project. ROXANNE is a research project funded by the European Union that intends to combine new speech technologies, face recognition, text analysis and network analysis into a novel platform that will assist law enforcement agencies to identify criminals in organised crime investigations. Further information is available on the project website: <https://www.roxanne-euproject.org/>

The ROXANNE partners are: Fondation de L’Institut de Reserche IDIAP (CH); Trilateral Research Ltd. (UK); Vysoke Uceni Technicke v Brne (CZ); Phonexia SRO (CZ); HENSOLDT Analytics (AT); Capgemini Consulting (FR); The International Criminal Police Organisation (INTERPOL, an international organisation); Univrsitat des Saarlandes (DE); Kentro Meleton Asfaleias (GR); Gottfried Wilhelm Leibniz Universitaet Hannover (DE); Universita Cattolica del Sacro Cuore (IT); Aegis IT Research UG (Haftungsbeschrant) (DE); Airbus Defence and



Space SAS (FR); Policejní Prezidium České republiky (CZ); Ministerul Afacerilor Interne (RO); Lietuvos Teismo Ekspertizės Centras (LT); Police Service of Northern Ireland (UK); Aditess Advanced Integrated Technology Solutions & Services Ltd (CY); Ministry of Interior (HR); Netherlands Forensic Institute (NL); Internet of Things Applications and Multi Layer Development Ltd (ITML, CY); Ministry of Public Security (IL); Hellenic Police (GR); An Garda Síochána (IE); Zentrale Stelle für Informationstechnik im Sicherheitsbereich (ZITiS, DE).

### **Purpose of the research activity**

The workshop you are attending intends to collect data in similar ways to how law enforcement agencies collect data during their investigations. There are three separate data collection activities occurring at this workshop that will gather: (1) voice data; (2) image and video data, and; (3) text data. These recordings will be used to expand a dataset created by the ROXANNE partners. This dataset is called ROXANNE Simulated Data, or ROXSD. These data will be used to develop, improve, and test computer models which will be used to research how people communicate and interact in crime-related settings; these models could later be used by law enforcement agencies for identifying and linking suspected criminals.

This research activity is being led by Phonexia. Phonexia will be the Data Controller for all data collected during the workshop. Equipment and infrastructure are provided by ZITiS for all data collections. Organisational support for the workshop is provided by HENSOLDT Analytics, who are also managing the voice, image, and video data collection; ITML are managing the text data collection (all ROXANNE partners).

### **Overall procedure of the research activities**

You will not be pressured or coerced into participating in any research activity. You are free to leave at any point, with no negative consequences.

If you choose to participate in a research activity, it will involve you being asked to play a fictitious criminal character and engage in creating data that will be recorded by ROXANNE researchers. You might be playing the same character you have previously portrayed. You should not reveal any personal details about yourself during the data collection activities.

The collected data you provide will be used by ROXANNE partners to either help improve, or test, computer models. Administrative information such as your name and email address will not be stored with the collected research data and the researchers using the research data will not be able to identify you directly, unless you are featured in an image/video and are known to other ROXANNE researchers.

You will be provided with any equipment needed to engage in the data collection activities. If you wish, you can also use your own devices; any information linking your devices to the collected data will be removed before the data is used for research purposes.

Your data will be stored securely at all times. After your data has been collected and pre-processed for incorporation into the ROXSD dataset, it will be transmitted to Phonexia who will collate all the collected data and distribute the expanded dataset to ROXANNE partners.

### **Procedure for voice data recordings**

In order that your voice can be recorded, ROXANNE researchers have set up mobile telephones to be intercepted using real law enforcement methods. You will be asked to use one of these mobile telephones to make a call to another mobile telephone that has also been set up for interception, or you can choose to call these telephones from your own telephone. If you choose to use your own telephone, information about it will be changed by researchers once the



voice data has been collected. You can consent to using your own telephone on the consent form below. You will be asked to engage in an unscripted telephone call whilst playing a suspected criminal character from the project scenario.

Collection of voice data through an interception will occur using real law enforcement methods where the data is requested by ZITiS, who are a law enforcement research agency, from a mobile telephone network provider. The data will then be collected by ZITiS (your data will be kept separately from those collected in real investigations) and transferred to Phonexia who will process the data so that it can be used for research by other ROXANNE partners.

### **Procedures for image and video data collections**

In order to collect images and videos, you will be asked to use a device provided by ROXANNE partners, or you can use your own device, to take photos and videos of locations and objects that are relevant to the project scenario. You will be instructed about what locations/objects ROXANNE researchers want included in the data.

ROXANNE partners would also like to collect image/video data from locations beyond the workshop location. This will involve you using your own device(s). If you would like to engage in collecting additional data on your journey home from the workshop, you should speak to ROXANNE researchers at the workshop and provide consent for this, and for using your own device(s) on the form below.

You can include yourself and other workshop participants in the images and videos if they have consented. If you record an image/video with featuring project participants, they should be playing characters from the project scenario. You should not include any person who has not consented, or are unconnected to the ROXANNE project.

### **Procedures for text data collection**

In order that the text data you provide in the conversations can be collected, ITML (ROXANNE partner) have set up a web-forum that you will be asked to type messages and replies into. You will be provided with access to a computer to contribute to these conversations, or you can choose to use your own device. You will be asked to write messages whilst playing a criminal character from the project scenario.

If you wish, you can also provide additional data after the workshop using your own device(s). If you are interested in this, you should speak to ROXANNE researchers during the workshop and provide consent for this, and for using your own device(s) on the form below.

Once all text data has been collected, it will then be expanded using text generation tools and collected using a web crawler that is like those available to law enforcement.

### **Use of the collected in other research activities**

Following the processing and collating of the collected data into an expanded ROXSD dataset, and using it to develop, improve, or test computer models during the ROXANNE project, ROXANNE researchers will report their results in a piece of work submitted to the European Commission and, potentially, scientific research papers. These results will not include any information that could identify you. By consenting to participate in this research activity, you are consenting to the results of this research activity being sent to the European Commission and published in scientific research papers.

In order to develop, improve, and test other computer models, ROXANNE researchers would like to make the ROXSD dataset available online to be used by other academic or industrial organisations for research purposes. If



you wish to opt in to making the data you provide available for other researchers, you can provide consent for this on the attached form. If you consent, information on your gender, age range, and the languages you speak will be included in this dataset. Your name, signature, and contact details will not be included in this dataset.

### **What personal data we will collect**

In this research activity, we will only process your personal data when you consent. You can remove your consent at any time. You can provide separate consent to each data collection activity below. By signing the informed consent form below, you consent to the processing of the following personal data:

- **Your name:** This is collected so that we can address you properly and we can facilitate the exercise of your rights about your personal data, it will be recorded on your informed consent form.
- **Your signature:** This is collected to confirm that you understand the meaning of giving your consent to participate in the research activity, it will be recorded on your informed consent form.
- **A recording of your voice:** These data are recorded so that we can process them to develop, improve, or test computer models and research how people communicate and interact in crime-related situations.
- **Images and/or videos featuring you:** These data are recorded so that we can process them to develop, improve, or test computer models and research how people communicate and interact in crime-related situations.
- **Text written by you:** These data are recorded so that we can process them to develop, improve, or test computer models and research how people communicate and interact in crime-related situations.
- **Your gender, age range, and details of the languages you speak:** These data are collected on your informed consent form so that we can correctly classify the data you provide in order to analyse them accurately.
- **Your email address:** This is collected in case we need to contact you about your participation. It will not be used for marketing purposes unless you consent.

Your personal data (i.e. data that could be used to identify you) will be processed according to applicable data protection law of the European Union (the General Data Protection Regulation, or GDPR) and of the Czech Republic (where Phonexia are based).

Physical copies of the informed consent form (below) will be stored under lock and key by Phonexia. Researchers who use the voice recordings will only have access to research data you provide and information on your gender, age range, and details of the languages you speak. They will not have access to administrative information including your name, signature, email address, or telephone number.

Where your data is processed on computers, they will only be processed on machines that are secured using password-protection and encryption. Your personal data will be destroyed when they are no longer needed, and in any case five years after the end of the project (whichever is sooner), unless you consent to your data to be made available for future researchers.

### **Rights regarding your personal data**

In accordance with principles of research ethics and EU data protection regulations, you have rights regarding how your personal data is processed. Here are your rights and how we can fulfil them:

- **Rights to access personal data processed about you, and the right for these data to be in a portable form** – If you request access to personal data that we hold about you, we will provide you with that data in an easily accessible format.
- **Right to rectify personal data held about you** – If you think the personal data that we hold about you is inaccurate or incomplete, you can correct or complete it.
- **Right to restrict the processing of your personal data** – If you want to restrict the way we process your personal data, you can request that we do so.



- Right to request your personal data is erased – If you want us to delete your personal data from our systems, you can request that we do so.
- Right to leave the research activity – If you wish to withdraw from having your data collected, you can do so at any time without negative consequences and we will delete any personal data you have provided.
- Right to complain to a supervisory authority – If you feel we have not adequately dealt with your requests, you can complain to the national data protection authority. This research activity is led by Phonexia in the Czech Republic, and you can find information on The Office for Personal Data Protection: <https://www.uoou.cz/en/>.

We aim to facilitate the rights of all data-subjects. But, we may reject some requests in accordance with data protection legislation.

You can find more information about our data processing here: <https://www.roxanne-euproject.org/data-protection-notice.pdf>

### **Potential risks to participating in this research activity**

Where you are known to ROXANNE researchers (or other researcher in future), you could, potentially, be identified from your voice, image or physical features, and writing style. The ROXANNE consortium foresees no physical or psychological risks which could materialise from your participation in this research activity. However, if your biometric data (e.g., your voice or image) is used to develop or improve a computer model, this could likely lead to a slightly increased chance of you being identified if you were to be suspected of a crime, placed under surveillance, and recordings of your activities are analysed by law enforcement officers using the ROXANNE tools. If you do not feel comfortable taking these risks, you should withdraw from the data collection activity. You will not face any negative consequences if you withdraw.

### **Who to contact if you have questions**

If you have any questions or queries about this research activity, or the processing of your personal data, you can use the following contact details:

- For more information on these data collection activities, you should contact Marek Kováč, Senior Pre-Sales Engineer at Phonexia [marek.kovac@phonexia.com](mailto:marek.kovac@phonexia.com)
- For more information on the processing of your personal data during these research activities, you should contact Mr. Michal Jirka, Data Protection Officer at Phonexia, [michal.jirka@phonexia.com](mailto:michal.jirka@phonexia.com)
- For more information on the ROXANNE project, you should contact Dr. Petr Motlicek, ROXANNE Project Coordinator: [petr.motlicek@idiap.ch](mailto:petr.motlicek@idiap.ch)
- For more information on the processing of personal data in the ROXANNE project generally, you should contact Dr. Joshua Hughes, Research Analyst at Trilateral Research [Joshua.hughes@trilateralresearch.com](mailto:Joshua.hughes@trilateralresearch.com)

## **INFORMED CONSENT FORM**

You are being asked to engage in providing research data for ROXANNE researchers whilst playing a fictitious character during a telephone call, being captured in photographs or videos, and/or writing text chat messages.

### **Research data:**



Gender:
Age range: <input type="checkbox"/> 18-24 <input type="checkbox"/> 25-30 <input type="checkbox"/> 31-35 <input type="checkbox"/> 36-40 <input type="checkbox"/> 41-45 <input type="checkbox"/> 46-50 <input type="checkbox"/> 51-55 <input type="checkbox"/> 56-60 <input type="checkbox"/> 61-65 <input type="checkbox"/> 66-70 <input type="checkbox"/> 71+
Native language:
Languages spoken at a conversational level:

**Research and data collection activities:**

	Yes	No
<b>Voice data collection</b>		
<i>Do you consent to having your voice recorded during this research activity?</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Do you consent to using your own mobile telephone during this data collection activity?</i>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Image and video data collection</b>		
<i>Do you consent to having your biometric information (e.g., voice, image, physical features) recorded during this research activity?</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Do you consent to using your own device(s) during this data collection activity?</i>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Text data collection</b>		
<i>Do you consent to having text data you provide recorded during this research activity?</i>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Do you consent to using your own device(s) during this data collection activity?</i>	<input type="checkbox"/>	<input type="checkbox"/>

**Use of data for future research:**

As explained in the Information Sheet, ROXANNE researchers would like to use the data collected during this activity to expand a dataset that is intended to be shared for research purposes. This could include the research data (information on gender, age range, and languages spoken), recordings of your voice, images and videos you have recorded, or featuring you, and text data you provide. Your name and email address will not be included in this dataset.

	Yes	No
<i>Do you consent to the research data you provide during this activity being included in a dataset for use in future research?</i>	<input type="checkbox"/>	<input type="checkbox"/>

By signing below, I confirm that I understand the following:



- I have read the information sheet explaining the project and understand how this research activity will collect and process my data.
- I have the right to ask questions about my participation and receive clear answers before making any decision.
- My participation is voluntary, and I have not been pressured or coerced in any way to participate in this research. I am free to end my participation at any time.
- I am being asked to participate in having my data collected, and these data will be stored securely.
- Administrative information such as my name, signature, and email address will be securely stored and kept separately from research data.
- My personal data will be destroyed when they are no longer needed, and in any case five years after the end of the project (whichever is sooner), unless I consent to my data being made available to other researchers in future.
- The results of processing my data might be included in the writing of a piece of work to be delivered to the European Commission, or scientific research papers.
- I have been made aware of any risks associated with participating in this activity.
- I have been made aware of my rights regarding my personal data.
- I have been given the contact details of the research team and I have been informed that I am free to contact them with any queries about the research or the project:

#### Administrative information:

Full Name:
E-mail address:
Date:
Signature:

## Annex 4 - License

*Note: The following text is used as tentative “License and Data Processing Agreement” and is enclosed to ROXSD v3 data set. In the future, there might be changes in the text. Please check the most up to date text in ROXSD v3 data set.*

# LICENSE AND DATA PROCESSING AGREEMENT

(Controller to Controller)

#### Between

**IDIAP**, having its registered office at Centre du Parc, Rue Marconi 19, CH - 1920 Martigny, Switzerland,  
**VYSOKE UCENI TECHNICKE V BRNE (BUT)**, having its registered office at Antoninska 548/1,601 90 Brno, Czech Republic,  
**GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER (LUH)**, having its registered office at Welfengarten 1, HANNOVER 30167, Germany,  
**HENSOLDT Analytics GmbH (SAIL)**, having its registered office at Mariannengasse 14, A-1090 Vienna, Austria,  
**Phonexia s.r.o.** having its registered office at Chaloupkova 3002/1a, 612 00 Brno, Czech Republic,  
**AEGIS**, having its registered office at 25 Humboldt Str. Braunschweig, 38106, Germany  
(“**Controllers**”)





And

**ORGANIZATION RECEIVING THE DATA  
("Partner")**

**PROCESSING DETAILS**

**Background**

ROXANNE was a research project funded by the European Union<sup>[1]</sup> that intended to combine new speech technologies, face recognition, text analysis and network analysis into a new platform that will assist law enforcement agencies to identify criminals in organised crime investigations. Further information is available on the project website: <https://www.roxanne-euproject.org/>.

Controllers recruited speakers (among their employees and external collaborators) that performed calls based on several prepared scripts resembling real calls as close as possible to the real conversations between suspected persons that were recorded ("Dataset"). All recruited speakers have given their consent and were informed about the purpose of recording, the way of storing the data, and the data validity time. The Dataset consists of audio recordings, metadata, transcriptions, translations, geolocation, and other types of information. All information is fictionally based on the artificially elaborated story. Dataset was pseudonymized.

Controllers are the copyright holders of the Dataset and joint controllers of the voice recordings as Personal Data.

Controllers has agreed to provide Personal Data to the Partner and the Partner has agreed to use the Personal Data only for the Purposes set out in these Processing Details and the parties shall each be a Controller of that Personal Data.

This Controller-to-Controller Processing Agreement sets out the terms on which Controllers may provide, and the Partner will use, the Personal Data. This Agreement consists of the below Processing Details and the Processing Conditions, including the Schedule.

<b>Parties</b>	
Controllers	<b>Phonexia s.r.o.</b> having its registered office at Chaloupkova 3002/1a, 612 00 Brno, Czech Republic, <b>VYSOKE UCENI TECHNICKE V BRNE (BUT)</b> , having its registered office at Antoninska 548/1, 601 90 Brno, Czech Republic, <b>GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER (LUH)</b> , having its registered office at Welfengarten 1, HANNOVER 30167, Germany, <b>HENSOLDT Analytics GmbH (SAIL)</b> , having its registered office at Mariannengasse 14, A-1090 Vienna, Austria, <b>IDIAP</b> , having its registered office at Centre du Parc, Rue Marconi 19, CH - 1920 Martigny, Switzerland, <b>AEGIS</b> , having its registered office at 25 Humboldt Str. Braunschweig, 38106, Germany
Partner	<i>[Note: Insert the name and registered address of the Partner to whom the Controllers will provide personal data under this agreement]</i>

<b>Processing Details</b>	
Purposes	• <i>Non-commercial research activities</i>
Sharing Start Date	<i>[Note: insert the data sharing start date. The agreement will apply from that date.]</i>
Personal Data provided by (or on behalf of) Controllers	<i>Audio recordings</i>
Categories of Data Subjects	<i>Audio recording participants</i>
Controllers's legal grounds to share Personal Data	<i>Informed Consent of Data Subjects</i>
Partner's legal grounds to Process Personal Data	<i>Informed Consent of Data Subject</i>
Access and Processing restrictions	<i>[N/A]</i>



Security Requirements	Personnel security requirements: <i>[N/A]</i> Technical security requirements: encryption, password protection Physical security requirements: <i>[N/A]</i> Any other agreed security requirements: <i>[N/A]</i>	
Review of this Agreement	once around the anniversary of the Sharing Start Date	
Controllers Contacts	<i>[Note: replace with Data protection Officer contact]</i>	<i>[Note: replace with Data protection Officer contact]</i>

Partner Contact	<u>Name:</u> <i>[Note: insert Partner contact name]</i> <u>Position:</u> <i>[Note: insert Partner contact role]</i> <u>Address:</u> <i>[Note: insert work address of contact]</i> <u>E-mail:</u> <i>[Note: insert Partner contact e-mail]</i> <u>Cc:</u> <i>[Note: insert Cc e-mail address for contact]</i> <u>Telephone:</u> <i>[Note: insert contact phone number]</i>
-----------------	--

Signed for and on behalf of  
THE CONTROLLERS

.....  
Authorised Signatory

by (Insert name)  
on (Date of signing)  
at (Place of signing)

.....

Signed for and on behalf of  
*[Note: Insert the full name of the Partner. This should be the same as above.]*  
.....

Authorised Signatory

by (Insert name)  
on (Date of signing)  
at (Place of signing)

## **LICENSE AND PROCESSING CONDITIONS**

### **1. License conditions**

- a. Controllers grant to the Partner non- exclusive, time-limited, royalty free and non-sublicensable right to use the Dataset only for non-commercial scientific research only. Controllers grant to the Partner the right to rework and build upon the Dataset, or any component thereof, as necessary or desirable for research or technology development activity and create derivative products or services for the Partner’s own internal research and development.
- b. The Dataset is provided AS IS and AS AVAILABLE without any warranty. Controllers make no representations or warranties of any kind concerning the Dataset, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to the Partner.
- c. To the extent possible, in no event will Controllers be liable to the Partner on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of the license and use of the Dataset, even if Controllers have been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to the Partner.

### **2. General compliance obligations**

- a. This Agreement sets out the framework for the provision of Personal Data by Controllers to the Partner. It sets out the principles and procedures that the parties shall adhere to and the responsibilities the parties owe to each other in connection with sharing of Personal Data.
- b. The parties are Controllers in common in connection with Personal Data provided under this Agreement, and are not joint Controllers.
- c. Each party shall comply with all the obligations imposed on a Controller under the Data Protection Laws. The Partner acknowledges that certain Data Protection Laws may apply to the Partner regardless of the location of the Partner.

### **3. Purposes and the types of Personal Data provided by (or on behalf of) Controllers**

- a. The Partner agrees to Process the Personal Data provided by (or on behalf of) Controllers only for the Purposes.
- b. The Personal Data provided by (or on behalf of) Controllers must not be irrelevant or excessive with regard to the Purposes.
- c. The types of Personal Data which may be provided by (or on behalf of) Controllers and the relevant categories of Data Subjects to whom such Personal Data relates are outlined in the Processing Details. Controllers is not obliged to provide Personal Data to the Partner.
- d. Controllers shall exercise reasonable endeavours to ensure that the Personal Data provided by it is accurate and up-to-date at the time of providing such Personal Data. Where the Partner becomes aware that the Personal Data provided by (or on behalf of) Controllers is no longer accurate or up-to-date, it shall promptly inform Controllers of such inaccuracy and provide the relevant accurate or updated Personal Data to Controllers.

### **4. General obligations**

- a. Without prejudice to clause 1.3 (*General compliance obligations*), the Partner shall:
  - i. comply with the access and Processing restrictions outlined in the Processing Details;
  - ii. subject to clause 3.1(a), not disclose or transfer the Personal Data provided by (or on behalf of) Controllers, unless such disclosure is:
    - to the Partner’s employees whose reliability the Partner has taken reasonable steps to ensure, on a need-to-know basis;
    - to any third party Processing Personal Data on behalf of the Partner; or

- to any other third party (other than the parties mentioned in clauses 3.1(b)(i) or 3.1(b)(ii) above) with Controllers's prior written consent

always provided that such disclosure or transfer is made for the Purposes, subject to an obligation of confidentiality and in compliance with Data Protection Laws and, in connection with clause 3.1(b)(ii) and 3.1(b)(iii), is made under a written legally binding agreement between the Partner and the relevant recipient of Personal Data which is no less onerous than this Agreement (including in connection with confidentiality);

- iii. ensure that it has all necessary notices and consents in place to enable lawful transfer of the Personal Data in accordance with clause 3.1(b);
- iv. assist Controllers in complying with all applicable requirements of Data Protection Laws;
- v. notify Controllers immediately if it becomes aware of any disclosure, loss or destruction, or Processing of any Personal Data that has been made or is likely to be made in breach of the terms of this Agreement or the Data Protection Laws; and
- vi. maintain complete and accurate records and information to demonstrate its compliance with this Agreement and allow for audits by Controllers or Controllers's designated auditor.

## 5. Privacy notices and legal grounds for Processing

- a. Controllers shall ensure that it has all necessary notices and consents in place to enable lawful provision of the Personal Data to the Partner for the Purposes.
- b. The Partner shall ensure that it Processes Personal Data provided by (or on behalf of) Controllers on the basis of one or more of the legal grounds outlined in the Processing Details.

## 6. Security and training

- a. The Partner shall ensure that:
  - i. it has in place appropriate technical and organisational measures (as defined in the Data Protection Laws), to protect against unauthorised or unlawful Processing of Personal Data provided by (or on behalf of) Controllers under this Agreement and against accidental loss or destruction of, or damage to, such Personal Data and shall, without prejudice to the generality of the foregoing, comply with the Security Requirements outlined in the Processing Details;
  - ii. its staff members are appropriately trained to handle and Process the Personal Data provided by (or on behalf of) Controllers in accordance with the required technical and organisational security measures together with applicable Data Protection Laws;
  - iii. it uses compatible technology in connection with Personal Data provided by (or on behalf of) Controllers, to ensure there is no lack of accuracy resulting from transfers of Personal Data; and
  - iv. it keeps the Personal Data provided by (or on behalf of) Controllers separate from all other data/information held, stored or otherwise Processed by or under the control of the Partner.

## 7. Rights of Data Subjects

- a. The parties agree to comply with their obligations to respond to Data Subject Access Requests and to give effect to other rights of Data Subjects in accordance with Data Protection Laws.
- b. The Partner shall:
  - i. promptly inform Controllers about the receipt of any Data Subject Access Request or other query, request, dispute or complaint from a Data Subject which relates to Personal Data provided by (or on behalf of) Controllers or otherwise Processed under this Agreement and shall provide Controllers with a copy of the same;



- ii. provide Controllers with assistance in complying with any Data Subject Access Request, query, request, dispute or complaint; and
- iii. in relation to Data Subject Access Requests, not disclose or release any Personal Data provided by (or on behalf of) Controllers in response to a Data Subject Access Request without first consulting Controllers, wherever possible, and shall retain a record of information provided to a Data Subject.

## 8. Personal Data Breaches

- a. The Partner shall provide Controllers assistance in complying with all applicable requirements of the Data Protection Laws relating to Personal Data Breaches. In particular, the Partner shall:
  - i. promptly (and in any case within 24 hours) notify Controllers as soon as it becomes aware of a Personal Data Breach or a suspected Personal Data Breach which relates to Personal Data provided by (or on behalf of) Controllers;
  - ii. take reasonable and prompt steps to mitigate the impact of the Personal Data Breach;
  - iii. co-ordinate with Controllers to investigate the actual or suspected Personal Data Breach, its extent and any parties responsible for the Personal Data Breach. Such investigation should consider if the processes outlined in this Agreement have been compromised and whether this Agreement has been breached. The Partner shall provide details of the Personal Data Breach to Controllers on request; and
  - iv. where in receipt of notification from Controllers in respect of a Personal Data Breach, provide reasonable assistance to Controllers to ensure Controllers is able to deal with such Personal Data Breach in an expeditious and compliant manner.

## 9. Dealing with Supervisory Authorities

- a. The Partner shall promptly notify Controllers of any dispute, claim or query brought by any Supervisory Authority concerning the Processing of Personal Data provided by (or on behalf of) Controllers and shall take into account any comments and representations made by Controllers when responding to such dispute, claim or query.
- b. The Partner agrees to co-operate and provide all reasonable assistance and information to Controllers in dealing with any dispute, claim or query brought by any Supervisory Authority in connection with this Agreement, with a view to settling them amicably and in a timely fashion.

## 10. Data Protection Impact Assessment

- a. Each Party shall be responsible for adopting and maintaining a data protection impact assessment relevant to its Processing activities in accordance with Data Protection Laws.
- b. The Partner agrees to provide Controllers reasonable assistance on request from Controllers in connection with preparation and updating of Controllers's data protection impact assessment.

## 11. Retention and deletion of Personal Data

- a. The Partner shall not retain or otherwise Process the Personal Data provided by (or on behalf of) Controllers for longer than is necessary for the Purposes.
- b. The Partner shall ensure that Personal Data provided by (or on behalf of) Controllers are either returned to Controllers or securely deleted or destroyed in accordance with Data Protection Laws as Controllers may instruct, in the following circumstances:
  - i. on request by Controllers;
  - ii. on termination or expiry of this Agreement (unless there is an obligation to Process such Personal Data under applicable laws); and
  - iii. once Processing of the Personal Data is no longer necessary for the Purposes.

## 12. Term and termination of this Agreement

- a. This Agreement will be deemed to come into effect on the date of the Sharing Start Date mentioned in the Processing Details (regardless of the date(s) of signing of this Data Processing Agreement) and will remain in full force and effect for so long as the Partner retains any of the shared Personal Data in its possession or control unless otherwise terminated in accordance with its terms.



- b. Controllers may terminate this Agreement by giving the Partner no less than 10 days' written notice to that effect.
- c. Controllers may terminate this Agreement immediately by giving the Partner written notice to that effect in the following circumstances:
  - i. the Partner notifies Controllers pursuant to clause 3.1(e) (*General Obligations*);
  - ii. the Partner has breached Data Protection Laws in connection with either this Agreement or the Personal Data provided by (or on behalf of) Controllers and such breach is either not capable of remedy or is not remedied within 10 days of written notice from Controllers;
  - iii. the Partner is not Processing the Personal Data provided by (or on behalf of) Controllers in accordance with this Agreement; or
  - iv. the Partner acts in any manner which brings or is likely to bring Controllers into disrepute or is materially adverse to the interests of Controllers.
- d. Any provision of this Agreement that expressly or by implication is intended to come into or continue in force on or after termination of the Agreement will remain in full force and effect. This includes clause 14 (*Liability and indemnity*).
- e. The termination of this Agreement will be without prejudice to any other rights or remedies of any party under this Agreement or at law and shall not affect any claims or rights which a party may have against the other which have accrued prior to such termination.

### 13. Review of Agreement

- a. The parties shall review the effectiveness of this Agreement at such times as indicated in the Processing Details, having consideration to the Purposes. The Parties shall continue, amend or terminate this Agreement depending on the outcome of such review. Such review will include:
  - i. assessing whether the purposes for which the Personal Data is being Processed are still the Purposes listed in the Processing Details;
  - ii. assessing whether the Personal Data provided by Controllers is still as outlined in the Processing Details;
  - iii. assessing whether the legal framework governing data quality, retention, and Data Subjects' rights are being complied with; and
  - iv. assessing whether Personal Data Breaches involving the Personal Data provided by (or on behalf of) Controllers have been handled in accordance with this Agreement and the applicable legal framework.
- b. In addition to the review carried out under clause 12.1, in case the Agreement is no longer adequate for the purpose of governing lawful data provision by Controllers, the parties agree to negotiate in good faith amendments to the Agreement in light of Controllers's reasonable requirements.

### 14. Cross-border Transfers of Personal Data

- a. The Partner (or any third party Processing Personal Data on behalf of the Partner) must not transfer any Personal Data provided by (or on behalf of) Controllers outside the EU or EEA territory unless the Partner:
  - i. complies with the provisions of Article 26 of the GDPR (in the event the third party is a joint Controller); and
  - ii. ensures that (i) the transfer is to a territory which is subject to adequacy regulations under the Data Protection Laws that the territory provides adequate protection for the privacy rights of individuals ; or (ii) the Partner participates in a valid cross-border transfer mechanism under the Data Protection Laws, so that the Partner can ensure that appropriate safeguards are in place to ensure an adequate level of protection with respect to the



privacy rights of individuals as required by Article 46 (Transfers subject to appropriate safeguards) of the GDPR; or (iii) one of the derogations for specific situations in Article 49 of the GDPR applies to the transfer.

- b. If any Personal Data transfer between Controllers and the Partner requires execution of SCC in order to comply with Data Protection Laws (where Controllers is the entity exporting Personal Data to the Partner outside the EU), the parties will complete all relevant details in, and execute, appropriate SCC provided by (or on behalf of) Controllers for such purpose, and take all other actions required to legitimise the transfer.

## 15. Liability and indemnity

- a. The Partner shall indemnify and keep Controllers indemnified in full and on demand in respect of any direct losses, claims, costs, expenses, proceedings and damages incurred by Controllers arising out of or in connection with the negligence, bad faith, and/or wilful misconduct of the Partner in relation to this Data Processing Agreement or Data Protection Laws.
- b. Nothing in this Agreement will:
  - i. limit or exclude any liability for death or personal injury resulting from negligence; or
  - ii. limit or exclude any liability for fraud or fraudulent misrepresentation; or
  - iii. limit any liabilities in any way that is not permitted under applicable law; or
  - iv. exclude any liabilities that may not be excluded under applicable law.

## 16. Notices

- a. Any notice given under this Agreement will be in writing and will be served personally, by first class post or (domestic/international) courier to the address of the relevant party's contacts outlined in the Processing Details (including both Controllers Contacts). Any notice so given will be deemed to have been duly served if:
  - i. personally delivered, on the day of delivery; or
  - ii. sent by post or courier within the same country, forty-eight (48) hours after posting; or
  - iii. sent by international courier, 7 days after posting.
- b. Any notice given under or in connection with this Agreement shall be in the English language.
- c. This clause 15 does not apply to the service of any proceedings or other documents in any legal action or other method of dispute resolution.

## 17. Language

- a. This Agreement is drafted in the English language. If it is translated into any other language, the English language version shall prevail.
- b. All other documents provided under or in connection with this Agreement shall be in the English language, or accompanied by a certified English translation. If such document is translated into any other language, the English language version shall prevail.

## 18. General

- a. The definitions in Schedule Part 1 to these Processing Conditions shall apply to this Agreement.
- b. The following rules of interpretation shall apply to this Agreement:
  - i. any reference to a clause or Schedule Part is to the relevant clause or Schedule Part of or to this Agreement and any reference to a sub-clause or paragraph is to the relevant sub-clause or paragraph of the clause or Schedule Part in which it appears;

- ii. the headings are included for convenience only and shall not affect the construction or interpretation of this Agreement;
  - iii. use of the singular includes the plural and vice versa and use of any gender includes the other genders;
  - iv. a reference to "person" includes a natural person, corporate or unincorporated body (whether or not having separate legal personality);
  - v. a reference to legislation or a legislative provision:
    - (i) is a reference to it as amended, extended or re-enacted from time to time; and
    - (ii) shall include all subordinate legislation made from time to time under that legislation or legislative provision.
  - vi. any phrase introduced by the terms "including", "include", "in particular" or any similar expression shall be construed as illustrative and shall not limit the sense of the words preceding those terms; and
  - vii. references to any act or omission or breach or non-compliance by or on the part of either party shall be deemed to include a reference to any act, omission, breach or non-compliance by any contractor, Processor, personnel or other representative of the relevant party.
- c. In the case of conflict or ambiguity between:
- i. any provision contained in these Processing Conditions and any provision contained in the Processing Details, the provisions in the Processing Details will prevail; and
  - ii. any of the provisions of this Agreement and any executed and applicable SCC, the provisions of the SCC will prevail.
- d. Each party shall perform its obligations under this Agreement at its own cost.
- e. This Agreement constitutes the entire understanding between the parties in relation to its subject matter and supersedes any prior arrangements, understandings, promises or agreements made or existing between the parties regarding its subject matter.
- f. No party shall be entitled to assign, novate, transfer, sub-license or sub-contract any of its rights or obligations under this Agreement without the prior written consent of the other party.
- g. No amendment of any term of this Agreement shall be effective unless it is in writing and signed by or on behalf of each party.
- h. No failure or delay by a party to exercise any right or remedy provided under this Agreement or by law shall constitute a waiver of that or any other right or remedy, nor shall it prevent or restrict the further exercise of that or any other right or remedy. No single or partial exercise of such right or remedy shall prevent or restrict the further exercise of that or any other right or remedy.
- i. Nothing in this Agreement shall be construed as establishing or implying any partnership or joint venture between the parties and nothing in this Agreement shall be deemed to constitute one party as an agent of another party or authorise a party to bind another party.
- j. A person who is not a party to the Agreement shall have no right to enforce any of its provisions which, expressly or by implication, confer a benefit on him and this Agreement does not give rise to any rights under the Contract (Third Party Rights) to enforce any term of this Agreement (but this does not affect any right or remedy of any person which exists or is available otherwise than pursuant to the Contract (Third Party Rights)).
- k. If any provision or part-provision of this Agreement is or becomes invalid, illegal or unenforceable, it shall be deemed modified to the minimum extent necessary to make it valid, legal and enforceable. If such modification is not possible, the relevant provision or part-provision shall be deemed deleted. Any modification to or deletion of a provision or part-provision under this clause shall not affect the validity and enforceability of the rest of this Agreement.
- l. This Agreement may be executed in any number of counterparts, each of which when executed and delivered shall constitute a duplicate original, but all the counterparts shall together constitute the one agreement.
- m. This Agreement shall be interpreted and applied in accordance with the laws of Germany and the parties hereby submit to the exclusive jurisdiction of the German courts.



This is the Schedule referred to in the foregoing Processing Conditions between Controllers and the Partner named in the Processing Details.

**SCHEDULE**  
**PART 1**  
**DEFINITIONS**

<b>“Agreement”</b>	this Controller-to-Controller data processing agreement between Controllers and the Partner consisting of the Processing Details, the Processing Conditions and this Schedule;
<b>“Controller”</b>	has the meaning given in the Data Protection Laws;
<b>“Data Processing Details”</b>	means the data processing details annexed to and forming part of this Agreement;
<b>“Data Protection Laws”</b>	means any law, statute, subordinate legislation regulation, order, mandatory guidance or code of practice, judgment of a relevant court of law, or directives or requirements of any regulatory body which relates to the protection of individuals with regard to the processing of personal data, electronic communications and privacy to which a party is subject including the Data Protection Act 2018 and any statutory modification or re-enactment thereof and the GDPR;
<b>“Data Subject”</b>	has the meaning given in the Data Protection Laws;
<b>“Data Subject Access Request”</b>	a request (or purported request) under Article 15 of the GDPR;
<b>“parties”</b>	Controllers and the Partner, and the word “party” shall be construed accordingly;
<b>“Personal Data”</b>	has the meaning given in the Data Protection Laws;
<b>“Personal Data Breach”</b>	has the meaning given in the Data Protection Laws;
<b>“Processing”</b>	has the meaning given in the Data Protection Laws (and cognate expressions shall be interpreted accordingly);
<b>“Processor”</b>	has the meaning given in the Data Protection Laws;
<b>“Purposes”</b>	means the purposes specified in the Processing Details;
<b>“Schedule”</b>	means this Schedule annexed to the Processing Conditions and forming part of the Agreement;
<b>“Security Requirements”</b>	means the security requirements outlined in the Processing Details;
<b>“Sharing Start Date”</b>	means the sharing start date set out in the Processing Details;
<b>“Standard Contractual Clauses” or “SCC”</b>	The European Commission’s Standard Contractual Clauses for the transfer of Personal Data from the EU to controllers established in third countries (controller-to-controller transfers) as set out in Schedule Part 2 (if required) (as may be amended from time to time);



<b>“Supervisory Authority”</b>	means the European Commission and any other supervisory authority under the Data Protection Laws;
<b>“GDPR”</b>	means Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data; and
<b>“Working Day”</b>	means Monday to Friday, excluding public holidays or bank holidays (ignoring 24 hour internet banking services).

## SCHEDULE

### PART 2

**[Insert Standard Contract Clauses if required]**

[\[1\]](#) Grant Agreement no. 833635