



---

## D1.4 Internal Progress and Quality Planning Report

---

<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: 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>	25 August 2020
<b>Due date:</b>	5 September 2020
<b>Deliverable lead:</b>	IDIAP
<b>Work package:</b>	WP1
<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	25 August 2020	IDIAP	All	Draft
	02 September 2020	KEMEA	WP2, T2.5 and WP8	Added Content
	02 September 2020	BUT, NFI, CAP, AIRBUS	All WPs	Added content
	02 September 2020	LUH	WP6	Draft
	02 - 03 September 2020	SAIL	All, WP4	Editing, Draft
	02 - 03 September 2020	INT	WP9, WP2, WP3	Review and editing
	03 September 2020	ADDITESS	T4.5, T7.4, T8.2	Added content
	03 September 2020	TRI	T1.5, WP3, WP10	Added content
	04 September 2020	IDIAP	All	Proof-reading, finalising missing sections
V1.0	05 September 2020	IDIAP	All	final updates

## Executive summary

This deliverable D1.4, **is part of Work Package 1 (WP1) project management work package.**

The aim of this Internal Progress Report is to present the current status of the ROXANNE project, twelve months (M12) of its activity (months 1 to 12 (September 2019 – August 2020)) from the kick-off of the project. During these 12 months, the ROXANNE consortium focused on launching the project, collecting end-user requirements (WP2), collecting and analysis of legal, ethics and societal aspects (WP3), and high level system architecture (WP7). Significant efforts were also invested in the organization and planning of the project (WP1). For the R&D work, an effort was devoted to all 3 WPs (WP4 on data management, WP5 on speech, text, video analysis, and WP6 on network analysis). The ROXANNE consortium also started the dissemination and exploitation activities, together with communication of project results to stakeholders and other interested partners (WP9). Finally, the whole consortium in large focused on WP10 to well prepare all the work for the first ethics check (coordinated by TRI partner).

The project directions in Y1 were significantly influenced by the lock-down (due to the COVID-19 pandemic), specifically during the March-June 2020 period. The major effect was postponing the first field-testing event (M9) to a later date (finally decided to be organised completely remotely in September 2020, M13 of the project). At the same time, many of the planned dissemination activities were also replaced, postponed or some of them cancelled. Nevertheless, the project continues according to the planned work in all 10 work-packages, although some of them with a delay.

This report is sectioned according to the project work packages. It also describes the status of the deliverables that were due in the first twelve months of the project and outlines any problems and delays experienced or expected in the ROXANNE work and the actions employed to resolve them.

The following sections are split according to work-packages (WP1 to WP10). The last section 11 contains a table describing problems occurred during the Y1 of the project, solutions and proposed deviations.

## Table of contents

Disclaimer	2
Copyright notice	2
Revision history	3
Executive summary	4
Table of contents	5
<b>1. Work carried out during the period M1-M12</b>	<b>9</b>
T1.1 Establishment of management structure	9
T1.2 Management, internal communication and reporting	10
T1.3 Scientific/Innovation coordination	11
T1.4 Quality assurance and risk management	11
T1.5 Elaborate the project's data management plan	11
Status of deliverables	11
Work planned for next Y2 period	11
<b>2. End-user requirements and use-cases</b>	<b>12</b>
T2.1 Collection of end-user requirements	12
T2.2 Analysis of end-user requirements	12
T2.3 Use-cases validation	12
T2.4 Technology requirements	13
T2.5 User training requirements	13
Status of deliverables	13
Work planned for next Y2 period	13
<b>3. Compliance with European societal values, fundamental rights and applicable legislation</b>	<b>14</b>
T3.1 Adhere to good ethical practices	14
T3.2 Comply with societal values	15
T3.3 Comply with fundamental rights	15
T3.4 Comply with applicable legislation, including in the area of free movement of persons, privacy and protection of personal data	16
T3.5 Examine the potential of using INTERPOL's existing global communications infrastructure and data storage mechanisms to facilitate rapid sharing of criminal evidence including digital evidence	16
T3.6 Develop a decision-making mechanism (in electronic form) to assist organisations in ensuring their compliance with the above objectives	16
T3.7 Validate how best to integrate considerations of societal values, fundamental rights and applicable legislation into the project's proposed solutions	17



Status of deliverables	17
Work planned for next Y2 period	17
<b>4. Data management</b>	<b>18</b>
T4.1 Inventory and analysis of lawfully intercepted data	18
T4.2 Overview of publicly available data resources for research, training and development	18
T4.3 Social media data ingestion	18
T4.4 Data pre-processing and enhancement	19
T4.5 Case management	19
T4.6 Target data simulation for development and demonstration activities	19
T4.7 Data and remote access platform for field-tests	19
Status of deliverables	19
Work planned for next Y2 period	20
<b>5. Speech, text and video data analysis</b>	<b>21</b>
T5.1 Initial speech/NLP/video technologies	21
T5.2 Speaker identification, diarization and role recognition in multiparty interaction	21
T5.3 ASR for network analysis: language use and vocabulary propagation	22
T5.4 Entity extraction and geo-information	22
T5.5 Relation extraction and discovery of new relation types	22
T5.6 Video data processing	23
T5.7 Production speech/NLP/video technologies for network analysis	23
Status of deliverables	23
Work planned for next Y2 period	24
<b>6. Network and relation analysis</b>	<b>25</b>
T6.1 Fusion of information from component technologies for network analysis	25
T6.2 Construction of crime related networks from different data sources	25
T6.3 Multilayer and cross network structural analysis	26
T6.4 Multilayer and cross network behavioral analysis	27
T6.5 Latent subnetwork detection	27
T6.6 Subnetwork shrinking	27
T6.7 Systematic assessment of the contribution of SLT and relation analysis to criminal NA	27
Status of deliverables	28
Work planned for next Y2 period	28
<b>7. Integration and visualisation of results</b>	<b>30</b>
T7.1 Design and definition of the ROXANNE system architecture M1-M6	30



T7.2 User management and access control M6-M30	30
T7.3 Run-time data visualisation and exploratory analysis M4-M30	30
T7.4 Secure data export and exchange M6-M30	31
T7.5 Integration, ROXANNE system/platform setup and maintenance M4-M36	31
T7.6 Integration of feedback from end-users M10-M36	31
Status of deliverables	32
Work planned for next Y2 period	32
<b>8. Field tests, user training and continuous testing</b>	<b>33</b>
T8.1 Development of end-user validation and performance test methodology	33
T8.2 End-User training	33
T8.3 Field test planning	33
T8.4 Design, implementation of field-tests	33
T8.5 Continuous testing	34
T8.6 Field test results analysis and evaluation	34
Status of deliverables	34
Work planned for next Y2 period	34
<b>9. Dissemination, exploitation and communications</b>	<b>36</b>
T9.1 Conduct stakeholder analysis and compile a stakeholder contact list	36
T9.2 Elaborate the dissemination and exploitation plan, IPR management	36
T9.3 Elaborate the plan of communications activities	36
T9.4 Promote the project identity and the project's website	36
T9.5 Prepare and use dissemination and communications materials	37
T9.6 Exploit the project's results	39
T9.7 Convene the project's final conference and participate in others	39
T9.8 Prepare articles for peer-reviewed journals and conference presentations	39
T9.9 Policy recommendations	41
Status of deliverables	41
Work planned for next Y2 period	41
<b>10. Ethics requirements</b>	<b>42</b>
D10.1 H - Requirement No. 2	42
D10.2 H - Requirement No. 3	42
D10.3 POPD - Requirement No. 6	42
D10.4 POPD - Requirement No. 7	43
D10.5 POPD - Requirement No. 8	43



D10.6 POPD - Requirement No. 9	43
D10.7 POPD - Requirement No. 10	44
D10.8 POPD - Requirement No. 11	44
D10.9 POPD - Requirement No. 12	44
D10.10 POPD - Requirement No. 13	45
D10.11 POPD - Requirement No. 14	45
D10.12 POPD - Requirement No. 15	45
D10.13 GEN - Requirement No. 16	46
D10.14 GEN - Requirement No. 17	46
D10.15 GEN - Requirement No. 18	46
D10.16 M - Requirement No. 19	46
D10.17 OEI - Requirement No. 20	46
Status of deliverables	47
Work planned for next Y2 period	48
<b>11. Problems, solutions and deviations</b>	<b>49</b>



## 1. Work carried out during the period M1-M12

### T1.1 Establishment of management structure

This task is mainly intended to facilitate the cooperation between partners while maintaining a strict control of gradual achievements of the action objectives. The management structure was prepared thoroughly during the kick-off meeting in M1 (September 2019, physically hosted by IDIAP). During the first months of the project, the project partners finalised the appointment of representatives into ROXANNE governing bodies. This specifically includes the Internal and External Ethics Boards, External Advisory Board (i.e. later referred to as the Stakeholder Board) and Security Advisory Board.

More details about management of the project is given in following points:

#### Teleconference calls:

IDIAP together with WP leaders have agreed on regular conference calls, which are organised through the various tele-conference systems. More specifically, we have following calls established:

- project general call (bi-weekly) organised by IDIAP - each project partner is required to participate
- WP2 call (monthly) organised by NFI, dedicated to discuss issues with internal LEAs
- WP3 and WP10 (monthly) calls including all partners and organised by TRI
- Technical calls (bi-weekly) organised by BUT for all discussions related to WP4, WP5, WP6 and WP7 work
- WP8 call (monthly) related to preparation of field-testing events and training platform, organised by KEMEA
- WP9 (monthly) calls organised by CAP for all aspects related to communication, dissemination and exploitation. Exploitation work is often discussed separately by all tech partners on calls organised by AIRBUS.

To support the internal communication in the project, IDIAP has set up several email lists (i.e. general, technical, LEAs, social-media, admin) to make sure the information is well communicated with all respective partners. More details were given in the D1.1 project handbook.

#### Organisation of consortium meetings:

Three consortium meetings were organised during the first twelve months of the project:

- M1 (physical meeting at IDIAP) - kick-off meeting, with participations of all partners, including several stakeholder board members (EUROPOL, U.N. and special guests from INTERPOL unit)
- M8 (remote meeting) - consortium meeting discussing the progress of the project, the meeting was also attended by ZITiS (stakeholder board member).
- M12 (remote meeting) - consortium meeting dedicated to the preparation of the field-test event, the meeting was also attended by ZITiS (the stakeholder board member).

#### Other meetings:

Meeting with stakeholder board members: a remote meeting was organised with all stakeholder board members, organised by INTERPOL.

TRI partner organises regular monthly remote meetings with the Internal Ethics Board, and meetings every 3 months with External Ethics Board members respectively.

Project board meeting - the meeting was organised by IDIAP in M12. This was a remote meeting.

The project coordinator also participates in CLUSTER meetings. CLUSTER<sup>1</sup> is a group of 8 FCT projects with goals to find synergies across the projects and increase the collaboration across these projects.

## T1.2 Management, internal communication and reporting

To support an easy exchange of material/documents, or pieces of SW, two cloud services are deployed by ROXANNE consortium as described in details in D1.2:

- SwitchDrive - a cloud based collaborative platform (offered by IDIAP) allowing to share any kind of document/data, or code with all partners. Every partner has a dedicated user-account allowing logging all the movements of the platform. Since the platform is operated by Swiss non-profit organisation, the project partners are assured that the information stays on the Swiss servers. Further, every partner can use its dedicated space, shared only with a certain partner.
- Gitlab - a cloud platform maintained by LUH allowing sub-versioning of the code shared with all the partners. The GIT is maintained directly by LUH, exploiting their server.

This task also takes care of the communication with the EC (involving the project coordinator), and administrative and financial coordination (IDIAP). IDIAP maintains centralised collection of information regarding the cost expenditure of all partners in the project, and transfers the EC advanced payment received by the Coordinator to all the consortium's partners.

### COVID-19 situation:

The situation related to lock-down brought several changes to the project, specified in more details below:

- both consortium meetings (M4 and M12) were replaced by remote teleconference calls. We found this solution quite efficient
- some of the partners did not have access to their computing systems for several weeks of the project. This was mainly due to the decisions imposed on the work from home. This situation got closely back to normal in ~June 2020
- PSNI has informed the coordinator to stop working on the project for some time as their effort was required to be fully dedicated to support the government in lock-down period. This was the case for ~4 months of the project.
- Due to the unclear situation with closed borders and availability to travel abroad, the project consortium hesitated about organising the first field-test event (originally planned in M9 (June 2020)). In May 2020, as the situation did not improve, the partners agreed on moving the field-test event to September 2020. During the second consortium meeting (M12), the project partners decided to keep the date but to organise the meeting completely remotely, as most of the project partners including stakeholder board members are still partially restricted in their travels. The 1st field-test event will be hosted through Zoom teleconferencing system and will take place on September 30, 2020. The meeting will be preceded by the preparation day (dry run).

## T1.3 Scientific/Innovation coordination

From the beginning of the project, the scientific and innovation coordination is done by IDIAP, although with help of other experienced partners (for scientific and research work mainly by BUT, for innovation work by AIRBUS, for issues related to LEAs by NFI, INTERPOL and KEMEA, while for all legal, ethical and societal activities by TRI, and for dissemination and communication work by CAP).

---

<sup>1</sup> <https://www.ccdriver-h2020.com/cluster>



The coordination of respective WPs is done by WP leaders. In case of technical WPs, the work is rather merged into one working group as there is a large overlap among the planned activities.

### T1.4 Quality assurance and risk management

IDIAP coordinated the work on two deliverables, D1.4 related to internal progress and quality planning report (i.e. this document) for Y1 of the project and D1.2 related to Risk assessment of the project.

### T1.5 Elaborate the project’s data management plan

TRI developed a questionnaire regarding important questions related to data management. This was distributed to all partners to complete. All partners provide the required information either through the questionnaire, or directly into the draft document. TRI used this information to develop the data management plan. This was internally reviewed by USAAR, IDIAP, AEGIS, AIRBUS, PHO, ITML, INTERPOL, SAIL, and LUH. TRI will coordinate updates to the data management plan at M18 and M36.

### Status of deliverables

Deliverable	Status
D1.1 Project handbook (IDIAP)	Submitted in M3
D1.2 Risk assessment (IDIAP)	Due in M6, delayed, and submitted in M12
D1.3 ROXANNE’s data management plan (TRI)	Submitted in M6
D1.4 Internal progress and quality planning report (IDIAP)	Submitted in M12
D1.5 Internal progress and quality planning report (IDIAP)	Due M24
D1.6 Final progress report (IDIAP)	Due M36

### Work planned for next Y2 period

WP1 will continue in managing the ROXANNE project, specifically for research and administrative work, and supporting AIRBUS and NFI for innovation end-user related management.

## 2. End-user requirements and use-cases

### T2.1 Collection of end-user requirements

The ambition of the project is to enhance criminal investigations. Technology experts and law enforcement agencies work together closely to make sure that the technology answers to the needs in the field. This task covers two ways to facility this collaboration.

First, we set up a periodic (monthly) meeting focused on knowledge exchange between LEAs and technology developers. In this meeting, we demonstrated technology and other intermediate results in order to get early feedback from the field. On the other hand, LEAs can share their insights from criminal investigations with developers, for them to get a feeling of what is important, what are the pitfalls, etc.

Second, we developed an end-user requirements survey. This survey was distributed among LEAs, as part of a longer survey which also includes questions about legal requirements and technology requirements (T2.4). Depending on the participant and their field of expertise, they could opt to answer all sections or only the sections in which they wish to share their opinion. The survey was translated to French, Spanish and Arabic, and distributed via LEAs in the project, via Stakeholder board members, and via INTERPOL's global law enforcement network (194 member countries). The requirements collection process is described in deliverable D2.1. The responses from participants are analyzed in T2.2 and reported in D2.2 (end-user training requirements) and D2.3 (ROXANNE end-user requirements).

The output of this task also contributes to T2.3 (use cases validation) and will affect other work packages as well.

This task was closed in month M6. However, the periodic meetings will continue in order to ensure that the project will meet LEAs' expectations. In month 12 we performed an internal evaluation of the collaboration between LEAs and technical partners. The results showed that LEAs were generally positive but they also gave some suggestions for improvement, such as LEAs' role in dissemination. We keep committed to achieving a productive collaboration in the next phase of the project.

## T2.2 Analysis of end-user requirements

In task T2.1 we designed and distributed a requirements survey, and collected results. This task covers the analysis of the responses, which is reported in deliverable D2.3. This task was closed in month 12, after the submission of this deliverable.

## T2.3 Use-cases validation

Three operational use-cases will facilitate the development and testing of the ROXANNE technology. These use-cases will reflect LEA's needs and will be developed in the context of criminal investigations and international police cooperation. The use-cases will provide the consortium a clear understanding of the end-user requirements and priorities for the development of ROXANNE outcomes and will serve as a basis for the implementation of appropriate legal safeguards (WP3), will drive technical developments (WP4-WP7), and prepare the Field Tests (WP8). The consortium will rely on three already elaborated operational use-cases which will be fine-tuned according to the results from task T2.2.

This task is planned to start in month M13 (September 2020) .

## T2.4 Technology requirements

In this task we have started translating end-user needs (from T2.1 as well as from direct discussions with LEAs) into technical requirements . This allowed us to provide the necessary information to WP7 in order to move forward on the specification of the platform as well as on its software and hardware architecture definition. We will then benefit from the results of task T2.3 in order to refine these technical requirements.

## T2.5 User training requirements

This task is related to the acquisition of end-user training requirements. These will be used for the creation and refinery of engaging curricula tailored for each of the targeted end-user categories, defining both the theoretical and practical training that will occur through e-learning, or physically at workshops aligned with field-test meetings. The task started at M1 of the project, parallel to T2.1. After several discussions and telco meetings, we managed to develop an end-user training requirements survey, both in online (via EU survey platform) and editable PDF format, which initially had been part of a one joint questionnaire with 49

questions. However, after the conducted pilot study among the consortium partners, 10 feedback responses had been received urging the need to be divided in three separate questionnaires, for the questions to be targeted at a more specific audience. For that reason, the end-user training questionnaire was distributed as a separate survey with 21 questions in total. The total number of responses added up to 31 participants stemming from ROXANNE Consortium practitioner partners. The analysis of the responses resulted in specific end-user training requirements, which were also prioritised according to the MoSCoW rating system. The final outcomes, as outlined in D2.2, will feed to the proper design of the training platform (WP8, T8.2). However, due to the fact that this analysis stemmed from the first iteration of the questionnaire among the Consortium partners, the outcomes will be continuously reviewed after the second iteration among the INTERPOL’s global law enforcement network, after each field test as well as during the continuous testing of the ROXANNE platform, taking into account all the relevant and update needs of the end-users in terms of their training and education.

### Status of deliverables

Deliverable	Status
D2.1 End-user requirements - collection methodology	Submitted in M8
D2.2 End-user training requirements	Submitted in M9
D2.3 ROXANNE end-user requirements	Submitted in M12
D2.4 ROXANNE use-cases	Due in 18

### Work planned for next Y2 period

The major part of activities in WP2 will be devoted to use-cases to be analysed in the ROXANNE project, with help of all internal LEAs, and supported by the ROXANNE platform built by technical partners.

### 3. Compliance with European societal values, fundamental rights and applicable legislation

As is explained in the following sections, D3.1 has been delayed owing to the impact of COVID-19 on the first field-test. However, the imposition of the WP10 ethics requirements on the project following acceptance of the proposal has caused many partners, and particularly those in WP3, to need to divert attention away from what was already a tight schedule. Consequently, some of the later aspects of tasks due to be completed by M12 are running over (specifics are explained by task below). In any case, these will be completed before the delayed D3.1 is submitted.

#### T3.1 Adhere to good ethical practices

##### ST 3.1.1. Logistics

According to the Grant Agreement, TRI contacted several experts to join the ROXANNE Ethics Board. 4 external experts joined the Board alongside 4 members of the consortium. After 2 meetings, the external members recommended that the internal members form a separate board to discuss issues arising from the project more regularly. The project did this, and information on the board memberships are provided in D10.12. TRI organises the meetings of both the internal and external ethics boards. 2 meetings were held as a joint board, and, following the separation, 1 meeting has been held with external experts (another is planned for after the upcoming field-test) and 6 have taken place with internal partners. WP3 partners CAP and INTERPOL have participated in ethics board meetings and contributed to discussions. In addition, in conjunction with WP10, TRI along with INTERPOL, SAIL and LUH have participated in a DPIA working group to develop a DPIA template that will work well for individual partners and allow for the completed templates to be combined into a composite-DPIA for the whole project.

This sub-task also required TRI to conduct an overview of national ethics committees for partners who could, potentially, use human participants in their research in order to ascertain whether these committees (12) could provide an ethical review of these activities. None were found, as most were focussed on bio-ethics or could only review research conducted by their institution. Following advice given in the first ethics check report, TRI contacted university partners to request them to speak with their university ethics committees and ask if they would be willing to review research activities involving human participants. USAAR's ethics committee reviewed information sheets and informed consent forms prepared by TRI, and information about the nature of human participation regarding each research activity and approved them. More detailed information is provided in D10.2, which also includes a record of communications between project partners and the institutional ethics committees of the partners.

##### ST 3.1.2. Identify and assess ethical issues arising from the project

In order to identify ethical issues in the project, TRI completed a 'Touchpoint Table', which lists each task and ethical risks that could be generated by them. These ethical risks are based around 7 ethics requirements from the SHERPA project.<sup>2</sup> TRI also suggested mitigation strategies which were then discussed with WP leaders, and a common understanding of both of the risks and solutions were reached. This agreed understanding of risks and mitigation strategies formed the basis of deeper analysis work conducted by TRI, CAP, and INTERPOL. Each covered the application of specific ethical requirements from the SHERPA guidelines to the project and potential use of the ROXANNE platform: TRI considered

<sup>2</sup> Philip Brey, Björn Lundgren, Kevin Macnish, and Mark Ryan, 'D3.2 Guidelines for the development and use of SIS', SHERPA project, 2019. Available at:

[https://dmu.figshare.com/articles/D3\\_2\\_Guidelines\\_for\\_the\\_development\\_and\\_the\\_use\\_of\\_SIS/11316833](https://dmu.figshare.com/articles/D3_2_Guidelines_for_the_development_and_the_use_of_SIS/11316833)





(1) human agency, liberty, and dignity, (2) technical robustness and safety, (3) privacy and data governance, and (4) individual, societal and environmental well-being. CAP covered (5) transparency. INTERPOL evaluated (6) diversity, non-discrimination and fairness, and (7) accountability. Each right was analysed across 5 phases of the project: requirement gathering; planning and designing; development; testing; evaluation; a 6<sup>th</sup> phase was added to consider issues raised during use of the platform. This analysis led to development of recommendations for the project and potential end-users of the ROXANNE platform (these will be discussed and implemented in T3.7, and results reported in D3.4). The work of both sub-tasks in T3.1 will be presented in the delayed D3.1.

### T3.2 Comply with societal values

CAP and TRI conducted a literature review of academic journal articles and books, industry reports and news articles. This generated over 60 items that were of immediate relevance to societal values in relation to the ROXANNE project and proposed platform. CAP analysed the following societal values: citizens' privacy; trust and the perception of safety; unintended consequences of technological solutions; social acceptability. TRI analysed: democracy and solidarity; equality and tolerance for other cultures; human rights; respect for human life; the rule of law.

CAP and TRI used these analyses to develop a briefing paper which contains recommendations both for the project and use of the platform. These recommendations will be discussed and implemented in T3.7, with results provided in D3.4. Also, 2 scenarios about possible uses of the ROXANNE platform that highlight potential societal issues that could arise. This briefing paper and the 2 scenarios are to be presented at the first field-test by CAP and TRI who will ask for feedback from attendees such as the stakeholder board and external ethics board. Owing to the field-test being delayed, this has meant that the presentation and evaluation of feedback has been delayed, and so submission of D3.1 has also been delayed. Following this presentation, the scenarios will be placed on the project website to gather feedback from citizens; this feedback will be included in D3.4.

### T3.3 Comply with fundamental rights

TRI conducted scoping work to evaluate the applicability of the EU Charter of Fundamental Rights and the relevance of the rights within it to the ROXANNE project, and its potential uses. TRI, CAP, and INTERPOL then divided up the most relevant rights and analysed them. This involved a 'business and human rights' approach to considering the applicability of the rights during the development of the platform, and also a comparative approach that included consideration of other human rights treaties and jurisprudence from other legal regimes where it brought further illumination to the issues.

In terms of specific rights considered, TRI evaluated: Human dignity; Right to the integrity of the person; Prohibition of torture and inhuman or degrading treatment or punishment; Right to liberty and security; Respect for private and family life; Protection of personal data; Freedom of Expression and Information; Freedom of assembly and association. INTERPOL considered rights of: Non-discrimination; Cultural, religious and linguistic diversity; Equality between men and women; The rights of the child; The rights of the elderly; Integration of persons with disabilities. CAP analysed the: Right to an effective remedy and a fair trial; Presumption of innocence and right of defence. This analysis has also generated recommendations for the project and potential use of the platform, which will be discussed and implemented as part of T3.7, with the results reported in D3.4.

In terms of the scenarios and dissemination, this has not yet been completed. At the point of writing TRI is developing initial scenarios for dissemination, which will be discussed with CAP and INTERPOL.

Dissemination of the work from T3.3 is also not completed, but discussions for how to do this are underway. Feedback from these dissemination activities will be included in D3.4.

### **T3.4 Comply with applicable legislation, including in the area of free movement of persons, privacy and protection of personal data**

INTERPOL, as task leader, evaluated the applicable legislation to be analysed. First the consideration of the free movement of persons was removed as it is not especially relevant to the core aspects of the ROXANNE project. Second, INTERPOL confirmed that legislation under consideration, and the partner analysing them, would be: EU General Data Protection Regulation 2016/679 (TRI); EU Law Enforcement Directive 2016/680 (TRI); INTERPOL Rules of the Processing of Data INTERPOL); Council of Europe Convention 108+ (INTERPOL); EU Directive on CopyRight in the Digital Single Market (INTERPOL); EU Network and Information Security Directive 2016/1148 (CAP).

Each piece of legislation was analysed in terms of the following provisions: Lawfulness of data processing; Special categories of data; Data processing principles; Individual rights; Accountability & transparency; Data security; Data storage & retention; Data transfer.

Analysis of these pieces of legislation according to each type of provision has been completed. Development of the online brochure to be disseminated is in progress. These will be provided in the delayed D3.1. Feedback from the online brochure, and information on how partners, stakeholders, and end-users can comply with the provisions across legislation, will be included in D3.4.

### **T3.5 Examine the potential of using INTERPOL's existing global communications infrastructure and data storage mechanisms to facilitate rapid sharing of criminal evidence including digital evidence**

This task is ongoing. INTERPOL has begun to explore the possibility of using its global communication infrastructure and data storage mechanisms to facilitate the rapid sharing of criminal evidence. This will entail an analysis of INTERPOL's legal framework, focusing on the Rules on the Processing of Data (RPD), which govern the functioning of INTERPOL's police information system and define the data processing regime. It will also consider potential options for storing information generated through the ROXANNE tools.

### **T3.6 Develop a decision-making mechanism (in electronic form) to assist organisations in ensuring their compliance with the above objectives**

This task starts in M12.

### **T3.7 Validate how best to integrate considerations of societal values, fundamental rights and applicable legislation into the project's proposed solutions**

This task starts in M12.



## Status of deliverables

Deliverable	Status
D3.1 Initial report on compliance with ethical principles (TRI)	Due in M12, delayed until M14
D3.2 Development of a decisions-making mechanism (TRI)	Due in M14, delayed until M16
D3.3 INTERPOL global communications infrastructure (INTERPOL)	Due in M18
D3.4 Final report on compliance with ethical principles (TRI)	Due in M36

## Work planned for next Y2 period

Year 2 of the ROXANNE project will first involve carrying out the societal values presentation at the delayed first field-test, and completing the uploading of scenarios to the project website in T3.2, and the dissemination activities in T3.3 and T3.4. Information about all of this will be included in D3.1. Next, taking the results of the analysis work in D3.1 to develop a decision-making mechanism in T3.6 will take place leading to submission of D3.2. Concurrently, examination of INTERPOL's global communications infrastructure will take place in T3.5 and lead to submission of D3.3. Finally, validation of recommendations made by WP3 partners through discussions with the External Ethics Board and Stakeholder Board.

## 4. Data management

The WP4 is related to data management activities of the project and consists of 7 tasks led by 5 different Consortium partners. As of the end of Project Year 1, 2 of these tasks are complete and the remaining 5 are active. The primary objective of WP4 is to find, collect, record, process and maintain data in order to carry out the research activities of the project, train/adapt/evaluate the technologies involved, and demonstrate the capabilities of the ROXANNE platform. The achievements obtained so far are in line with this objective and the initial work plan.

### T4.1 Inventory and analysis of lawfully intercepted data

The objective of this task is to prepare a list of “real” datasets containing criminal investigation data (mainly lawfully intercepted telephone calls), which are held/used by the LEA partners of the Consortium and which can potentially be used for evaluating the ROXANNE Platform at a later stage in the project. The FRIDA (Telephone conversations in Dutch) and the BALSAS\_200LT (Telephone conversations in Lithuanian) datasets were offered to the Consortium by the partners who hold the rights to these data. Brief information about these datasets is provided in D4.1 Overview and analysis of lawfully intercepted and publicly available data, submitted in M4. Not all LEA partners were able to provide data to the Consortium, due to the sensitivity of such data and the legal constraints involved. This task ended in M12.

### T4.2 Overview of publicly available data resources for research, training and development

The objective of this task is to: (i) prepare a list of publicly available resources including commercial databases, partners' own data, multimedia files on web and social media, which are already accessible by the project partners, and (ii) consider their use for training, adaptation and evaluation of the technologies to be employed in the ROXANNE solution. The D4.1 listed 19 of such databases with their potential application domains. 3 data sets were eventually selected to be used for the first version of the ROXANNE Platform: Selected episodes from the “CSI: Crime Scene Investigation” TV series, a subset of the telephone conversation data released by the NIST Speaker Recognition Evaluations, and the “Enron corpus” which contains company emails and conversational telephone speech, which were made available to the public. This task ended in M12.

### T4.3 Social media data ingestion

This task aims to ingest and process other data sources into the ROXANNE Platform, in particular the social media channels and data types identified as relevant by end-users through the end-user requirements survey carried out as part of WP2. Since the results of this survey were only available at the end of M12, the initial efforts in this task concentrated on a prototype system definition. This prototype considers collecting the publicly available data from Twitter using some specific keywords based on the use-cases, passing them through a data normalization and anonymization component, and making the anonymized collection available for training and development activities in WP5, WP6 and WP7.

### T4.4 Data pre-processing and enhancement

This task is involved in data-related actions such as determining the encodings and formats, employing cleaning and enhancement methods and considering data protection measures, in order to make the data commonly usable and securely accessible by the project partners. In line with the activities in T4.2,

extensive analyses have been carried out on the selected datasets to bring the data into the desired format and to determine the metadata (ground truth) information. Third party sources such as Google Trends API that can reflect topics of public interest at different geo-locations are also considered and analysed with respect to its applicability. A text anonymization component is currently under development, which will allow to filter out sensitive content from written documents while they are being ingested by the ROXANNE platform.

#### T4.5 Case management

The main objective of this task is to prepare a content management system for the efficient storage, retrieval, management, archiving, processing and logging of data (content, in general) to be used in the ROXANNE platform. The hardware setup for hosting is currently under development alongside the component setup, which will be finalized after the discussions on integration, user roles and data provision. Furthermore, ADDITESS has also investigated the option of re-factoring its components to be deployed using Docker and utilizing CI/CD pipelines. The task is scheduled to finish on M32 and the outcomes are to be reported on D4.4.

#### T4.6 Target data simulation for development and demonstration activities

The aim of this task is to prepare a dataset simulating the behaviour of a criminal network which would resemble “real” investigation material as much as possible. Since the “real” datasets listed in T4.1 were insufficient, efforts in this task started earlier than planned. Based on real-life cases, a criminal activity scenario built around a drug dealing network was prepared. According to the scenario the criminal network members are located in different locations in Europe and communicate with each other over the phone in different languages. Researchers of the ROXANNE project worked as voice actors and a target telephone speech dataset of 100 calls were recorded.

#### T4.7 Data and remote access platform for field-tests

This task deals with defining the data and the use-cases for field testing, and with enhancing continuous testing operations at the LEA sites throughout the whole project. Due to the rescheduling of the first field test, the efforts for this task started only recently. The CSI and NIST datasets of T4.2 and the simulated dataset of T4.6 were selected as the data to be processed and presented in the first field test. Based on the outcome of this field test, along with the requirements analysis in T2.2, we will create a plan for the next field test which will be around month 25. This task is scheduled to end after month 30, but will likely be extended up until the date of the third field test.

#### Status of deliverables

Deliverable	Status
D4.1 Overview and analysis of lawfully intercepted and publicly available data (SAIL)	Submitted in M4
D4.2 Simulated data for development and demonstration (PHO)	Due in M16
D4.3 Final report on ROXANNE data (BUT)	Due in M30
D4.4 ROXANNE case management and data preprocessing (ADITESS)	Due in M32

#### Work planned for next Y2 period

For the second year of the project, the following work is planned in WP4:



- Detailed analysis of the end-user survey results, with emphasis on their languages, data types, and social media platforms of interest
- Completing the non-target calls of the simulated dataset and extending it with other data modalities
- Improving the scenario with different use-cases, hidden connections and additional evidence
- Development of a new audio segmentation system specifically for the ASR component, with improved speaker turn detection and sentence segmentation
- For the social media data collection, individual data analysis based on topics/events of interest and individual network construction based on users/group of users
- Experiments with audio data augmentation for robustness to noise
- Data selection for the second field test

## 5. Speech, text and video data analysis

### T5.1 Initial speech/NLP/video technologies

Initial speech/text/video technologies were gathered across the project and summarized immediately at the outset of the project. Deliverable D5.1 summarizes this work, first giving a non-expert overview of such technologies and concentrating on their use in LEA framework. It also lists technologies available at partners, and presents experiments done so far in the ROXANNE project on three experimental data-sets drawn from public sources: Crime Scene Investigation (CSI) series, National Institute of Standards and Technology (NIST) speaker recognition evaluation data, and the ENRON data-set. Some of these speech, text and video technologies were adapted and prepared for integration into the ROXANNE platform, in the form of integrable DOCKER images.

### T5.2 Speaker identification, diarization and role recognition in multiparty interaction

Experimental data-sets were processed by current speaker identification engines available at the partners and processing results were adapted to be usable in subsequent network analysis. A technique for speaker clustering was developed facilitating grouping speakers by their voice, instead of routinely used meta-information, such as telephone numbers. These works were performed on all three experimental data-sets and are now extended towards ROXANNE simulated data (Task 4.6). Speaker Identification / Language Identification and audio diarization engines were adapted for the ROXANNE platform. ROXANNE teams participated in several international research efforts and international technology evaluations such as NIST (U.S. National Institute of Standards and Technology) Speaker Recognition Evaluation (SRE) 2019 and Short-duration Speaker Verification (SdSV) Challenge 2020. We were successful in both these efforts. Work in T5.2 led to four publications supported by ROXANNE :

1. ALAM Jahangir, BOULIANNE Gilles, BURGET Lukáš, DAHMANE Mohamed, DIEZ Sánchez Mireia, GLEMBEK Ondřej, LALONDE Marc, LOZANO Díez Alicia, MATĚJKA Pavel, MIZERA Petr, MOŠNER Ladislav, NOISEUX Cédric, MONTEIRO Joao, NOVOTNÝ Ondřej, PLCHOT Oldřich, ROHDIN Johan A., SILNOVA Anna, SLAVÍČEK Josef, STAFYLAKIS Themos, ST-CHARLES Pierre-Luc, WANG Shuai and ZEINALI Hossein. Analysis of ABC Submission to NIST SRE 2019 CMN and VAST Challenge. In: Proceedings of Odyssey 2020 The Speaker and Language Recognition Workshop. Tokyo: International Speech Communication Association, 2020, pp. 289-295.
2. SILNOVA Anna, BRUMMER Niko, ROHDIN Johan A., STAFYLAKIS Themos and BURGET Lukáš. Probabilistic embeddings for speaker diarization. In: Proceedings of Odyssey 2020 The Speaker and Language Recognition Workshop. Tokyo: International Speech Communication Association, 2020, pp. 24-31.
3. MOŠNER Ladislav, PLCHOT Oldřich, ROHDIN Johan A. and ČERNOCKÝ Jan. Utilizing VOiCES dataset for multichannel speaker verification with beamforming. In: Proceedings of Odyssey 2020 The Speaker and Language Recognition Workshop. Tokyo: International Speech Communication Association, 2020, pp. 187-193.
4. Sarfjoo, Seyyed Saeed and Madikeri, Srikanth and Motlicek, Petr and Marcel, Sebastien, Supervised domain adaptation for text-independent speaker verification using limited data, Interspeech 2020<sup>3</sup>.

---

<sup>3</sup> [http://publications.idiap.ch/downloads/papers/2020/Sarfjoo\\_INTERSPEECH\\_2020.pdf](http://publications.idiap.ch/downloads/papers/2020/Sarfjoo_INTERSPEECH_2020.pdf)

### T5.3 ASR for network analysis: language use and vocabulary propagation

Automatic Speech Recognition (ASR) was deployed as well for processing of experimental data-sets. An English ASR engine was prepared for integration in the ROXANNE platform and successfully tested on experimental data. Current efforts include adding more languages to the Platform (mainly in view of the simulated data that is in English, Czech, German, Russian and Vietnamese). In ASR, ROXANNE teams also participated in international evaluations, namely in NIST OpenSAT 2020, that is finishing at the time of the submission of this report and where our results look very favorable.

Khonglah, Banriskhem and Madikeri, Srikanth and Dey, Subhadeep and Boulard, Herve and Motlicek, Petr and Billa, Jayadev, INCREMENTAL SEMI-SUPERVISED LEARNING FOR MULTI-GENRE SPEECH RECOGNITION, Proceedings of ICASSP 2020, 2020<sup>4</sup>.

### T5.4 Entity extraction and geo-information

The named entity recognition (NER) module deployed in the ROXANNE platform automatically recognizes important entities appearing in the text such as [PERSON], [LOCATION], [ORGANIZATION] etc. The current ROXANNE platform contains the LSTM-based deep learning NER module. An advanced LSTM-CNN-CRF based model is also finished and ready for deployment. At the same time, a transformer-based model trained on more challenging datasets is under development. This new model should not only perform better, but will also support more labels.

### T5.5 Relation extraction and discovery of new relation types

As current datasets don't fit the relation extraction scenario, we decided to work on relation extraction more in the next phase. Instead, we have been focusing on the topic detection task in which we want to infer the topic of a given document. More specifically, We support two topic-detection modules. The first one is LDA based which defines topics by a list of words appearing in the documents. Such topics are latent yet very flexible. The second one is related to information extraction: an end-user can choose several sentences from the documents which describe the topic in his mind, then our module will retrieve all documents with that topic. We also studied the feasibility of fraud detection using topic detection and preliminary results on the Enron dataset are available.

The topic detection module deployed in the ROXANNE platform can automatically extract the topic from the text and tested on different dataset such as CSI, NIST, and ENRON. The topic detection module uses an unsupervised approach which doesn't require labelled data to train the models.

The authorship attribution module uses Bert fine tuning with a consideration of stylometric and hybrid features to identify different authors from the text data.

### T5.6 Video data processing

During the first year, a modular video processing engine has been adapted in order to ease the integration of existing state of the art image analytics models aimed at providing various axes of analyses of videos. The video processing engine includes a shot detection module to segment a long video in shorter shots. Each shot can then be described based on the outputs of the following possible analytics:

- “object” detection modules where objects can be faces, persons, vehicles...

<sup>4</sup> [http://publications.idiap.ch/downloads/papers/2020/Khonglah\\_ICASSP2020\\_2020.pdf](http://publications.idiap.ch/downloads/papers/2020/Khonglah_ICASSP2020_2020.pdf)

- “object” signature extraction modules where signatures embed the visual aspect of each object to enable similarity searches and clustering for diarization purposes
- “keyword” extraction modules where the general content of the images of a given shot are described with keywords derived from image classification or object detection modules (indoor / outdoor, day / night, military / civil, ...)

The modularity of the developed engine enables the integration of open sourced state of the art models, as well as models developed by the partners. The following models have been integrated for testing purposes:

- FaceNet / MTCNN face detector
- EfficientDet5 object detector
- EfficientNetB7 image classification
- Place365 Resnet18 place classification

Besides, some first place and face diarization tests have been performed on the CSI dataset allowing for instance to derive, based on face clustering, a network whose nodes consist of face clusters (ideally, a single individual), and edges corresponding to various proximity granularities such as:

- the two individuals have been observed in a same frame
- the two individuals have been observed in a same shot (defined by video cuts)
- the two individuals have been observed in the same scene (or conversation where conversation time stamps come from audio analytics)
- the two individuals have been observed in a same video

The objective of next months is to extract from end-user requirements questionnaire video analytics needs that could have been reported to better focus the integrated analytics with respect to end user pain points, to further work on the automatic summarization of videos through people / object / places clustering and diarization, and work on the complementarity between audio and visual information, for instance to enable the automatic association between face embeddings and voice prints to associate each person of interest with a multi-modal signature (audio + image).

## T5.7 Production speech/NLP/video technologies for network analysis

This task is planned to start in M25.

### Status of deliverables

Deliverable	Status
D5.1 Initial speech/text/video technologies (PHO)	Submitted in M8
D5.2 Interaction of speech/text/video and network analysis (USAAR)	Due in M18
D5.3 Final set of speech/text/video technologies (BUT)	Due in M33

### Work planned for next Y2 period

In Y2, the following work is planned in WP5:

- extensive work on ROXANNE simulated data, as well on possible other data-sources from or close to the LEA application domain.





- in speaker identification, tighter integration with the requirements of network analysis, for example, in combination with (probabilistic) meta-information, adaptive thresholding for speaker clustering, etc.
- in ASR, work on languages relevant to LEAs and on vocabulary adaptation towards the end domain.
- for entity extraction and geo-information, USAAR and BUT plan to extensively work on this topic. For the named entity recognition (NER), a state-of-the-art transformer based model is planned which will not only perform better, but also support more labels. Moreover, we will support new labels on demand.
- in relation extraction, we plan to integrate the module into the platform, while exploring meaningful use cases. Otherwise we will be very flexible and can deliver other NLP modules on demand, e.g. topic detection modules as we did in the first phase.
- in topic detection, zero-shot classifiers will be implemented which let humans to define the labels.
- in authorship attribution, will be implemented on the automatically generated transcripts (i.e. by automatic speech recognition system).
- in video processing, end-user requirements related to video analytics will be analysed and exploited to refine the types of image analytics to be integrated in the modular video processing engine. Additional work on face clustering and diarization will contribute to the elaboration of multi-modal descriptions of individuals (audio and image) to ease their re-identification in available source documents. Finally, the developed analytics and processing engine will be integrated in the ROXANNE platform to contribute to next field tests.



## 6. Network and relation analysis

The main results in the first year of the project includes the following:

- Data fusion bus construction: we have investigated available technologies and designing options for constructing a data fusion bus that is more appropriate for ROXANNE
- Network analysis review: we conducted a comprehensive systematic review of network analysis functionalities, particularly the ones closely related to ROXANNE, the existing methods and promising research direction for each of them. We have also surveyed the open source libraries and commercial software that provide those functionalities.
- Network API construction: we leveraged open source libraries to construct an API server for providing the network analysis functionalities that are essential for the tasks in this work package. The first versions of the server have been released and successfully integrated into ROXANNE's pipeline.
- Network visualizer construction: we have developed a web-based interface for importing, analyzing and visualizing networks and the analysis results. This visualizer allows users to perform the essential tasks as the above API server, but through a much visually friendly interface that helps non-technical users can examine and investigate more efficiently. The first versions of the visualizer have been released and updated frequently.
- Investigation of assessment of different information sources: we investigated and examined methods for assessing the contribution of information provided by different technologies components for network analysis
- Deliverable D6.1: Lastly, we have summarized our reviews and surveys on network analysis and described our work in developing the above systems and software components in the deliverable D6.1. We have submitted the deliverable on time, at the end of month 9 of the project.

In the following sections, we will describe in more detail the progress in each task of this work package, the on-going works, and our plan for the next year of the project.

### T6.1 Fusion of information from component technologies for network analysis

In the overall effort of WP6 to use multimodal sources in order to represent criminal networks in a static but also dynamic way and therefore effectively support the investigation procedures, T6.1 aims to create an active interface between the outcomes of WP5 based on speech, text and video data analysis and the network analysis procedures. Therefore, the efforts on T6.1 were targeted on how the real-time data aggregation and alignment will be delivered. The creation of an application programming interface (API) for merging, filtering and aggregation functions on the dataset schemas provided by WP6 Leader (LUH) summarizes the ongoing technical and development activities but also stands as the first goal of T6.1 in order to start using the ITML's data fusion bus services. Alongside, ITML as leader of T6.1 is working towards the development of an initial containerized version to be provided in WP6 and in the ROXANNE Platform in general.

### T6.2 Construction of crime related networks from different data sources

#### Work progress:

At the beginning months, when real datasets are not yet available while the simulated data generation work is in progress, we have investigated a wide range of public datasets for developing and testing the network analysis system described above. These include:

- **Covert networks:** a collection of past cases' networks which are collected and made publicly available by researchers from the University of Manchester. Here, we focus on a subset of the collection where not only the network among the involved individuals is available but also some of their attributes (e.g., name, gender, role, etc.). These include the networks of the Madoff fraud case and 911 attack.
- **CSI networks:** a collection of conversational networks among characters in selected episodes of the Crime Scene Investigation series. Detailed information of this dataset is given in Section 4 of the deliverable D5.1 Initial speech/text/video technologies. Partners in WP5 have run extensive experiments on this dataset and helped to extract “who talked to whom” information from the videos. Using this information, we have constructed several networks with tens of individuals and tens of conversational links
- **Enron network:** a dataset among a group of employers in Enron who have exchanged some emails or were involved in some phone calls. Similarly, detailed information of this dataset is given in Section 4 of the deliverable D5.1 Initial speech/text/video technologies. Preprocessing steps in WP5 have helped to extract from this dataset a network of 69 employers and 350 communication links (i.e., email or phone call) among them.
- **NIST datasets:** consists of two clusters of the phone call network dataset that was published under NIST Speaker Recognition Evaluations and acquired by the ROXANNE partners. These two network clusters have 2044 nodes & 5866 links and 354 nodes & 468 links, respectively.

We have also developed a toolbox for constructing the networks from the outputs of speech/text/video processing chains in WP5. The toolbox is provided as an API as described above, and has been successfully integrated into the ROXANNE's pipeline.

#### On-going work:

Recently, partners in this WP6 have been contributing extensively to the simulated data collection and preparation work in WP4. We are working on constructing different networks from this simulated dataset and integrating these networks into our system components. We are also working on improving the toolbox and API function for constructing networks from WP5's outputs, so as to make them work more smoothly in the ROXANNE pipeline and also to take more information from the outputs into account (e.g., using topic and entities mentioned in the conversations to type the links).

### T6.3 Multilayer and cross network structural analysis

**Work progress:** We have reviewed a wide range of methods for network analysis, both within and cross networks. Based on the review, we have leveraged open source libraries to build a toolbox for performing those analyses. The toolbox is integrated into our API server and network visualizer as described above. In the first versions of the toolboxes, we focus on the analysis methods that purely rely on the structure of the networks, i.e. considering only links among the networks' individuals. As for now, for each essential analysis, our toolbox provides several method choices for performing the analysis. We provide more detail about the methods in the deliverable D6.1

**On-going work:** We are investigating network embedding methods and their applications in network analysis. These methods have been shown to outperform other traditional methods while not requiring much domain knowledge. Furthermore, these methods can be extended easily to also take non-structural information into account, e.g., individuals' and links' properties.

### T6.4 Multilayer and cross network behavioral analysis

**Work progress:** We have conducted an extensive survey on methods for analyzing behavior of individuals in the network. We focused on outlier detection methods which are potentially useful for criminal network investigations as they may help to suggest suspicious individuals or behaviors. Again, we

summarized the survey in the deliverable D6.1. As the detection of outliers is highly domain specific, we decided to leave the development and integration of these methods for the later phase of the project when more dataset is available. Instead, we first centered on implementing, in our visualizer described above, the visual effects that help users to examine and investigate individuals' behavior more easily.

**On-going work:** As more data are collected and shared among project partners, we are investigating, identifying and adapting methods for outlier detection that are more appropriate for the project. We are also working on implementing and integrating these methods into our network analysis toolbox.

## T6.5 Latent subnetwork detection

**Work progress:** We have investigated a wide range of methods for network latent subnetwork detection. These include methods for community detection and link prediction tasks. We have leveraged open source libraries to implement, adapt and integrate these methods into our network analysis toolbox. As for now, for each of the tasks, our toolbox provides several method choices for performing the task. We provide more detail about the methods in the deliverable D6.1

**On-going work:** Again, we are investigating network embedding methods and their applications in community detection and link prediction. As mentioned above, these methods have been shown to outperform other traditional methods and highly extensible to leverage side information (e.g., individuals' and links' properties) to improve the performance of the task.

## T6.6 Subnetwork shrinking

**Work progress:** We have investigated a wide range of methods for social influence analysis, e.g., to quantify the relative importance of individuals within the same network. These importance might be helpful in identifying notable individuals in criminal networks, and hence help investigators to "shrink" the network to focus on those individuals. Based on the review, we have leveraged open source libraries to build and integrate into our network analysis toolbox the function to perform social influence analysis. As for now, our toolbox provides several method choices for performing the analysis. Again, we provide more detail about the methods in the deliverable D6.1

**On-going work:** We are investigating methods for also considering individuals' and links' properties in quantifying the social influence. These methods will be potentially useful in the later phases of the project when more data is available.

## T6.7 Systematic assessment of the contribution of SLT and relation analysis to criminal NA

As specified, the goal of T6.7. is to compare the "classic" criminal networks (based on traditional sources such as telephone contacts/logs, wiretaps, or meetings) with the enhanced networks based on the additional information provided by the research activities of WP5 and WP6. Consequently, the systematic assessment is dependent on the development of the techniques and methodologies for the collection of additional information and the definition of those which will be included, also in consultation with end-users and LEA partners.

Furthermore, research has often emphasized that network topology and characteristics vary significantly across types of relations and types of data sources. ROXANNE has so far mostly worked on sample networks based on activities, which are relatively far from actual criminal networks (e.g. the NIST data, the CSI episodes). While these networks were helpful for the development of ROXANNE platform and to test its components, there are limited purposes in systematically testing the contribution of SLT as they result on actual criminal networks may significantly differ due to the characteristics of the data and relations. With the advancement of the project, the consortium hopes to be able to test the methods and techniques on

actual data or at least on more realistic simulations (work currently ongoing with other partners). This should enable to systematically assess the contribution of the project as required in T6.7.

## Status of deliverables

This WP6 has only one deliverable in the first year of the projects, the D6.1 Preliminary report on network analysis, which has been submitted on time, i.e at the end of month 9.

Additionally, partners in this work package have contributed significantly to deliverables in other ones. Those include the deliverables on data inventories (in WP4), ethics and legal issues (WP3 and WP10), user requirement analysis (WP2), the integration into ROXANNE’s pipeline (WP7), and the dissemination and exploitation of the project’s results (WP8 and WP9).

Deliverable	Status
D6.1 Preliminary report on network analysis (LUH)	Submitted in M9
D6.2 Criminal entity network construction (LUH)	Due in M18
D6.3 Analysis of criminal networks (LUH)	Due in M24

## Work planned for next Y2 period

The main focus in our plan for the second year of the project consists of the following:

- **Research work:** We will spend an effort in Y2 on investigating the network embedding methods and their applications in network analysis tasks. As these methods have resulted in state-of-the-art performance in a wide range of tasks, while being robust to domain knowledge presence, those are potentially useful for this project. More specifically, we will investigate the methods for the embedding of uncertain networks, which are often the case of the network constructed from WP5’s outputs.
- **System development:** We will extend our API server by adding more method choices for the existing network analysis tasks in our network analysis toolbox. The additional choices will focus on considering individuals’ and links’ properties in performing the tasks. Technically, we will provide these additional method choices through extending our current API server, and adding more user interactions into our network visualizer. Besides, we are planning to use the Data Fusion Bus (DFB) existing components (orchestrator, Kafka, ES). The provided API (work carried out and in progress) can be used besides for merging and filtering of different datasets (e.g. CSI matrices, social influence analysis datasets, community detection datasets, etc.) for visual data exploration delivered in WP7 by AEGIS using AEGIS FVT. Real-time prediction and analysis over constructed and existent networks of WP6, while respecting and gathering the user requirements from AEGIS, will be offered via the data fusion framework of DFB in order to support the data visualization and exploratory analysis.
- **Evaluation:** We will conduct the empirical evaluation of network analysis methods using a wide range of publicly available datasets, particularly the methods for embedding of uncertain networks. Regarding T6.7, we plan to develop the assessment methodology during Y2 and test it on the most suitable of the pilot/mock data currently in use.

## 7. Integration and visualisation of results

### T7.1 Design and definition of the ROXANNE system architecture M1-M6

During the first six months of the project, we focused on the definition of the architecture of the ROXANNE platform which will allow the integration of the technical components provided by the WP5 and WP6 and thus build the systems that will be tested during the field tests.

To do this, we translated the functional requirements (collected by WP2) into technical requirements, which then allowed us to define a software and hardware architecture;

The architecture defined in this task aims to be generic and flexible and allows to fulfil the criteria defined at the start of the project:

- **Agility:** separated teams could develop separate services
- **Scalability:** processing tasks are automatically load balanced between several services instances
- **Easiest deployment:** Each service is self-sufficient in its own container and will not affect other services

Finally, we also worked on the definition and specification of the API allowing interactions between the backend and the frontend (provided by T7.4)

This work was disseminated in both D7.1 and D7.2.

### T7.2 User management and access control M6-M30

This task consists of defining common, secure and central authentication mechanisms; For the first version of the Roxanne system, since we have decided to focus on the integration of a maximum number of technical components, we shift the work on this task to the second period.

### T7.3 Run-time data visualisation and exploratory analysis M4-M30

The objective of this task is to develop an environment that visualises social network relationships derived from multi-modal data (i.e., text, audio and video) and presents statistical indicators (e.g., number of individuals identified, the gender distribution, etc.) and other insights (such as an assessment of the importance of each individual, identification of communities, etc.) in an easy-to-digest way. Furthermore, this visualisation framework enables the end user to explore data in a high level through several interconnected, interactive charts that also allow filtering and drilling into more detailed information to reveal hidden relationships and insights.

During the first period of ROXANNE, a Graphical User Interface was prepared and integrated with the rest ROXANNE platform for a) retrieving information related to the raw multimedia files that were processed under the selected criminal case, and b) retrieving information about the social networks that were synthesised based on the available multi-modal data. A wide set of visualisation formats are currently supported, including networks, bar charts, pie charts and word clouds. Furthermore, a set of filters can be used for making the discovery and presentation of these insights as natural for the analyst as possible regardless of the amount of information.

Next steps include enhanced configuration and collaboration capabilities, as well as visualisation features based on the feedback collected from the first field test and the insights gained from the survey on user requirements.



## T7.4 Secure data export and exchange M6-M30

This task entails the design and implementation of a generic mechanism to export data for the use in external applications. At this phase of the project, we performed an investigation of possible data securing mechanisms that can be applied. Furthermore, access rights with respect to user roles have been investigated in order to implement data transformation and minimization. In the next period, ADDITESS will start implementing the secure data exchange and export approach. The task is expected to finish on M30 and the documentation of the work will be provided in D7.4.

## T7.5 Integration, ROXANNE system/platform setup and maintenance M4-M36

Once the work on the specification and design of the platform was finalized, we provisioned the necessary hardware for the implementation of the integration work and the development of demonstrators. We have followed an incremental integration process: once a component is ready it is installed, integrated and tested in a global processing chain. Below is a summary of the work carried out during the first year :

- Setup of two hardware platforms on Airbus and KEMEA sites
- Installation of the Kubernetes cluster on both platforms
- Implementation of APIs with technical partners (WP5 and WP6)
- Integration of a first version of Audio processing components (WP5) :
  - Voiceprint and Language-print extraction
  - Language identification
  - Age estimation
  - Gender identification
  - Speech to text
  - Named Entity Extraction on text
  - Topic detection on text
  - Voiceprints comparison
- Integration of a first version of Network Analysis components (WP6) :
  - Network construction
  - Social influence analysis
  - Community detection
  - Link prediction
- Integration of a first version of FVT (Forensic Visualisation Toolkit provided by T7.3)visualisation GUIs

This first work will allow the first ROXANNE system (prototype) to be set up, which will be presented to end-users during the first field test.

## T7.6 Integration of feedback from end-users M10-M36

This task is aligned with the execution of field tests. The start of the work is postponed by four months as is the case for the first field-test.

## Status of deliverables

Deliverable	Status
D7.1 Technical specifications and detailed architecture report	Submitted in M6
D7.2 Description of the integration toolkit, guidelines and plan	Submitted in M9
D7.3 Data visualisation and exploratory analysis toolkit V1	Due in M18
D7.4 Description of ROXANNE platform V1, data export and exchange	Due in M20

D7.5 Data visualisation and exploratory analysis toolkit V2,  
ROXANNE platform V2

Due in M32

## Work planned for next Y2 period

Our plan for the next period of the project consists of the following:

- Integration of video processing components
- Make processing chains configurable
- Integration of new version of components from WP5 and WP6
- Propose and implement secure and central authentication mechanisms
- Propose and implement a secure data exchange and export approach
- Guideline to install ROXANNE platform on LEA premise
- Enhance configuration and collaboration capabilities, as well as visualisation features
- Take into account LEAs feedback on the first version of the ROXANNE system
- Prepare the ROXANNE system V1 for the second field-test.

## 8. Field tests, user training and continuous testing

### T8.1 Development of end-user validation and performance test methodology

This task aims to evaluate the developed speech/text/video analytics, network analysis tools and technology both separately and combined ensuring in that way the proper evaluation of the right system features with the appropriate components. D8.1 has been the outcome of a joint effort among KEMEA and all related consortium partners, which outlined the exact validation framework and the test methodology that will be followed throughout the ROXANNE project. The whole process is based on the Structured planning framework of the European Operational Concept Validation Methodology (E-OCVM) that consists of 6 concrete steps, with certain revisions if the results are not satisfactory. It also entailed the validation management structure, the specific validation criteria as well as the process of performance testing and the specific metrics used for ROXANNE platform, underlining also the KPIs that must be measured in proportion to their objectives, as set by ROXANNE consortium. The above processes will take place both in the field tests and in the final implementation phase. With them as timestamps this dynamic process will be carried out by the technical managers almost continuously until the end of the project.

### T8.2 End-User training

The purpose of this task is to develop a set of training materials along with an e-learning training platform. At this phase of the project, ADDITESS has developed the training platform in order to support the training of end-users before each pilot. Regarding the first field test, ADDITESS in a collaboration with technology providers prepared the first volume of training material which is reported in D8.2 (submitted on M10). ADDITESS will continue the maintenance of the training platform and will start working on the second volume of training material (D8.7) scheduled for M19. The task will end on M32 where the field tests are expected to be finalized.

### T8.3 Field test planning

This task was about setting-up the field-tests plan for the field-tests of the ROXANNE project, including the methodology of the field-tests, the complete set-up of the tests for the end-users, as well as the schedule of the field-tests activities. The resulted Deliverable, D8.3, thoroughly outlined the Field Test Guidance Methodology (FTGM)<sup>5</sup> produced by KEMEA with the valuable feedback of all respective partners, which consists of 3 distinctive phases (preparation, execution, and evaluation), including also the reporting of potential risks and mitigation measures. This methodology will be applied before, during and after each field test, for their proper organisation, implementation and evaluation of their outcomes. It will be revised and thoroughly updated after the first field test, reported in the revised version of deliverable D8.5 ROXANNE integrated solution - field test planning in Month 18 of the Project (February 2021).

### T8.4 Design, implementation of field-tests

This task includes the organisation, implementation and evaluation/feedback integration of the 3 field tests around ROXANNE platform. During year one of the project, this task was involved in the organisation of the first field test of ROXANNE platform, including the overall implementation of the respective training activities. The unforeseen COVID-19 situation delayed all the activities included in this task, postponing for that reason the realisation of the first field test (planned for June 2020). KEMEA actively participated in all relative technical discussions of WP4-WP6 including the platform integration (for which we acquired the

---

<sup>5</sup> This methodology was based on the the Trial and Guidance Methodology principles (TGM) for the Driver+ (Driving Innovation in Crisis management for European Resilience) EU Project (<https://tgm.ercis.org/>)



respective server and installed it in KEMEA’s premises), as well as for the training platform integration (also in KEMEA’s premises). In addition to that, initial evaluation questionnaires have been sent for further feedback to all respective partners for the organisation of the first field test, which will be realised in September 2020.

### T8.5 Continuous testing

The installation of the platform at LEAs could not be done because of delays related to the supply of hardware as well as the delay related to the COVID19 pandemic. Nevertheless, we are applying continuous testing on the development platform (located at Airbus premises) in order to check the behaviour of components combined together and to evaluate the compliance w.r.t end-users requirements.

During the next period, we will support the LEAs in the installation of their internal ROXANNE platform so that they begin to test their operational data.

### T8.6 Field test results analysis and evaluation

This task was set to start in M9, after the initial planning of the first field test. However, due to COVID-19 restrictions, the field test has been moved to take place virtually on 30th of September 2020, thus this task will start producing substantial results after M14 (October 2020).

### Status of deliverables

Deliverable	Status
D8.1 Validation criteria list and performance test methodology	Submitted in M5
D8.2 Training manual Volume I	Submitted in M8
D8.3 Field Test Planning Guide	Due in M8, submitted M10
D8.4 1st Field Test report and recommendations	Due in M10, delayed to M13
D8.5 ROXANNE integrated solution - field test planning	Due in M18
D8.6 Final field test report and recommendations	Due in M34
D8.2-II Training Manual Volume II	Due in M19
D8.2-III Training Manual Volume III	Due in M29
D8.4-II 2nd field test report and recommendations	Due in M20

### Work planned for next Y2 period

For Y2, WP8 will focus on the:

- a. 1st Field Test organisation and implementation (organised on 2020-09-30, virtual session)
- b. Training platform integration and organisation of training sessions for the first field test
- c. Setup trainers and trainees accounts on the training platform
- d. Accommodate training material on the training platform
- e. Evaluation and feedback collection and analysis of the first field test
- f. Constant communication with ROXANNE technical team for any updates on the components, use cases and platform integration
- g. Organisation of second field test for the ROXANNE platform (implementation of FTGM cycle)
- h. Evaluation and feedback collection and analysis of the second field test
- i. Updates on the training platform and the sessions designed for the end users
- j. Prepare Training Manual Vol II



- k. Continuous testing of ROXANNE platform both on the development platform but also on end-user platforms (use of validation criteria and performance test methodology as described in D8.1)
- l. Submission of respective deliverables D8.4, D8.5, D8.7 and D8.9.

## 9. Dissemination, exploitation and communications

### T9.1 Conduct stakeholder analysis and compile a stakeholder contact list

The stakeholder list has been compiled and it includes inputs from the entire consortium. The list and the analysis is described in D9.4 which has been submitted to EC in M12. This stakeholder list is going to be a live document with limited access by partners (TRI and IDIAP). The contact list will be further used for dissemination and communication activities (e.g. sending newsletters).

### T9.2 Elaborate the dissemination and exploitation plan, IPR management

The consortium has finalised the communication and dissemination plan for the project (D9.3 - The project's dissemination and exploitation plan - submitted at M5). We have identified key target groups along with key messages for each year of the project. This plan includes various avenues to be leveraged for effective communication to our key stakeholders. As for the exploitation plan, partners are already involved in discussions relating to specifics to feasible business plan and optimum way to exploit this platform as a whole. Technical partners have also been in the process of completing exploitation questionnaires in this regard.

Other than this, CAP organises a WP9 meeting on a monthly basis and keeps track of all communication and dissemination activities by the entire consortium. This includes contributions to digital platforms (e.g. social media), participation in events, scientific publications, press releases etc.

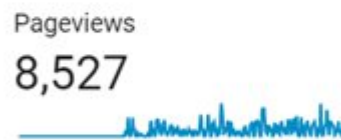
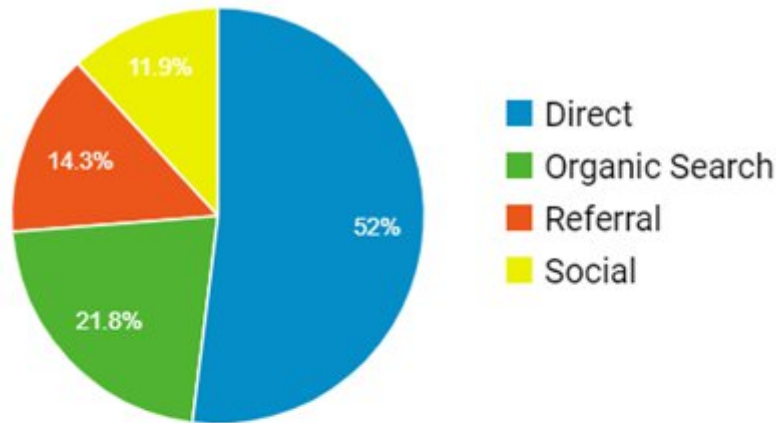
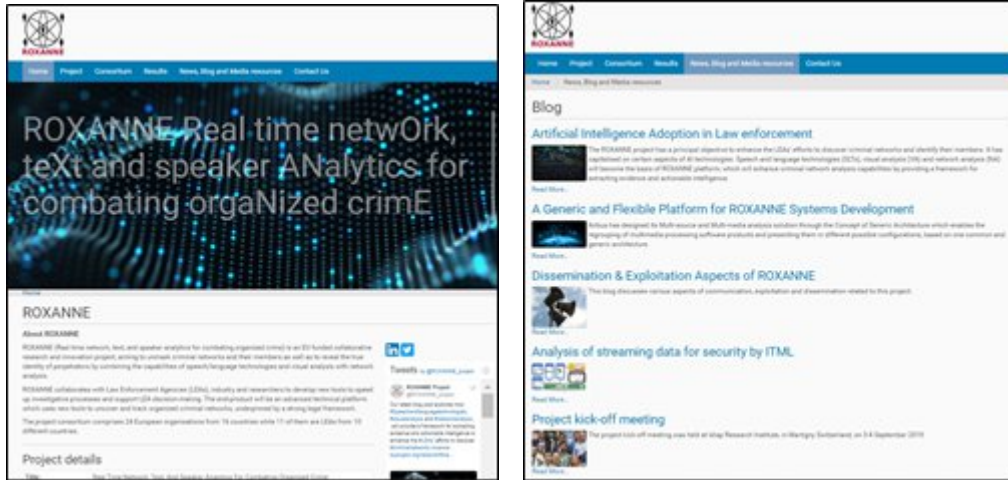
### T9.3 Elaborate the plan of communications activities

CAP and TRI have identified a custom plan for communication activities (D9.2 - The project's communications plan - submitted at M4) to target relevant stakeholder groups. This includes identifying and categorising stakeholders and target audiences, tools and channels of communication for each type of audience, along with the timelines for the same.

### T9.4 Promote the project identity and the project's website

The ROXANNE project now has a [LinkedIn account](#), [Twitter account](#) and [official website](#), all with a common graphic line. The social media handles are updated regularly and are gaining a significant. As for the website, it is updated at least on monthly basis. The website has sections relating to latest news, newsletter, blogs, project information etc.



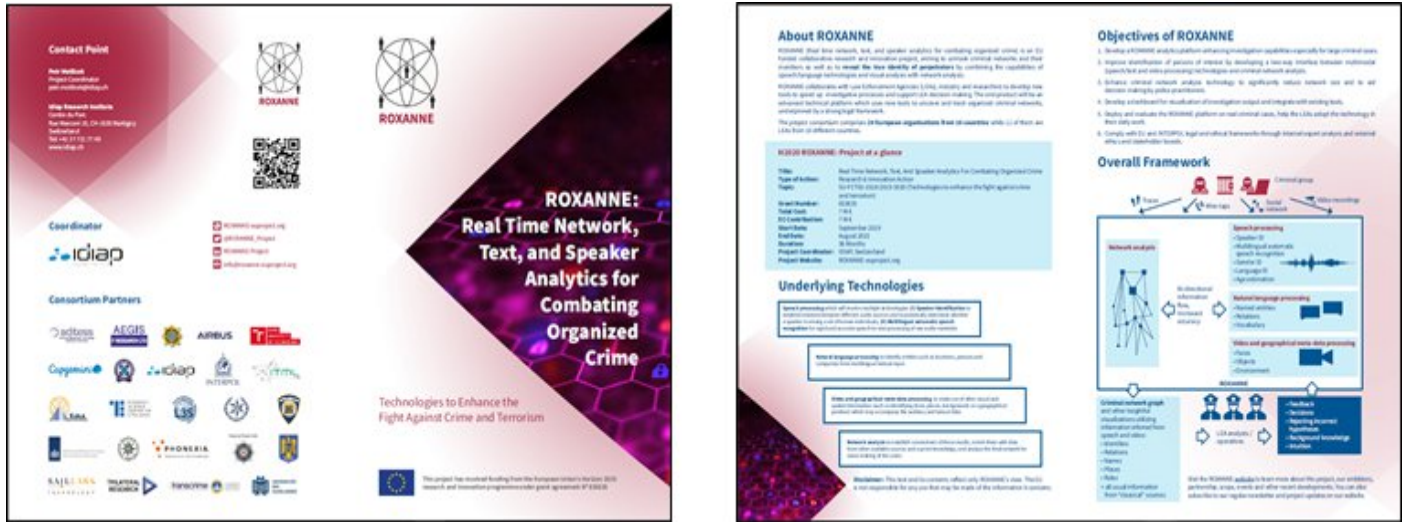


## T9.5 Prepare and use dissemination and communications materials

We have created a project brochure and a PPT presentation which is used for dissemination purposes. Furthermore, we have prepared and shared the first newsletter (sent to 70+ stakeholders in April 2020). Also, the second newsletter pending at M12 is under preparation and would be published soon. Further, TRI has been working closely with the consortium to finalise the script for the project video (to be ready for the first field-test event), which would be released in coming months.



Following are some of the screenshots of our dissemination material:



## T9.6 Exploit the project's results

This task starts in M20.

## T9.7 Convene the project's final conference and participate in others

This task starts in M30.

## T9.8 Prepare articles for peer-reviewed journals and conference presentations

Even though the project has just completed one year, we have successfully prepared multiple articles which are in the accepted/published/submitted stage in various events (with details below):

Partne	Date	Title	Paper Status
IDIAP	April 2020	Idiap NMT System for WAT 2019 Multi-Modal Translation Task	Published WAT Workshop, EMNLP2019 Hongkong
IDIAP	April 2020	Overview of the 6th Workshop on Asian Translation	Published WAT Workshop, EMNLP2019 Hongkong
IDIAP	April 2020	OdiEnCorp 2.0: Odi-English Parallel Corpus for Machine Translation	Published WILDRE Workshop LREC 2020
IDIAP	April 2020	German News Article Classification: A Multichannel CNN Approach	ETAERE 2020 conference, India, March 2020
IDIAP	April 2020	Idiap Submissions to Swiss German Language Detection Shared Task	Published Swiss-German Language Detection Shared Task, SwissText and KONVENS 2020
IDIAP	April 2020	Inferring Highly-dense Representations for Clustering Broadcast Media Content	Paper accepted to Prague Bulletin of Mathematical Linguistic (PBML) Journal
IDIAP	April 2020	INCREMENTAL SEMI-SUPERVISED LEARNING FOR MULTI-GENRE SPEECH RECOGNITION	Accepted in ICASSP 2020 conference, Barcelona, May 2020
USARR	April 2020	Distant Supervision and Noisy Label Learning for Low Resource Named Entity Recognition: A Study on Hausa and Yorùbá	Accepted by the workshop PML4DC under ICML 2020



IDIAP	July 2020	Idiap & UAM participation at GermEval 2020: Classification and Regression of Cognitive and Motivational Style from Text	Published GermEval 2020 Shared Task on the Classification and Regression of Cognitive and Motivational style from Text, SwissText and KONVENS 2020
IDIAP	July 2020	Idiap and UAM Participation at MEX-A3T Evaluation Campaign	Accepted by MEX-A3T@IberLEF2020
IDIAP	June 2020	Improving Speaker Identification using Network Knowledge in Criminal Conversational Data	Re-submission to ICASSP 2021
UCSC	June 2020	Robust link prediction in criminal networks: A case study of the Sicilian Mafia	Published
IDIAP	July 2020	BertAA: Bert fine-tuning for Authorship Attribution	Submitted to CoNLL 2020
BUT	July 2020	Analysis of ABC Submission to NIST SRE 2019 CMN and VAST Challenge.	Proceedings of Odyssey 2020 The Speaker and Language Recognition Workshop, available in iSCA archive.
BUT	July 2020	Probabilistic embeddings for speaker diarization	Proceedings of Odyssey 2020 The Speaker and Language Recognition Workshop, available in iSCA archive.
BUT	July 2020	Utilizing VOICES dataset for multichannel speaker verification with beamforming	Proceedings of Odyssey 2020 The Speaker and Language Recognition Workshop, available in iSCA archive.
LUH	July 2020	Experimental Evaluation of Scale, and Patterns of Systematic Inconsistencies in Google Trends Data	Workshop on Evaluation and Experimental Design in Data Mining and Machine Learning (EDML 2020) @ ECML PKDD

## T9.9 Policy recommendations

This starts in M30.



## Status of deliverables

Deliverable	Status
D9.1 Creation of the project's identity and website (TRI)	Submitted in M2
D9.2 The project's communications plan (CAP)	Submitted in M4
D9.3 The project's dissemination and exploitation plan (CAP)	Submitted in M5
D9.4 Stakeholder analysis and contact list (TRI)	Submitted in M12
D9.5 Report on the dissemination, exploitation (CAP)	Due in M18
D9.6 Report on the project's final conference (CAP)	Due in M36
D9.7 ROXANNE's recommendations (TRI)	Due in M36
D9.8 Final report on the dissemination, exploitation (CAP)	Due in M36

## Work planned for next Y2 period

For year 2, we will focus on increasing the contribution to dissemination activities from partners and recording their impact through such activities. In terms of dissemination material, we plan to have two more newsletters, one project video, 12 new blogs, one new brochure, and one leaflet (if needed) for the upcoming year. Moreover, we will continue to have regular updates on the project website and social media handles. We also intend to make the exploitation plan clearer after having the required discussions within the consortium. As for the deliverables, the only deliverable pending in Y2 for WP9 is D9.5 (Report on the dissemination, exploitation) [M18].

## 10. Ethics requirements

Following the imposition of WP10 onto the ROXANNE project, TRI took the lead in completing them owing to their experience in dealing with ethical and data protection issues. Many of the requirements were required by M1, owing to the large amount of information to be provided in a short time it resulted in several documents being incomplete.

### D10.1 H - Requirement No. 2

This deliverable relates to procedures for human participation in the ROXANNE project. Procedures and criteria for recruiting human participants in research, along with information sheets and informed consent forms were first drafted by TRI and included in D10.1. It was initially submitted in M1. It was reviewed at the 1<sup>st</sup> Ethics Check as being partially completed as the participant information sheet and informed consent form was not judged to be compliant with Article 13 of the GDPR. TRI re-drafted these documents, which were approved by the USAAR ethics review board. This document was internally reviewed by PHO, IDIAP, USAAR, SAIL, MOPS-INP, and IDIAP. The document was re-submitted in M9 and the requirement was considered completed by the reviewers.

### D10.2 H - Requirement No. 3

This deliverable relates to the opinions of ethics committees. TRI reviewed all research activities in the ROXANNE project to determine where human participation would take place. TRI's *prima facie* analysis was then discussed by email with task leaders to confirm the nature of human participation in these tasks, and whether each partner had an ethics committee that could review the use of human participants. None were found from the partners involved in using human participants, and nor were any national research ethics committees found to be appropriate to request an ethical review from. In light of this, TRI discussed human participation with the project's External Ethic Board in order to provide independent oversight. This document was submitted at M3, and reviewers at the 1<sup>st</sup> Ethics Check found it to be partially completed as TRI had not considered the ethics review committees of partners who are not involved in research using human participants. TRI contacted university and research organisation partners regarding the possibility of their research ethics committees reviewing the use of human participants in ROXANNE research activities. USAAR's Ethics Review Board approved these activities. This was reported in v2.0 of the document at M9. This document was internally reviewed by IDIAP, INTERPOL, and USAAR. At the 2<sup>nd</sup> Ethics Check, this document was considered completed and open for monitoring.

### D10.3 POPD - Requirement No. 6

This deliverable relates to special derogations on the rights of data-subjects. TRI developed a table for partners to fill out regarding special derogations regarding rights of data subjects. This was due in M1 and was submitted on time. This was regarded as partially completed at the 1<sup>st</sup> Ethics Check owing to some partners not providing information by M1. TRI continued to collect information for this document, and submitted information from all partners in v2.0 of the document at M9. It was internally reviewed by ADITESS, LTED, and IDIAP. At the 2<sup>nd</sup> Ethics Check, this document was assessed as completed.

### D10.4 POPD - Requirement No. 7

This deliverable relates to details of the partner's data protection officer or privacy policies. TRI developed a table for partners to provide information on their data protection officers, or privacy policy. This was due in M1 and was submitted on time. Reviewers at the 1<sup>st</sup> Ethics Check, regarded as partially completed and open for monitoring as some partners had not provided the required information. TRI continued to collect required information and submitted v2.0 at M9. It was internally reviewed by INTERPOL, AEGIS, ITML, and IDIAP. Reviewers at the 2<sup>nd</sup> Ethics Check noted that some partners had not provided information about their data protection officers and that web links to their privacy policy were no longer accessible (some weblinks had expired). TRI is now in the process of acquiring the missing information to be submitted in v3.0 of this document.

### D10.5 POPD - Requirement No. 8

This deliverable relates to details of anonymisation/pseudonymisation techniques and technical and organisational measures taken to safeguard data-subjects. TRI contacted project partners for information about their use of anonymisation/pseudonymisation and the technical and organisational measures they are taking in the project to protect the rights of data-subjects. This was due in M1 and submitted on time. Some partners did not provide information on time. Reviewers at the 1<sup>st</sup> Ethics Check regarded this document to be partially complete and requested more information on anonymisation/pseudonymisation and for partners to provide the required information. TRI reviewed what information about different types of technical and organisational measures should be included (use of anonymisation/pseudonymisation, use of encryption, system resilience/cyber security, ability to resort data, regular evaluations of technical and organisational measures), IDIAP consulted the Project Officer who also suggested including information regarding which persons at each partner could have access to research data. BUT coordinated the collection of this information, particularly from technical partners. TRI consulted technical partners and re-drafted the sections regarding use of anonymisation/pseudonymisation techniques across the project. This was re-submitted as v2.0 at M9. It was contributed to by all partners, and was discussed by both Internal and External Ethics Boards. At the 2<sup>nd</sup> Ethics Check, reviews considered the requirement to be completed.

### D10.6 POPD - Requirement No. 9

This deliverable relates to the lawfulness of international data transfers. TRI developed a table for partners to provide information regarding any international data transfers. This was due at M1 and was submitted on time. Some partners did not provide information before submission. At the 1<sup>st</sup> Ethics Check, reviewers considered that, owing to the missing information and some information not immediately relevant to the ROXANNE project, that this requirement was partially completed. TRI contacted partners for the missing information. This was submitted as V2.0 at M9. It was internally reviewed by PSNI, AGS, IDIAP, CAP, INTERPOL, RMIA, and IDIAP. At the 2<sup>nd</sup> Ethics Check, the reviewers stated that it was not clear whether data on surveys distributed by INTERPOL was in conformity with the laws of the countries where the data itself was collected. This document will be revised by TRI and the requested information will be provided in V3.0.

### **D10.7 POPD - Requirement No. 10**

This deliverable relates to informed consent procedures. TRI provided initial information regarding informed consent procedures for this document based on discussions with partners. It was submitted on time at M1. At the 1<sup>st</sup> Ethics Check, reviewers noted that it was not clear which legal bases would be relied upon in different research activities, and that informed consent forms were not Article 13 GDPR compliant. TRI revised this document in its entirety and submitted V2.0 at M9. It was internally reviewed by UCSC, KEMEA, AGS, SAIL, and IDIAP, and was discussed by both the Internal and External Ethics Boards. At the 2<sup>nd</sup> Ethics Check, reviewers considered the requirement partially completed and determined that there was a need to clarify that partners who serve a public function will be processing data in the project and are relying on legitimate interests outside the scope of their public function. TRI will update this document at V3.0 to provide the requested information.

### **D10.8 POPD - Requirement No. 11**

This deliverable relates to secondary use of data. At the beginning of the project, TRI provided partners with a table to provide information on their secondary use of data. This was not completed by all partners, and, owing to submitting this deliverable at an early stage of the project, the legal basis for processing was not definitively decided upon. At the 1<sup>st</sup> Ethics Check this requirement was judged to be partially completed. TRI completely revised this document following discussions with partners about their re-purposing of data. This was re-submitted as V2.0 at M9, it was internally reviewed/contributed to by MIA, PSNI, UCSC, INTERPOL, IDIAP, and SAIL. This document was also discussed at Internal and External Ethics Board meetings. At the 2<sup>nd</sup> Ethics Check, reviewers noted that members of the External Ethics Board had suggested continuing discussions on legal bases for processing and agreed on continuing this discussion, and that it was not clear how data to be repurposed would be used. TRI will review this document and discuss it with the External Ethics Board (along with any other documents that need discussion prior to being re-submitted) at the next External Ethics Board meeting. This will be submitted at V3.0 to provide the requested information.

### **D10.9 POPD - Requirement No. 12**

This deliverable relates to profiling. TRI conducted an initial evaluation of profiling in the ROXANNE project. This was submitted on time at M1. At the 1<sup>st</sup> Ethics Check, this document was considered to not have been completed. This was due to a misunderstanding between whether identification of persons potentially engaged in criminality constituted profiling. TRI completely revised this deliverable to discuss the presence of profiling in the ROXANNE project. V2.0 was internally reviewed by UCSC, LUH, USAAR, INTERPOL, and IDIAP, and was resubmitted at M9. This document was also discussed at Internal and External Ethics Board meetings. This requirement was considered to have been completed at the 2<sup>nd</sup> Ethics Check.

### **D10.10 POPD - Requirement No. 13**

This deliverable relates to processing of data related to criminal convictions and offences. TRI conducted a survey of all partners, and particularly LEAs, to determine whether they intended to process data related to criminal convictions and offences. Although some partners did not provide information, of those that did, none expressed an intention to process such data; this document was submitted at M4. Owing to the missing information, this document was assessed as being partially completed at the 1<sup>st</sup> Ethics Check. TRI

continued to gather information from partners, with one (LTEC) noting an intention to process data related to criminal convictions/offences. At the point of re-submitting this document, LTEC were still in consultation about whether they could process these data. The document was internally reviewed by KEMEA, HP, INTERPOL, and IDIAP, and was re-submitted at M9. At the 2<sup>nd</sup> Ethics Check, the reviews accepted that this document will be updated once more information is available. TRI has continued to discuss this issue with LTEC and have now acquired the necessary information. TRI will continue discussions with other LEAs within the project regarding personal data relating to criminal convictions and offences and update this document as V3.0 with the requested information.

#### **D10.11 POPD - Requirement No. 14**

This deliverable relates to the need for a data protection impact assessment (DPIA). TRI conducted an overview of the tasks in ROXANNE to determine which involved high-risk processing and so required a DPIA to be conducted. This was internally reviewed by IDIAP and KEMEA, and submitted at M4. It was regarded as completed and open for monitoring by the reviewers at the 1<sup>st</sup> Ethics Check. The reviewers asked for the DPIAs for the high-risk processing to be included in the next iteration of the Data Management Plan (M18). TRI has developed a DPIA template for partners to use and is coordinating the DPIA process for the project.

#### **D10.12 POPD - Requirement No. 15**

This deliverable relates to the composition of the Ethics Board. TRI provided information about the appointment and membership of the Ethics Board. This was submitted at M1. At the 1<sup>st</sup> Ethics Check, this requirement was regarded as completed, and the reviewers recommended appointment of other external persons with expertise in the GDPR and Law Enforcement Directive, TRI recruited another member of the External Ethics Board with appropriate expertise. The External Ethics Board also requested that this document be altered to accurately reflect their unpaid and voluntary role TRI did this, and provided additional information about the nature and membership of both the Internal and External Ethics Boards. This document was revised to explain new appointments to the boards, and the changes requested by the External Ethics Board. This document was also discussed at Internal and External Ethics Board meetings. The document was internally reviewed by AIRBUS, UCSC, INTERPOL, and IDIAP, and was re-submitted at M9. At the 2<sup>nd</sup> Ethics Check, this document was again considered to be completed.

#### **D10.13 GEN - Requirement No. 16**

This deliverable is a report by the Ethics Board. Owing to the unpaid and voluntary role of the External Ethics Board members, the project has sought to reduce the workload which it asks of Board members. TRI drafted a report for the External Ethics Board, which was approved by them and submitted at M4. At the first Ethics Check, this was rejected as the report was not written by the members of the External Ethics Board. TRI consulted with the External Ethics Board as to whether they would be willing to write this document which they declined owing to the substantial efforts required in a voluntary and unpaid role. For re-submission in M9, TRI collated minutes of meetings with the External Ethics Board and a record of letters between this Board and the project. At the 2<sup>nd</sup> Ethics Check, the reviews determined that this record of discussions did not constitute a report by the External Ethics Board and again determined that the requirement was not completed. IDIAP is going to discuss the best way forward for this requirement with

the Project Officer in order to find a solution that is not overburdensome for the members of the External Ethics Board.

### **D10.14 GEN - Requirement No. 17**

This requirement is due at M12. Owing to the first field-test being delayed until the end of M13, and the project wanting to hold a meeting with the External Ethics Board after the field-test, this report will be delayed. It will be submitted as soon as possible. The nature of this report will be outlined in conjunction with the previous report (D10.13).

### **D10.15 GEN - Requirement No. 18**

This is due to be submitted in M30.

### **D10.16 M - Requirement No. 19**

This deliverable relates to managing the risks of misuse and mass surveillance generated by the ROXANNE platform. TRI drafted this document to analysis risks of misuse and mass surveillance. This was discussed with partners who provided input and internal review, including UCSC, INTERPOL, SAIL, MUP, and IDIAP. This document was also discussed at Internal and External Ethics Board meetings. The document was submitted at M9. At the 2<sup>nd</sup> Ethics Check, reviewers considered that the requirement was completed.

### **D10.17 OEI - Requirement No. 20**

This deliverable relates to details of the AI/data mining system to be developed/used in the ROXANNE project. This document was drafted by TRI, AIRBUS, and IDIAP. It was submitted at M1. At the 1<sup>st</sup> Ethics Check the requirement was regarded as having been completed. At the 2<sup>nd</sup> Ethics Check, the reviewers noted a discrepancy in the terminology of 'closed case' and 'cold case'. TRI will revise this document to clarify that it is unforeseeable that 'cold case' data will be accessed by the consortium.

### **Report on the meaning of 'privacy preserved'**

At the 1<sup>st</sup> Ethics Check, the reviewers noted several uses of the term 'privacy preserved' and requested a report to explain the meaning of this. TRI drafted an initial report, which was contributed to by SAIL, LUH, UCSC, and IDIAP. This was submitted at M9. At the 2<sup>nd</sup> Ethics Check, the reviewers regarded this requirement as complete.

### **Detailed report on profiling methods**

At the 1<sup>st</sup> Ethics Check, the reviewers requested a detailed report on how the rights and freedoms of data-subjects would be safeguarded during any profiling activities. This was discussed at Internal and External Ethics Board meetings. LUH created an initial draft which TRI, UCSC, LUH, INTERPOL, and IDIAP also contributed to. This was discussed at Internal and External Ethics Board meetings. This was submitted at M9. At the 2<sup>nd</sup> Ethics Check, the reviewers regarded this requirement as complete.

## Incidental Findings Policy

At the 1<sup>st</sup> Ethic Check, the reviewers requested an incidental findings policy. TRI drafted an initial policy. UCSC, KEMEA, INTERPOL, NFI, and IDIAP also contributed and provided internal review. This was discussed at Internal and External Ethics Board meetings. This was submitted at M9. At the 2<sup>nd</sup> Ethics Check, the reviewers regarded this requirement as complete.

## Status of deliverables

Document	Status
D10.1	Completed
D10.2	Completed
D10.3	Completed
D10.4	Partially Completed
D10.5	Completed
D10.6	Partially Completed
D10.7	Completed
D10.8	Not Completed
D10.9	Completed
D10.10	Partially Completed
D10.11	Completed
D10.12	Completed
D10.13	Not completed
D10.14	Not yet submitted
D10.15	Not yet submitted
D10.16	Completed
D10.17	Completed
Report on the meaning of 'privacy preserved'	Completed
Detailed report on profiling methods	Completed
Incidental findings policy	Completed



## Work planned for next Y2 period

In Year 2 of the ROXANNE project, TRI will coordinate the revisions to ethics requirements as requested by reviewers at the 2<sup>nd</sup> Ethics Check and submission of D10.14. TRI will also continue to organise Internal and External Ethics Board meetings. TRI will also manage revisions to WP10 deliverables that are 'living documents' and will be subject to revisions as more information is provided by partners, particularly in relation to datasets to be processed by LEAs.

## 11. Problems, solutions and deviations

Problems encountered	WP	Resolving Action	Impact/Delay/Deviations
Some of LEAs announced an interruption in their project contribution due to the lock-down situation in Spring 2020.	WP 2	The message was taken into account by the coordinator and the consortium tried to get larger involvement of other internal LEAs during that period.	These issues did not impact and delay any of the work in the project.
Lock-down in March - June 2020	All	The COVID19 caused significant problems during the Y1 of the project, especially related to sudden home-office work, unexpected family situations of many consortium members, impossibility to access work PCs during the first weeks of lock-down, ban for travelling, etc. For many partners, the situation got complicated in hiring experienced staff (e.g. postdoc) as they often come from non-EU countries. This situation is still ongoing.	Some of the project tasks were delayed, specifically the works related to the 1st field-testing planned for M10, but moved to some of next months. In M12, we decided for the whole remote organisation of the event, as the problems with travelling abroad did not really improve for many of the partners. Several deliverables were delayed due to that. Hiring experienced staff is still an issue for many EU countries, especially for academic partners.
Problems with accessing real-data, shared by LEAs	WP 2	The consortium did not expect such huge complications in accessing real-case data during the first phase of the project.	The situation got partially resolved by a decision to collect the data (motivated by real use-case) by the consortium. For Y2, we have a promise from several internal LEAs to have access to their data (under the conditions defined by WP10 deliverables). At the same time, several LEAs agreed to test the ROXANNE platform by them. Eventually, ROXANNE consortium agreed to include another LEA partner (ZITiS) with an

			agreement to collect another set of multimodal data through real interception systems.
Exploitation activities are not moving fast enough	WP 9	Due to the large consortium (24 partners), a common agreement on how the final platform could be exploited by LEAs at the end of the project seems complicated and requires more work than expected.	The consortium realized this issue and keeps organising regular telcos to discuss this issue so that an agreement will be reached shortly.
WP10 not planned during the proposal	WP 10	Work for WP10 requires significant involvement of all ROXANNE partners. This is specifically the case for the WP10 leader - TRI. Further, significant work is also required from the external ethics board.	This work was not planned during the project proposal and thus does not occur in the budget associated with each partner. Since TRI had to invest significantly more effort to WP10 than other partners, a possible budget reshuffling will be anticipated at the end of the project. Further, although very experienced and motivated members took part of the external ethics board, the consortium has an issue to request from them extensive amounts of work (e.g. writing a complete deliverable) as we would not be able to remunerate them. The consortium, as well the members of the Board have noted, it does not seem reasonable to ask unpaid volunteers to write extensive reports in their spare time. We therefore wish to reduce the work we are asking Board members to perform, whilst trying to respect the requirement



			<p>from the Commission. Our intention for revising the D10.13 report is to partially draft it with an outline of several topics, while exploiting a record of our discussions from meetings, and then allow External Ethics Board members to write their own comments in addition to this.</p>
--	--	--	--