This project that has received funding from the European Union's Horizon 2020 - Research and Innovation Framework Programme, H2020-FCT-2015, under grant agreement no 700381.

Analysis System for GAthered Raw Data

ASGARD

Instrument: Research and Innovation Action proposal

Thematic Priority: FCT-1-2015

D4.1. Demonstrations Planning

Deliverable number	D4.1	
Version:	1.0	
Delivery date:	January 2019	
Dissemination level:	Public	
Classification level:	Public	
Status	Final	
Nature:	Report	
Main author(s):	Dimitrios Kavallieros	(KEMEA)
Contributor(s):	Pavel Gladyshev Youssef Bouali, Vicari Salvatore, Ernesto La Mattina Anabela Filipe, Antonio Fonseca Armin Vogl	(NUI UCD) (ENG) (PJ) (BMI)

DOCUMENT CONTROL

Version	Date	Author(s)	Change(s)
0.1	15/10/2018	Dimitrios Kavallieros	ToC, information on sections
0.2	19/10/2018	Bouali Youssef, Vicari Salvatore	Updated version of ToC
0.3	6/11/2018	Dimitrios Kavallieros	Circulation of new version for input
0.4	9/11/2018	Armin Vogl, Anabela Filipe	Section 2.2, Section 2.3
0.5	16/11/2018	Dimitrios Kavallieros	Section 3 new version
0.6	4/12/2018	Dimitrios Kavallieros	Section 1, 2 new versions
0.7	14/12/2018	Pavel Gladyshev	Section 2.1
0.8	15/12/2018	Dimitrios Kavallieros	Final version for review circulated
0.9	18/12/2018	Romaios Bratskas	Quality Assurance of the deliverable
1.0	21/12/2018	Dimitrios Kavallieros	Review comments addressed-final version

DISCLAIMER

Every effort has been made to ensure that all statements and information contained herein are accurate; however, the Partners accept no liability for any error or omission in the same.

This document reflects only the view of its authors and the European Commission is not responsible for any use that may be made of the information it contains.

Table of Contents

Annexes

Tables

Figures

1 Introduction

1.1 Overview

The DoA describes this deliverable as:

D4.1 - Demonstrations Planning. [month 27]

The main objective of this document is to capture the high-level information regarding the internal trials that will be conducted in Ireland, Austria and Portugal.

- In Ireland, UCD will describe the "*Open Source Intelligence Monitoring with ASGARD tools*" scenario.
- In Austria, AT/MOI will describe the "*International migration threat*" scenario.
- In Portugal, PJ will "*Cybercrime Investigation with ASGARD tools*" scenario.

1.2 Relation to other deliverables

This deliverable is related to the following other ASGARD deliverables:

Receives inputs from:

Deliv. #	Deliverable title	How the two deliverables are related
D3.1	Use cases definition and end-user requirements report	D4.1 provides high level (Public) information based on information depicted in D3.1.

Table 1 – Relation to other deliverables – receives inputs from

Provides outputs to:

Deliv. #	Deliverable title	How the two deliverables are related
D4.2	Interim Trial Results	D4.1 will provide the framework for the interim trials

Table 2 – Relation to other deliverables – provides outputs to

1.3 Structure of the deliverable

This document includes the following sections:

- Section 2: In this section the overall time plan of the interim trials and final demonstrations is depicted. Furthermore, the three use cases are described and information regarding each one is provided (e.g. actors, dataset, hardware and software requirements etc.)
- Section 3: In this section the final demonstrations of ASGARD are described in high-level
- Section 4: This section concludes D4.1

2 ASGARD interim trials and final demonstration planning

This section provides information regarding each of the three interim demonstration exercise of ASGARD (e.g. high level of the description of the scenario, actors, dataset etc,). UCD, AT/MOI and PJ will lead each interim demo exercise respectively. Figure 1 depicts the timeline of the interim trials and the final demonstrations of the project.

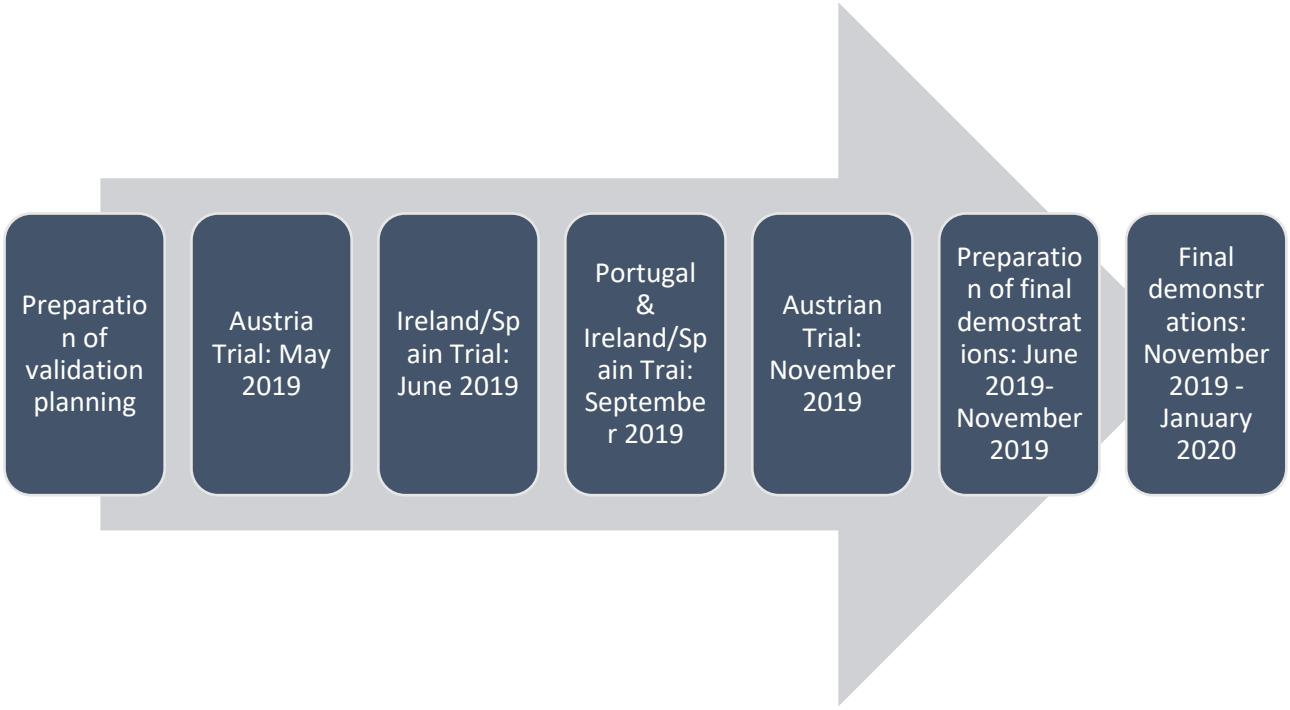

Figure 1: ASGARD Demonstrations Timeline

The following sections will describe the three interim demo exercises respectively. Each section will provide information regarding the description of each scenario, the actors that will be employed to run the scenario, the dataset needed and requirements in regard to software and hardware equipment.

2.1 Forensic evidences from Big Data

Name of the Scenario

Open source intelligence monitoring toolkit for digital forensics

Description

Many investigations require correlation of information extracted from stationary and mobile computing devices (digital forensics) with information extracted from open online sources, such as social media platforms. In some cases, monitoring of social media outlets together with wiretapping and other forms of lawful interception is necessary for obtaining additional evidence.

Given significant amounts of data arising from wiretapping and online investigative activities, ASGARD tools

are well positioned to assist Law Enforcement Agencies (LEAs) in automatic monitoring and processing of information extracted from online sources and correlating it with the information extracted from digital evidence.

This scenario is based around a simulated instance of dismantling a human trafficking net:

In a country where pimping is a crime, a woman reports to the police after being assaulted. An initial investigation determines that the is a street prostitute likely brought into the country by human traffickers. Although the victim refuses/afraid to divulge any information about the people she is working for, the investigators collect information about her contacts and her environment using street Closed-circuit television (CCTV) cameras and Social Media (Facebook, Twitter…), which leads to identification of several suspects. The activities of suspects are monitored for a period of time using online media, wiretapping, and other forms of legal interception. Eventually, sufficient amount of information is collected that leads to arrests and successful prosecution. ASGARD technology is used to process unstructured and semi-structured data collected during investigation to automatically extract useful actionable information.

Scenario Actors

Digital forensic examiners and investigators.

Envisioned Datasets

Datasets will be created for social media data, RAT logs, intercepted audio, forensic disk images. This data will be subject to WP12 (SELP) approval.

Hardware and Software requirements

PC/Laptop, Server capable of running ASGARD, additional requirements posed by relevant ASGARD software.

Time plan

Two trials are planned in June 2019 and September 2019 respectively.

Validation and Evaluation

The evaluation of the demo exercise will measure the functionalities of the respective tools in the framework of the scenario and against the user requirements depicted in WP3.

Legal and data privacy requirements

During the demo exercise only, simulated data will be used.

Operational support and logistics

Technical support will be required from VICOM and all the technical partners providing tools used in the trial to set-up and fine-tune ASGARD infrastructure.

2.2 Data exchange and situational awareness

Name of the Scenario

International migration threat.

Description

To comprehensibly analyse and understand the current irregular migration wave, it is necessary to understand the concept of a qualified migration of many humans as well as the individual, correlating migration causing factor from a specific - non-legal - migratory point of view. Therefore, a specific analysis model has been created.

The model enables the systematic survey of the specific locations that migrants are using as their migration route stations, starting with the migration source country, along their migration routes, and up to the destination country. Simultaneously, the model takes reasons that are under constant developing into consideration, and which are finally causing and triggering irregular migration and secondary movement (secondary migration). That is done by redefining the concept of push and pull factors, and by systemically connecting each factor with the relevant socio-psychological implications considering the different values of each society, region, or ethnic traditions, and – of course – the different factor packages with each other.

The migration analysis model for a country or location is built through the continuous processing of open source information mixed with internal data. The evaluation of the actual and future situation in an area is determined by the Push and Pull factor list (PPF). Each area has its own PPF list, which is created by experts, who have sufficient knowledge about a location. Each term in the PPF list is used to automatically detect topic specific information within the collected content (e.g., within the documents). As a first result, the officer/user gets a list of found snippets. The officer/user can decide very quickly if the content is important and if it should be further used. In a second step, the confirmed documents are used to update the key risk indicators in the PPF list for that area. The key risk indicators in the migration model provide an early signal of increasing risk exposures in that area. Finally, the results from the migration model are used to generate reports to support strategic decision processes for the migration topic.

Scenario Actors

Analysts and researchers

Envisioned Datasets

GDelt-Project, Sputnik News, Relief web and RSS feeds.

Hardware requirements

Server landscape, PC/Laptop; internet access (wireless). Additionally, the requirements that go with the envisioned tools.

Software requirement

The following functionalities/requirements should/must be enabled by tools and software (some already being available by ASGARD). E.g., crawler/source identification; Uniform Resource Locator (URL) ; scraping (source) data repository; RSS feed reading (IN rss URL – out .xml), e-mail ingestion (IN e-mail account – OUT text with link extract); social media monitoring - Ubicity (AIT) (Hashtag filter, key word filter); site map extractor, pdf to txt; text classification; weak signal analysis using linguistic markers/classifier/topic miner; text analytics pipeline (mulit lingual translation, natural language processing; video and audio analytics (meta data extractor, text concept detection; graph database (video analytic pipeline (video summarization, person identification); analysis tools (time line, network analysis, Geographic Information System (GIS) for maps). Figure 2 bellow depicts the process of the use case based on the tools, services and data that will be employed to smoothly execute it.

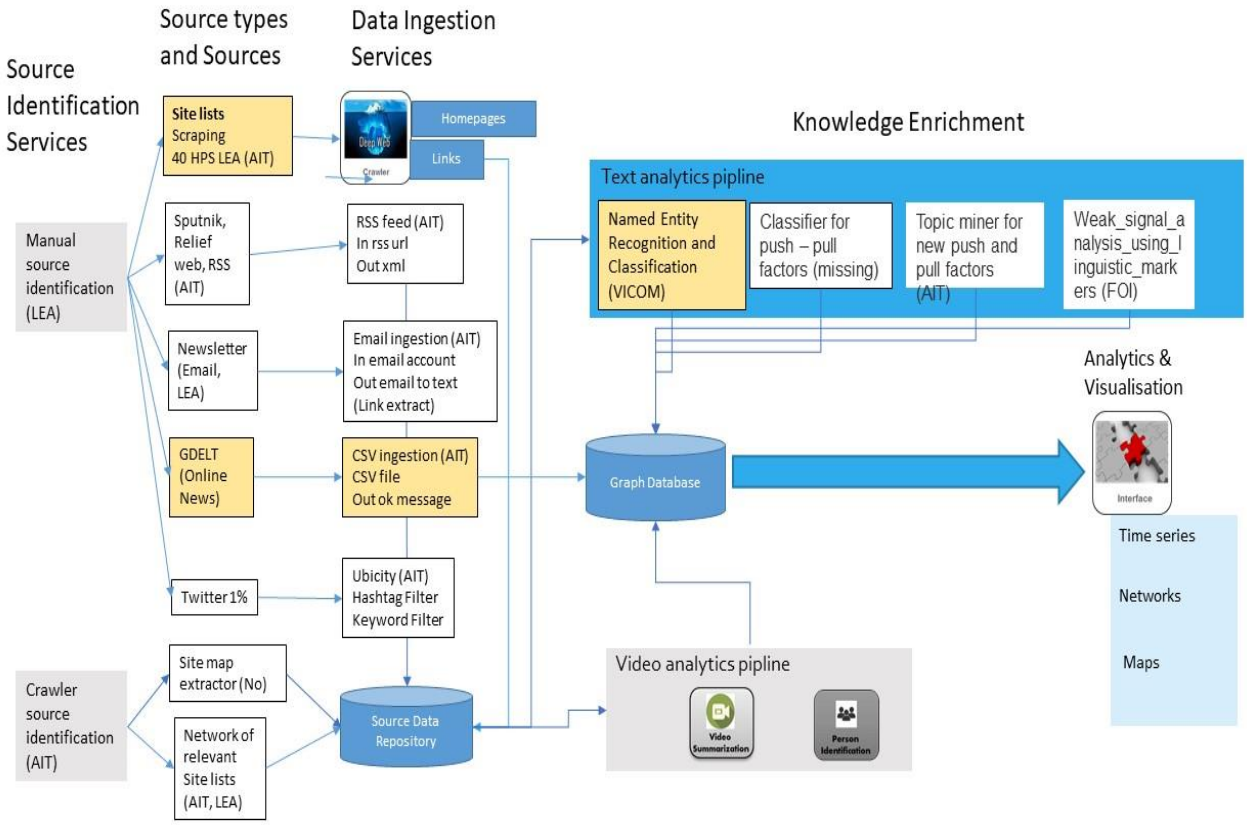

Figure 2: International migration threat use case execution

Time plan

The Austria Trial will take place in two different time periods. The 1st one will be at May 2019 while the second will take place at September 2019.

Validation and Evaluation

The evaluation of the demo exercise will measure the functionalities of the respective tools in the framework

of the scenario and against the user requirements depicted in WP3. Furthermore, the functionalities of ASGARD will be tested in order to evaluate the data exchange between LEAs and network operators as well as if it is enchasing the operational and situational awareness of the end-user.

Based on the aforementioned tools and services the following high-level software requirements (goals) are identified.

Goals:

- Pipelines automized as much as possible (all mentioned tools have to interact)
- Improvement of the situation awareness picture
- Improvement of information exchange

Legal and data privacy requirements

For the execution of the use case only open source data (no personal data) will be used.

Operational support and logistics

Technical support will be needed during the preparation and the operation/trial from all the partners owing the tools needed for running the demonstration, as well as support for installing the tool landscape at the BMI/trial location premises.

2.3 Analysis of cyber offences

Name of the Scenario

Cybercrime Investigation with ASGARD tools.

Description

Cyber offenses (and many other cyber cases) typically involve a huge amount of data to process, analyze, and correlate. Such data includes both structured data and unstructured data and typically comes from third parties' platforms. The uniformisation and automation of collection and/or analysis of such data is of upmost importance to speed up the investigative process. In this sense, Investigators who are not experts in digital investigations will focus the PT ASGARD trial on the usage of tools for analysis of digital artifacts.

This will require easy-to-use tools with user-friendly interfaces and provide LEA's with the possibility to substantially decrease the number of cases that require analysis by an expert (or in lab analysis), thus contributing to make digital investigations more accessible and reduce their cases backlog.

In this demo investigations we will approach four different use cases of criminal investigation in the scenario of cyber offences.

Our trial use cases will be the following:

- Investigation of cybercrime in a telecom provider. Pre-processing and analysis of 6 computers and 8 mass storage devices
- Investigation of a cybercrime in a telecom provider. Processing of metadata from audio interceptions output of preformatted transcription documents
- Cascade fraud case. Processing of huge quantities of PDF files containing commercial invoice information
- Telecommunication data from organized crime. Processing of huge quantity of PDF files.

Scenario Actors

Polícia Judiciária's officers – around 20 persons

Envisioned Dataset

We will provide four datasets with all the data concerning the cases for this scenario. This data will be subject to WP12 (SELP) approval. This data will be manipulated in a restricted isolated environment.

Hardware and Software requirements

The hardware will consist of two servers already allocated to the ASGARD project and a set of desktop computer belonging to the PJ IT infrastructure.

The software will consist of the, at the time, current configuration of the ASGARD software plus some small software utilities that maybe needed. MS Office and Adobe Acrobat reader will also be needed.

Time plan

Presentation of the use cases and the tools followed by participant's evaluation and testing. The demonstration exercise will take place in September 2019, 2 days with morning and afternoon sessions: 2 x 7 hours.

Validation and Evaluation roadmap:

The evaluation of the demo exercise will measure the functionalities of the respective tools in the framework of the scenario and against the user requirements depicted in WP3. Furthermore, ASGARD solution will be tested in order to evaluate its effectiveness in the entire prosses of analysing cyber-crimes.

The tools for each use case will vary:

- All the Main User Interface and Orchestration Framework will be needed
- Legacy Adapter
- PDF to text
- Metadata Extractor
- NERC tagger
- Keyword Search and Discovery
- Text Analysis
- Location Geotagger for documents
- File Recovery System
- Knowledge Export

We preview that some format transformation tools will be needed to translate for example eXtensible Markup Language (XML) into tabular formats, or to convert txt files into simple word documents.

Each use case will be dwelt separately. After a first demonstration of the tools before the participants, the participants will evaluate the same tools according to their own requirements that should be adjusted to the tool performance levels.

Legal and data privacy requirements

All the datasets that PJ will provide concern already closed cases. However, concerns of data ownership and privacy will be evaluated according to the WP 12 rules.

Operational support and logistics

PJ should provide installations and technical requirements for the trial.

3 ASGARD Final Demonstrations

This section will present the framework of the final demonstrations of the project. Three final demonstrations have been envisioned, in the Netherlands, Belgium and Greece hosted by TNO/EUROPOL, NICC and KEMEA/ Hellenic Police Forensic Science Division respectively. Furthermore, except the aforementioned events, synergies and participation in EU events will be explored in order to maximize the impact of the project and the demonstration of the results.

All three demonstrations will have two main objectives:

- Demonstrate and disseminate the tools developed throughout the project
- Acquire experts' feedback in order to evaluate the ASGARD tools.

Following the feedback received from the interim demonstrations, the technical partners will update and enhance the tools accordantly in order to employ the new version of the tools for the final demonstrations.

For the collection of the feedback tailored questionnaires will be drawn based on the questionnaires that will be developed in T4.2 for each tool and trial. Nevertheless, the questionnaires will measure the time needed to setup and configure each tool appropriately, the added value compared to the tools used now from LEAs, usability, user friendliness and measure the specific objectives set for each tool.

4 Conclusion

4.1 Summary

In this document we have described the overall time-plan of the interim demo exercises and final demonstrations as well as the respective description of each scenario.

In Section 2 we have described the interim demo exercises. To this extend, the description, the actors needed, requirements as well as the validation criteria have been presented among other information.

In Section 3 we have presented the plans and objectives of the final demonstrations of the project.

In summary, we have set up the framework of the ASGARD demonstrations for each of the interim trials as well as the framework of the final demonstrations.

4.2 Evaluation

This report provides inputs from all the core partners that will be involved in the interim trials and demonstrations, identifying all the needs for the execution of these activities.

4.3 Future work

This document will be the basis to carry out the interim trials in T4.2 as well as for the evaluation criteria of each trial.

ANNEX I. GLOSSARY AND ACRONYMS

Term	Definition / Description
LEAs	Law Enforcement Agencies
CCTV	Closed-circuit television
PPF	Push and Pull factor
XML	eXtensible Markup Language
URL	Uniform Resource Locator
GIS	Geographic Information System

Table 3 – Glossary and Acronyms