1. Needs

Current workflow is defined by:

1. Types of data - structured and unstructured, mostly (1) location data, (2) social media structured data (understood as relationships established between 2 entities and their frequencies), (3) textual data and (4) images.

2. Types of analysis

1. network analysis
2. profiling
3. pattern of life
4. content analysis & sentiment analysis

3. Domain/ working field - national security

1. Preventing and countering espionage, terrorism, cross-border organized crime (activities of international criminal organizations) that by nature and extent affect national security (this could be of interest to LEAs as well)
2. Protection of democratic values (preventing violent extremism, far-right and far-left extremism, radicalization etc.)
3. Insider threats

Summary of job description:

The analyst has in-depth knowledge and understanding of a particular domain, knows and employs several methods of analysis to deliver evidence-based reports. Analysis is mostly descriptive (with diagnostic purpose) or explicative (interpretation of causality and correlation relations).

He is responsible with generating insight from data, but **is not** responsible with collecting or cleaning data.

The analyst usually works with deep data rather than big data - meaning he integrates data about one subject of interest, but from various data sources.

Detailed workflow and decision points:

1. Receives requests - discusses, clarifies and works alongside teams within the organization or the management team to define specific and clear needs for information.

2. Consults available data sources (the data is clean at this point).

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

3. Evaluates data sources (credibility & relevance).
4. Evaluates whether the data is enough. If not, makes formal or informal requests for information and/or proposes new collection.
5. Once he/she has enough information, pursues to correlate, analyze and integrate data.
6. Delivers a text report which emphasizes most relevant aspects, while distancing the facts from assumptions in an evidence-based reasoning model.
7. The report is peer-reviewed.
8. When the situation requires it, the analyst should propose indicators to be monitored.

2. Expected results/ learning objectives

The main focus of the project is on nontechnical skills. Maximizing the business value of data analysis requires more than the ability to code, an ability to clearly understand the kind of data he/she could benefit from, if they can be obtained, as well as possible correlations between them; it is therefore important for the analysts to be aware of the categories of data he/she could obtain, how they can be correlated and how they can be obtained and what kind of real life situations they describe; this is determined rather by a clear understanding of how data fits in the organization and how the data implementation process should be deployed.

Given the various educational and professional backgrounds of the trainees, THESEUS should provide the general provisions for understanding the main concepts governing data analysis, and then the trainee would learn the skills on the job (as models of understanding differ across business areas, even in the same organizations, software has short life cycles and data analysis poses a challenge when trying to include it in a conventional academic syllabus, as it is constantly updating and changing the main algorithms).

Learning objectives

1. Understanding concepts related to data analysis and how data can be used within the organization.

2. becoming aware of the main legal constraints when it comes to (big) data analysis.

*as each intelligence organization has a legal department that approves or denies requests for data collection and storage in accordance with the legal provisions. The trainee must only know and understand the main constraints in his/her activity, so it can plan and allot resources accordingly.

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

3. Ability to work within a defined range of routinized queries and across multiple data sources (at least 2 different types of data[1]).

4. Report writing and presenting findings - strong verbal and written communication skills.

3. Selection of candidates

Eligibility criteria for students enrolled by MVNIA:

- bachelor degree;
- minimum 1 year professional experience and/or good knowledge of the intelligence community mission and vision;
- internal evaluation results;
- time management skills;
- has a sense of initiative, curiosity, flexibility and creativity.

Prerequisite knowledge:

- knows the main indicators for evaluating the credibility of a data source
- knows, understands and applies standards of quality when drafting a report
- introductory level in Excel – can manage small data sets and filter them based on given criteria
- critical thinking – knows and can identify some of the most frequent cognitive errors
- factsheet on how to do online searches
- factsheet with introductory concepts on statistics (e.g. sample, sampling methods etc.)

Assessment:

Given the project's scope, it is relevant to evaluate the ability of the candidate to work out new concepts and abstract ideas and observe underlying logic in a pattern. This is usually independent of education experience and cultural background and is relevant for jobs where problem-solving and initiative are important, or where candidates will need to deal with complex data, or perform non-routine tasks where initiative is required, as it is our case.

Thus, in the assessment stage, we propose to evaluate the following:

1. Abstract thinking (inductive reasoning/ diagrammatic reasoning)

[1] understood as biographical data, communication data, location data, social media data

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Proposed exercises:

- Watson Glaser critical thinking/ error checking test
- Provide the candidate with the legend for abstract representation of a process and request him to graphically represent an activity described in the given text

2. Availability to working with data

Proposed exercises:

- Provide the candidate with a short data report and ask him to explain that report to the evaluator
- Provide the candidate with 3 examples of data interpretation and ask him to identify and explain in which of them a thinking/interpretation error has been made.
- Provide the candidate with small data sets from various sources (relevant for a specific question) and request the candidate to identify informational needs, select relevant data and rank them based on credibility.

4. Topics for scenarios

* so as to ensure applicability across business areas and partners

- Pattern of life - Insider threat assessment (applicable in all intelligence organizations, extended control of data sources)
- Profiling - CT (applicable in all intelligence organizations, low control of data sources, existent model of understanding in professional literature, can be extrapolated to other business areas).
- Network analysis (1) – ego network, regardless of the dataset
- Network analysis (2) – flow of data/money within a network (applicable in CT, CI, organized crime networks for both intelligence services and LEAs, useful for the study of propaganda)
- Content analysis – propaganda analysis (relevant for both target groups)

Observation point:

Data analysis makes a distinction between **collection**, which is gathering data from various sources and in different formats and data **processing**, which is cleaning of data (defined in green). However, as the understanding of the two concepts across disciplines can cause

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

ambiguity and considering the project does not aim to teach the trainees how to process data (in the sense understood below), we could redefine the modules as you consider fit.

- **Data processing:** A series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use. Methods of processing must be rigorously documented to ensure the utility and integrity of the data.

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).