SECURING TOMORROW: NARCOVIEW PROJECT - INNOVATIONS IN DETECTING AND MANAGING DRUG DUMPING

FINAL THESIS

INTERNSHIP CYBERSECLAB | NARCOVIEW

B.R.J. BASTERS|476790
HBO ICT | HOGESCHOOL SAXION | ENSCHEDE
Ton de Bruyn | Wilco Engelsman

Names and addresses

Student intern:

Berit Basters
476790@student.saxion.nl

Company representatives:

MSc. Dimitar Rangelov
d.g.rangelov@saxion.nl

MSc. Tatjana Kuznecova
t.kuznecova@saxion.nl

Saxion coaches:

Ton de Bruyn
afm.debruyn@gmail.com

dr. Ir. Wilco Engelsman
w.engelsman@saxion.nl

Herman Voortman
h.w.voortman@saxion.nl

Preface

Before you is the bachelor's thesis "Securing Tomorrow: NarcoView Project - Innovations in Detecting and Managing Drug Dumping”. This thesis was written to meet the graduation requirements of the HBO-ICT Business course at the Saxion University of Applied Sciences in Enschede. I was researching and writing my thesis from September 2023 to February 2024.

I noticed during my studies that I like to step outside my comfort zone. That is why I have chosen a topic for this graduation assignment in an area wich I am unfamiliar with. I have worked with new innovations, AI models, different data sources and important stakeholders. I also gained more experience with materials I was already familiar with. That is why this thesis has taught me a lot on both a personal and professional level.

I would like to thank my supervisor, Dimitar Rangelov, Tatjana Kuznecova, Wilco Engelsman en Ton de Bruyn for the pleasant guidance and support during the process. I am glad that you challenged me to take my research to a higher level. This has maximized my learning opportunities, for which I am grateful. I would also like to thank all the team members from the NarcoView project for their contribution to the data collection for this research.

Finally, I would like to thank my family and friends for being there for me during my research process.

I hope you enjoy reading.

Berit Basters

Summary

NarcoView aims to improve finding narcotic waste dumpsites by utilizing a methodology that reduces reliance on current reporting methods. The project faces challenges in data- and usersecurity requirements. The main question is: How does the NarcoView project ensure secure handling of data used on the platform and customize it to meet the data security requirements of the stakeholders?

The main design goal is to create a comprehensive data management plan that addresses potential risks associated with data handling. The research questions focus on assessing data risks, understanding data origin, data intention, specific risks, stakeholder security requirements, and potential solutions to minimize these risks in the NarcoView platform.

The research design method by Piet Verschuren and Hans Doorewaard involves a systematic approach with several phases. The methodology emphasizes the iterative nature of research, allowing for adjustments and careful documentation of methods and decisions throughout the process.

This research delves into understanding and managing data risks, identifying five key theoretical perspectives: data profiling, threat modelling, data modelling, data security legislation, and best practices. Abedjan's data profiling methodology, emphasizing systematic data classification, is employed. Threat modelling combines OWASP and STRIDE for an example see 4.4 Results question 4. STRIDE categorizes threats comprehensively. While no laws directly apply to NarcoView's data collection, best practices like 2FA and ISO 27001 can ensure data security and personal screening can avoid that possible malicious individuals get access to sensitive data . These frameworks serve as a guide in effectively managing data risks.

The NarcoView platform uses a mix of internal and external data sources, primarily for monitoring, predictions, and crime detection. The platform's data is analyzed using OWASP and STRIDE methodologies, revealing various risks, related to API and external data sources. Stakeholders prioritize a secure platform that aligns with legal standards and effectively mitigates identified risks. The NarcoView project should integrate a mix of theoretical approaches and practical solutions, aligning with stakeholder expectations and requirements.

This research provides an understanding of data security in innovation projects. It focusses on regulatory frameworks like the EU's AI Act. The research has limitations, such as reliance on the research group's control over data security measures and limited stakeholder engagement. Future research should explore the integration of advanced AI and machine learning techniques for predictive analytics in data security and the human aspect of data security, considering user behaviour and organizational culture. The research provides a comprehensive understanding of data security challenges and opportunities within the NarcoView project.

This research recommends regularly updating and assessing data sources for risks, staying informed about laws, and implementing user access levels. These measures ensure data relevance, legal compliance, and enhance data security. Server security involves both online and physical aspects, with an emphasis on avoiding personal data storage. These recommendations are crucial for handling non-personal data.

Contents

Names and addresses ...... - 1 -

Preface ...... - 2 -

Summary ...... - 3 -

Contents ...... - 4 -

Abbreviations ...... - 6 -

List of figures ...... - 7 -

1. Introduction ...... - 8 -

1.1 Project rationale ...... - 8 -

1.2 Project subject ...... - 8 -

1.3 Project scope ...... - 8 -

1.4 Project relevance ...... - 8 -

1.5 Problem statement ...... - 8 -

1.6 Project objectives ...... - 9 -

1.8 Research questions ...... - 9 -

1.9 Research method ...... - 10 -

1.10 Reader ...... - 10 -

2. Theoretical framework ...... - 11 -

2.1 Data profiling ...... - 12 -

2.2 Threat modelling, OWASP and STRIDE ...... - 13 -

2.3 Data security law and legislation ...... - 14 -

2.4 Semistructured interview for expert opinions ...... - 16 -

2.5 Unified Modelling Language ...... - 17 -

3. Methodology ...... - 18 -

3.1 Conceptual Design ...... - 19 -

3.1.1 Subquestion 1 ...... - 19 -

3.1.2 Subquestion 2 ...... - 19 -

3.1.3 Subquestion 3 ...... - 20 -

3.1.4 Subquestion 4 ...... - 20 -

3.1.5 Subquestion 5 ...... - 21 -

3.1.6 Subquestion 6 ...... - 21 -

3.2 Choice of products ...... - 21 -

3.2.1 Theoretical framework ...... - 21 -

3.2.2 Data management plan ...... - 22 -

3.2.3 Data-profiling list ...... - 22 -

3.2.4 Risk analysis - 22 -
3.2.5 Advise report - 22 -
3.2.6 Visual threat and mitigations - 22 -

3.3 Research process - 23 -

3.4 Validity and reliability - 24 -

4. Results - 25 -

4.1 Results question 1 - 25 -
4.1.1 Data profiling - 25 -
4.1.2 Threat modelling - 25 -
4.1.3 Security legislation and best practices - 26 -

4.2 Results question 2 - 26 -

4.3 Results question 3 - 28 -

4.4 Results question 4 - 31 -

4.5 Results question 5 - 33 -

4.6 Results question 6 - 34 -

5. Conclusion - 36 -

5.1 Conclusion question 1 - 36 -

5.2 Conclusion question 2 - 36 -

5.3 Conclusion question 3 - 36 -

5.4 Conclusion question 4 - 36 -

5.5 Conclusion question 5 - 36 -

5.6 Conclusion question 6 - 36 -

5.7 Final conclusion - 37 -

6. Discussion - 38 -

7. Recommendations - 39 -

7.1 Data handling - 39 -

7.2 Server security - 39 -

7.3 Digital security - 39 -

8. Reflection - 40 -

Appendix - 42 -

Abbreviations

Abbreviation	Description
NVWA	Nederlandse voedsel- en Warenautoriteit
TCI	Technologies for Crime Investigation
DMP	Data management plan
2FA	Two-Factor Authentication
CDN	Content Delivery Network
HTTPS	Hypertext Transfer Protocol Secure
API	Application Programming Interface
AI	Artificial Intelligence
UML	Unified Modeling Language
OWASP	Open Web Application Security Project
STRIDE	Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege
DFD	Data Flow Diagram
GDPR	General Data Protection Regulation
NIST	National Institute of Standards and Technology
DOS	Denial of Service
BRP	Basisregistratie Personen
DPA	Data Protection Act

List of figures

Figures:

Figure 1 project products and competencies .... - 9 -
Figure 2 Research layout Verschuren and Doorewaard .... - 18 -
Figure 3 UML class diagram API .... - 30 -
Figure 4 Example STRIDE API threat .... - 31 -
Figure 5 Risk register .... - 32 -
Figure 6 Risk matrix .... - 32 -
Figure 7 Roadmap .... - 35 -
Figure 8 Visual Summary Thesis .... - 37 -

Tables:

Table 1 Literature subquestion links .... - 17 -
Table 2 Research question list .... - 23 -
Table 3 Current data list .... - 27 -
Table 4 Data sources included in API .... - 29 -
Table 5 Data sources excluded in the API .... - 29 -

1. Introduction

1.1 Project rationale

Drug dumping is becoming a larger problem than ever for Belgium and the Netherlands (Ree, 2023). This illegal activity not only poses significant environmental risks but also represents a substantial threat to public health and community safety. It is evident that traditional methods for detecting and addressing drug dumping have become inadequate. That is why the NarcoView project is working on a new way to detect drug dumping locations.

1.2 Project subject

The new methodology is designed to enable law enforcement to discover dump sites more effectively and quickly. This is accomplished through the creation of a risk map that identifies different areas with varying levels of risk. These identified risk areas can then be subjected to regular drone-based investigations. These drones, equipped with sensing technology and deep learning algorithms using satellite data, scan the designated locations and detect potential anomalies such as the presence of a jerry can or unusual discolorations in the terrain. The NarcoView platform generates reports about these dump sites, which can subsequently be utilized for future predictions and proactive enforcement efforts.

1.3 Project scope

The project is focused on researching and designing different ways to upscale the project to targeted users and put in place the security measures necessary to secure data handling within the project. In particular, the research will involve requirement analyses in a prototyping approach. In addition, the risk assessment needs to cater to these requirements. This project aims to determine and begin implementing the requirements necessary to bring the project to the next stage of prototyping. Finally, the data used in the platform needs to conform to the security standards of the stakeholders.

1.4 Project relevance

This research is relevant to the project, bringing it one step closer to its completion. It aids in the efficient identification of dumping sites, demonstrating the potential of robotics and machine learning for workplace optimization. Moreover, it plays a role in preventing food crop contamination by identifying high-risk dumping areas. In addition, this project offers valuable opportunities for students' professional growth, fostering expertise in project management, goal setting, and time management.

1.5 Problem statement

NarcoView problem statement: Create a methodology to make finding and searching for narcotic waste dumpsites more efficient and less dependent on current ways of reporting.

Research problem statement: In the project there is a lack of clear understanding of how to ensure the security of the data processed within the platform as well as the user security requirements necessary for the project's integration into the workforce.

User problem experience: Currently, narcotic waste is only noted when a passerby notices it and reports it to the appropriate authorities. The users of the future platform will not have to rely anymore on passersby to report the waste; they can inspect certain risk area and suspicious objects with the help of drones and self-learning algorithms.

1.6 Project objectives

Main design goal

The main goal is to develop a comprehensive data management plan that encompasses storage and data communication, while addressing the potential risks associated with data handling.

The final products are broken down into subproducts. In Figure 1 below it is portrayed how these subproducts interact with the different competencies.

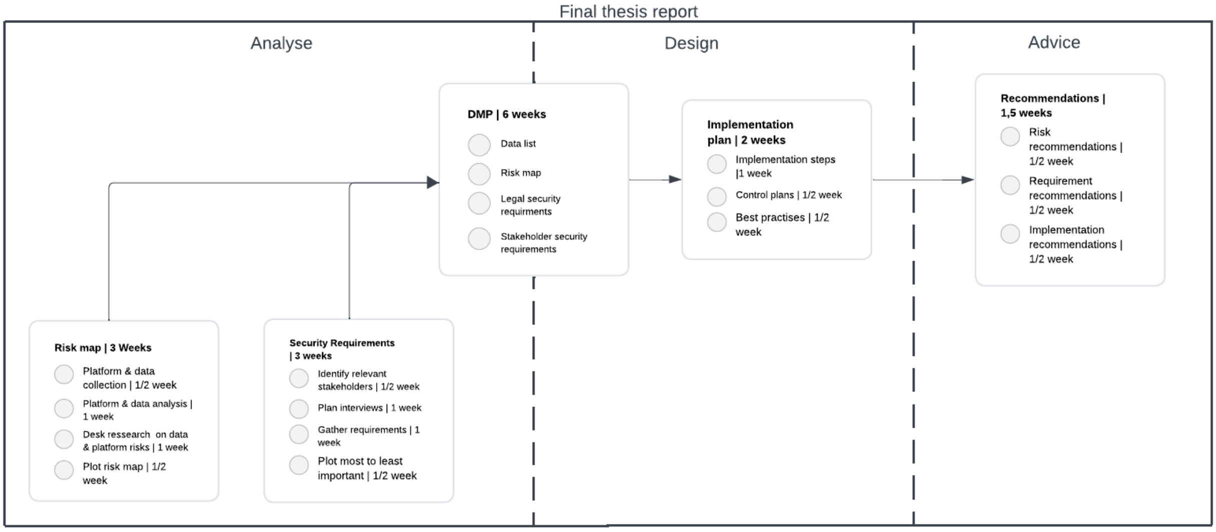

Final thesis report
Analyse
Design
Advice
DMP | 6 weeks
Data list
Risk map
Legal security requirments
Stakeholder security requirements
Implementation plan | 2 weeks
Implementation steps |1 week
Control plans | 1/2 week
Best practises | 1/2 week
Recommendations | 1,5 weeks
Risk recommendations | 1/2 week
Requirement recommendations | 1/2 week
Implementation recommendations | 1/2 week
Risk map | 3 Weeks
Platform & data collection | 1/2 week
Platform & data analysis | 1 week
Desk ressearch on data & platform risks | 1 week
Plot risk map | 1/2 week
Security Requirements | 3 weeks
Identify relevant stakeholders | 1/2 week
Plan interviews | 1 week
Gather requirements | 1 week
Plot most to least important | 1/2 week

Figure 1 project products and competencies

This figure shows that the research report is strategically structured and that the components of the report correspond with the competencies from the study. This figure is intended to provide a visual representation of the different products.

1.8 Research questions

Main question: How does the NarcoView project ensure secure handling of data used on the platform and customize it to meet the data security requirements of the stakeholders?

Subquestions:

1. Which theoretical perspectives and concepts can help understand how to assess data risks?
2. What is the origin of the data necessary in the use of the NarcoView platform?
3. What is the intention of the data that will be used in the NarcoView platform?
4. Which specific risks and threats are associated with the collection, storage, and processing of the data used within the NarcoView platform?
5. What are the security requirements and priorities of the stakeholders?
6. What are possible solutions to eliminate or minimize these risk that can be implemented in the NarcoView platform?

1.9 Research method

The methodology uses the research design method developed by Piet Verschuren and Hans Doorewaard. This method was chosen because it allows more room for iterative change, and the approach is revised in every phase. This method better aligns with the project and this thesis, as both the project and research focus on innovation, where changes play a significant role and pose a risk. There is no guarantee that the project will not undergo changes during the assignment period that could impact the research.

The method is set up in two groups of activities: conceptual design and technical research design (Verschuren & Doorewaard, z.d.). Designing a research project according to the methodology of Verschuren and Doorewaard involves a systematic approach to research with several phases. The orientation phase is where the problem is explored and the objectives are formulated. Following is the strategic phase; this phase creates the theoretical framework, refines the research questions, and describes the research methodology. Next is the work plan phase; in this phase the planning is carried out, and the data is collected. The analysis phase follows; in this phase the data is analysed. The penultimate phase involves reporting. Here, the results as well as the reflection and conclusion are presented. The last phase is the evaluation phase, including the critical reflection, dissemination of the results, and feedback to practice.

The Verschuren and Doorewaard methodology emphasizes the iterative nature of research, where one can revisit earlier phases to adjust as necessary. It is important to carefully document and be transparent about methods and decisions throughout the entire research process. Any changes of decisions can be found in the methodology chapter. See Methodology.

1.10 Reader

Chapter 2 serves as the theoretical framework, offering an in-depth explanation of the theories, concepts, and models that underpin the project. Chapter 3 delves into the methodologies and techniques applied throughout the project, providing insights into their practical use. Chapter 4 presents the research findings, organized according to the subquestions posed. Chapter 5 builds upon the preceding chapter by drawing conclusions derived from the subquestions. Chapter 6 initiates a discussion where expectations are analysed in relation to the results and new insights are introduced. Chapter 7 offers recommendations resulting from the research, providing practical guidance for future actions. Finally, Chapter 8 offers a reflective account of the project's progression and the lessons learned along the way.

2. Theoretical framework

This theoretical framework is aimed at laying a foundation for the research by exploring core concepts and methodologies such as data profiling, STRIDE (spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege) threat modelling, data security legislation and regulations, and expert opinions. It serves as a basis for the analysis and interpretation of the data.

Scientific literature:
The Open Web Application Security Project (OWASP), STRIDE methodology, and works by Abedjan et al. (2017), along with analyses from experts, form the foundation of this research. These sources were selected due to their comprehensive analyses and influence in data analysis and security.

Core concepts of the research:

- Data profiling: An approach to analyse the structure, quality, and relevance of data.
- Threat modelling with STRIDE: Focusing on identifying and mitigating security risks.
- Expert opinions: Insights and analyses from field experts that strengthen methodologies and conclusions.
- Data security legislation and regulation: Legal and best-practice frameworks for data security.

Definitions and justification:

- STRIDE: A methodology for security analysis, chosen for its comprehensive approach.
- Expert opinions: Perspectives from professionals that support the interpretation of data and research direction.

Interactions between core concepts:
Data profiling, STRIDE, and expert opinions are guided by data security legislation and regulation. The interaction between the different core concepts ensures a multifaceted approach to the research.

Central theories and models:

- OWASP and STRIDE threat modelling: Methodologies for systematically identifying and mitigating security risks.
- UML: Method for visualy representing information systems.

These theories and models were chosen due to their relevance and effectiveness, and they reinforce each other when applied together and supported by expert opinions.

2.1 Data profiling

This subsection outlines the process of data profiling as an effective method for compiling a list of data sources and understanding their intended usage within a platform. Data profiling involves systematically analysing data to gain insights into its characteristics, quality, and relevance (Abedjan et al., 2017).

Data source identification:

Create a comprehensive list of potential data sources, encompassing both internal and external options that could be integrated into the platform. This includes various sources like databases, APIs, third-party data providers, and user-generated content.

Data profiling:

Conduct data profiling for each identified data source to gather insights into the following:
Data structure, including schema, tables, and fields;
Data quality, assessing accuracy, completeness, consistency, and reliability;
Data volume, determining size and growth trends;
Data relevance, evaluating alignment with platform objectives.

Usage intent analysis:

Categorize usage intents such as reporting, analytics, machine learning, and real-time processing.

Data integration feasibility:

Assess the technical feasibility of integrating each data source based on profiling results. Identify potential data transformation or cleaning requirements.

Documentation and cataloguing:

Create a comprehensive catalogue or inventory of data sources, including their profiles and usage intents.

Validation and feedback:

Seek validation and feedback from relevant stakeholders to ensure that the compiled list accurately reflects reality. Iterate on the list based on feedback.

Communication and reporting:

Present the compiled list of data sources and their intended usage to stakeholders and decision-makers, offering a clear roadmap for data integration and usage within the platform.

This process is highly applicable in the research because it ensures that data sources are selected and integrated based on quality, relevance, and alignment with platform objectives. It is flexible, and regular updates and maintenance of the list adapt to changing data needs and stakeholder requirements. This aligns well with the evolving phase in which the NarcoView project currently finds itself. The process is used in answering subquestions 2 and 3.

2.2 Threat modelling, OWASP and STRIDE

Threat modelling using the OWASP methodology is a structured approach to identifying and mitigating security threats and vulnerabilities in web applications and services. OWASP provides a comprehensive set of resources and guidelines for threat modelling. The objective of threat modelling using OWASP is to systematically identify, analyse, and address security threats and vulnerabilities in web applications, following best practices and standards provided by OWASP (*Threat Modeling - OWASP Cheat Sheet Series*, z.d.).

Create a data flow diagram:

Develop a data flow diagram that illustrates how data flows through the application. Identify data sources, data sinks, and the paths data takes within the system.

Identify threat agents:

Identify potential threat agents, such as attackers, insiders, or malicious users, who may exploit vulnerabilities or compromise assets.

STRIDE analysis:

Apply the STRIDE model to systematically analyse and identify threats and vulnerabilities associated with each asset.

Mitigation strategies:

Develop mitigation strategies for each identified threat or vulnerability.

Review and validation:

Review the threat model with security experts, developers, and relevant stakeholders to ensure its accuracy and completeness.

Integration into software development lifecycle:

Integrate the threat model into the software development lifecycle. Ensure that security considerations are addressed in the design, coding, and testing phases.

Ongoing monitoring:

Continuously monitor the application for new threats and vulnerabilities, and update the threat model as needed.

Education and training:

Train development and security teams on OWASP best practices and threat modelling techniques to foster a security-aware culture.

OWASP is highly applicable in the research because threat modelling using OWASP provides a structured approach to identifying and mitigating security risks in web applications. This method proactively addresses security concerns, reduces the likelihood of security breaches, and enhances the overall security posture of the platform. This approach aligns closely with the research goal and is applied to subquestions 4, 5, and 6 (*OWASP Top Ten | OWASP Foundation*, z.d.).

2.3 Data security law and legislation

The objective of this desk study is to comprehensively review and analyse legislation and best practices related to data security. This methodology provides a structured approach to gathering, evaluating, and synthesizing information from existing sources in the context of data security.

Define research objectives:

Clearly define the research objectives, focusing on the data security aspects to be studied such as legal requirements, industry standards, and best practices.

Scope definition:

Determine the scope of the study, specifying the geographic region (e.g., national, international), types of data (e.g., personal, sensitive), and relevant time frames (e.g., recent legislation).

Information sources:

Identify and compile a list of information sources for data security research.

- Data protection laws and regulations (e.g., GDPR).
- Government cybersecurity guidelines and frameworks.
- Industry-specific data security standards (e.g., ISO 27001).
- Academic articles, research papers, and case studies.
- Reports and publications from cybersecurity organizations.
- Best-practice guidelines and frameworks (e.g., NIST cybersecurity framework).

Search strategy:

Develop a systematic search strategy tailored to the research objectives. Utilize academic databases, online libraries, government websites, and specialized search engines to gather relevant documents and sources.

Data collection:

Collect data security-related documents and sources systematically. Organize the gathered materials in a logical and accessible manner, categorizing them based on relevance.

Data validation:

Verify the accuracy and relevance of the collected data by cross-referencing multiple sources and validating information against authoritative documents and legal texts.

Document review:

Review legislative documents and data security standards thoroughly, focusing on key aspects such as data protection principles, requirements, security controls, and enforcement mechanisms.

Best-practices assessment:

Analyse best-practice guidelines and frameworks related to data security. Identify common recommendations, methodologies, and frameworks for securing data.

Comparative analysis:

Conduct a comparative analysis of data security legislation and best practices. Identify areas of alignment, gaps, inconsistencies, and potential areas for improvement.

Recommendations and conclusions:

Formulate recommendations based on the analysis of data security legislation and best practices. Offer insights into potential areas for improvement, compliance strategies, and alignment with best practices.

Documentation and reporting:

Document the entire desk study process, including the sources used, methodologies applied, and analysis results. Prepare a comprehensive report with clear findings, conclusions, and actionable recommendations for enhancing data security.

Peer review and validation:

Seek peer review and validation of the research methodology and findings from experts or stakeholders in the field of data security to enhance credibility and accuracy.

Regular updates:

Keep the study updated by periodically reviewing new legislation, regulations, and emerging best practices in the field of data security.

This research employs the above methods to systematically examine legislation and best practices in the field of data security, gain insights into legal requirements and industry standards, identify areas for improvement, and contribute to informed decision-making in the realm of data security. This contributes to answering subquestions 4 and 5.

2.4 Semistructured interview for expert opinions

The objective of semistructured interviewing is to gather qualitative data through guided yet flexible conversations with participants. This methodology aims to explore complex topics, gain in-depth insights, and generate rich narratives. According to (Adams, 2015)

Participant selection:

Identify and select participants who have relevant knowledge, experience, or insights related to the research objective. Ensure diversity in participant backgrounds if applicable.

Informed consent:

Obtain informed consent from participants, explaining the purpose of the interview, how the information will be used, and assuring confidentiality.

Interview guide preparation:

Develop a semistructured interview guide that includes a list of open-ended questions and prompts related to the research objectives. Ensure flexibility for in-depth exploration.

Pilot testing:

Pilot test the interview guide with a small group of participants to refine questions, assess clarity, and ensure that the guide elicits the desired insights.

Open-ended questions:

Commence the interview with a broad, open-ended question to encourage participants to share their thoughts, experiences, and stories freely.

Probing and clarification:

When participants provide rich insights, ask probing questions to clarify and expand on their responses.

Data validation:

Employ techniques like member checking by sharing interview findings with participants to validate the accuracy and interpretation of their responses.

Recording:

Record the interview with the participant's consent, whether through audio, video, or detailed notetaking. This ensures accuracy and allows for later analysis.

Debriefing:

After the interview, thank the participant and provide an opportunity for them to ask questions or share additional insights.

Reporting and documentation:

Document the semistructured interview process thoroughly, including the interview guide, pilot testing results, transcripts, and analysis notes. Elaborate the documentation by reporting the findings, including quotes or anecdotes that illustrate key points. Present the results in a clear and organized manner, drawing conclusions related to the research objective.

2.5 Unified Modelling Language

Purpose: Unified modelling language (UML) can be used for modelling software architecture and design, but it also has features such as activity and sequence diagrams that can be useful for modelling processes.
Use: UML activity diagrams can be used to model the flow of a process, while sequence diagrams are useful for showing the interaction between different objects.
Benefits: UML is versatile and can be used for both process and software design.
Drawbacks: UML can be complex and is more focused on software development than on business processes.
Use in project: UML is used when dataflows need to be described.
(*UML Class Diagram Tutorial*, z.d.)

Conclusion:
This research uses semistructured interview because of the flexibility they give to an interview. The more open context of the interview leads to qualitative data and contributes to a nuanced exploration of the research objective. This contributes to answering subquestion 5.

Application and integration of theories and models
The collaboration of data profiling and STRIDE threat modelling, strengthened with expert opinions, provides a comprehensive approach for analysing the security aspects of data. This integration ensures a robust security system and thorough data analysis.

Conclusion of the theoretical framework and outlook
Insights from literature and experts provide a detailed understanding of how data profiling, STRIDE threat modelling, and data security legislation and regulations collectively address the subquestions. Below is a table to show how the literature and the subquestions interact Table 1.

Literature	**Sub question**
Data profiling	2 3
Threat modelling	4 5 6
Semistructured interview	5
Data security law and legislations	4 5
Unified modelling language	3

Table 1 Literature subquestion links

This table is a visual summary of the literature used for the corresponding sub-question. It is intended to help the reader understand how the literature and the sub-questions are related.

3. Methodology

The research methodology employed in this study is based on the research design method formulated by Piet Verschuren and Hans Doorewaard. This specific method was selected due to its flexibility, enabling iterative adjustments at each stage of the research process. This choice aligns more effectively with both the project and the thesis, as they both center around innovation, where change is a pivotal element carrying inherent risks. Given the dynamic nature of the project, it is essential to acknowledge that there may be unforeseen changes during the assignment period that could potentially influence the research outcomes.

This method is often associated with designing and conducting research in a systematic and structured manner. It was chosen for this study because it can be effectively applied to innovation research such as this one. This method divides the research process into two main parts: conceptual design and technical design Figure 2. Below, each component is explained in terms of how it would be applied in the research.

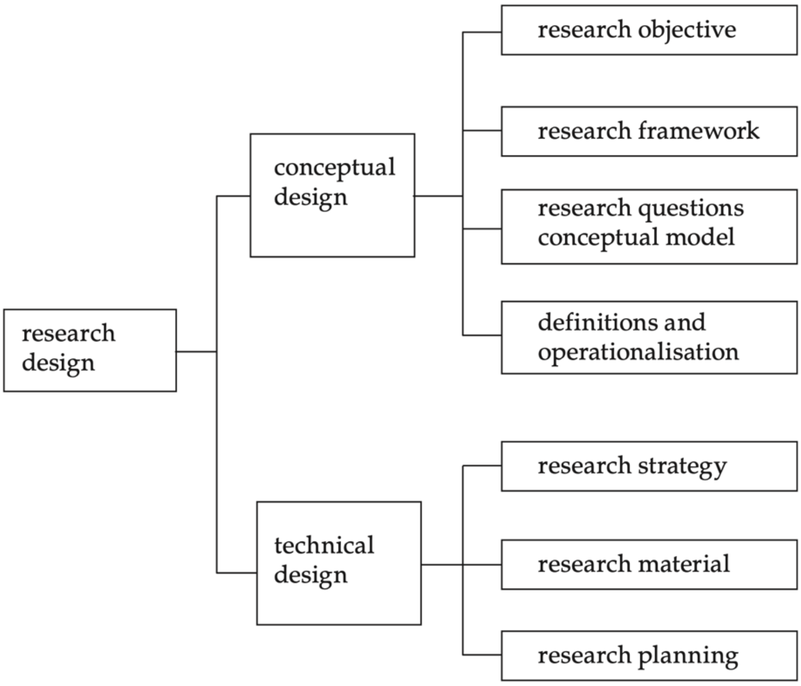


Figure 2 Research layout Verschuren and Doorewaard

This figure provides a presentation of what the research design looks like and where the various components belong. This model has also been adopted in this research and gives the reader the necessary foundation to understand the structure of the research.

3.1 Conceptual Design

The conceptual design provides a clear theoretical foundation for the research and clarifies the aim of the study. The primary objective of the research, along with the main and subsidiary questions, are discussed below. The subsidiary questions are explained along with their definitions and operationalization, referring to the framework used.

Research objective: To assess the effectiveness of NarcoView project's data security measures and their alignment with stakeholder requirements, aiming to identify and recommend improvements for data protection.

Main question: How does the NarcoView project ensure secure handling of the data used on the platform and customize it to meet the data security requirements of the stakeholders?

3.1.1 Subquestion 1

Which theoretical perspectives and concepts can help understand how to assess data risks?
This is an explanatory question. The aim of it is to facilitate a comprehensive understanding of the theoretical foundations that can guide the assessment of data risks, ultimately contributing to more effective risk management strategies and decision-making. It is focused on gaining a deeper understanding and insight into data risk assessments.

Methodology: Desk research to explore which models and techniques can be utilized during the study to answer the research questions. This will be achieved by examining various methods to address the different sub-questions. Subsequently, the most suitable methods for each specific aspect of the research are determined. The selected methods are then integrated and elaborated within a theoretical framework, which allows for the development of a comprehensive understanding and the answering of the research questions.

Products: Theoretical framework encompassing all models and techniques used in the research to answer the subquestions.

3.1.2 Subquestion 2

What is the origin of the data necessary in the use of the NarcoView platform?
This is an exploratory question. It aims to determine the sources or origins of the data that is required for the functioning of the NarcoView platform. This question is intended to gain an understanding of where the data comes from, whether it is collected, generated, or obtained from external sources, which can be crucial for data management, validation, and ensuring the platform's data quality and accuracy.

Methodology: Interviews with various team members to identify which data sources are still being used in the development of the platform, which are no longer applicable, and the origins of these data sources. This will be achieved by identifying the team members handling the data. Subsequently, arrangements are made for interviews, and an interview structure is created and shared with the team members. Interviews are then conducted, followed by the collection of data. Finally, a new data list is generated.

Products: Data source list with all updated data sources, their origins, and how they are utilized.

3.1.3 Subquestion 3

What is the intention of the data that will be used in the NarcoView platform?
This is an exploratory question. It aims to explore and explain the purpose and intended use of the data within the context of the NarcoView platform. This question seeks to gain a better understanding of why the data is being utilized, ensuring that the data serves its intended purpose effectively.

Methodology: Data profiling, combined with interviews, to determine how each data source should be used and who is authorized to use it. This will be achieved by seeking someone within the team that is responsible for the user interaction in the application. Subsequently, arrangements are made for an interview, during which an interview structure is created and shared, focusing on intended users and usage discussions. Interviews are then conducted, followed by the compilation of a data usage list using the information gathered from the interviews.

Products: Data usage list detailing the purpose of each data source.

3.1.4 Subquestion 4

Which specific risks and threats are associated with the collection, storage, and processing of the data used in the NarcoView platform?
This is an evaluative question. It aims to evaluate and assess the potential risks and threats related to the data handling processes in the NarcoView platform. This question is concerned with identifying and understanding the negative aspects and vulnerabilities in data management to make informed decisions and take preventive measures to address these risks.

Methodology: Enhance threat modelling by integrating OWASP and STRIDE methodologies to provide a detailed analysis of potential threats and their corresponding mitigations. This approach is systematically applied to each data type collected in the preceding sub-questions. Mitigations for the identified threats are then gathered through a comprehensive review of prior literature. The risks are compiled into a threat list, and their severity is assessed. Finally, the identified threats are presented in a threat matrix.

Products: A comprehensive threat list and a detailed threat matrix to systematically display the identified threats and their respective mitigation strategies.

3.1.5 Subquestion 5

What are the security requirements and priorities of the stakeholders?
This is an exploratory question. It aims to explore and explain the specific security needs and priorities of the stakeholders in the project. This question seeks to understand what the stakeholders consider essential in terms of security, which can be crucial for tailoring security measures and solutions to meet their requirements effectively.

Methodology: Expert opinions to gather insights on identified threats and mitigations, ensuring alignment with stakeholder requirements. This is done by identifying experts who possess knowledge about the data types and the associated risks. Interviews are then scheduled with these experts. During the interviews, inquiries are made regarding the risks and mitigations that they deem most suitable for the data in question. Subsequently, the responses obtained from the experts are compared to the findings from the initial research. The insights gathered from the interviews are used to enhance and refine the answers. Finally, a list of mitigations and implementations for the project is formulated.

Products: List of mitigations and a visual representation that clearly illustrates the threats, their corresponding mitigations, and implementation strategies within the project.

3.1.6 Subquestion 6

What are possible solutions to eliminate or minimize these risk that can be implemented in the NarcoView platform?
This is an advisory question. It aims to design and propose solutions to address the identified risks within the NarcoView platform. This question is focused on finding practical approaches to mitigate or eliminate risks and improve the platform's security, which falls under the problem-solving and advisory category.

Methodology: Continuous improvement to implement a cycle of ongoing evaluation and enhancement to identify and mitigate risks, ensuring the NarcoView platform's security evolves with emerging threats. This will be accomplished by compiling an advisory report based on the previously obtained answers and insights. Utilizing the identified mitigations, a roadmap is developed to outline the time and cost estimates associated with their implementation. Subsequently, the advice and roadmap are discussed with company supervisors to validate their feasibility.

Products: Advisory report outlining identified risks, recommended solutions, and practical approaches to strengthen the platform's security, serving as a guide for implementation and future reference.

3.2 Choice of products

In the next section, the choices for the different products will be discussed in detail, specifically addressing how each product will contribute to the improvement and success of the project.

3.2.1 Theoretical framework

The theoretical framework in this study is considered a unique product as the research primarily consists of new innovations, and there are no existing theories that seamlessly align with this case. The theoretical framework is formed by methods and techniques that correspond with a specific part of the research or have significant relevance to an aspect of the study. The theoretical framework serves as a guideline for the course of the research.

3.2.2 Data management plan

The data management plan provides a comprehensive description of all the specific details of the research. This plan details the origin of the data, how it is used and by whom, as well as the potential risks associated with this data and the methodology applied. Furthermore, the data management plan enables compiling data-profiling lists and conducting risk analyses.

3.2.3 Data-profiling list

The data-profiling list is used to gain a clear understanding of the current data, data flows, and how they are handled. It also links users to the data and preassigns future user levels to the data sources. However, the data list should be updated first before the data can be profiled.

3.2.4 Risk analysis

The risk analysis is based on two theories: OWASP and STRIDE. The combination of these two ensures that a detailed risk analysis can be created that covers all aspects of the research. From this analysis, two subproducts emerge: a risk register and a risk matrix. The risk register acts as a summary of the risk analysis, while the risk matrix displays it in a visual form. This allows for a quick determination of which risks should be prioritized for mitigation.

3.2.5 Advise report

The advisory report is a concise document that highlights all risks, including mitigation strategies for these risks. It places extra emphasis on the most significant threats to the project and provides a detailed description of how these mitigations can be implemented, including an estimated timeframe for each approach. The report also offers an estimation of the potential costs associated with improving the situation around data protection and handling. This report is designed to ensure that all stakeholders of the project understand the severity of the threats and are motivated to take the necessary actions.

3.2.6 Visual threat and mitigations

A visual representation will be created to present the threats and associated mitigation strategies in a simplified manner. This visualization is intended to convey the severity of the situation and raise awareness of the need for change. It will also provide insight into the benefits of the future situation.

No.	***Research question***	***Data collection***	***Data analysis***	***Professional products***
1.	Which theoretical perspectives and concepts can help understand how to assess data risks?	Literature collection	Literature analysis	Theoretical framework
2.	What is the origin of the data necessary in using the NarcoView platform?	Data source collection	Data source collection	Data source list

3.	What is the intention of the data used in the NarcoView platform?	Semistructured interviews with product owner	Data profiling	Data usage list
4.	Which specific risks and threats are associated with the collection, storage, and processing of the data used in the NarcoView platform?	Research risk studies, Semistructured interviews with security expert, Security webinars	Threat modelling, Data flow diagrams	Risk map
5.	What are the security requirements and priorities of the stakeholders?	Semistructured stakeholder interviews	Use Moscow to scale requirements	Security requirement list
6.	What are possible solutions to eliminate or minimize these risk that can be implemented into the NarcoView platform?	Best-practices research, Semi-structured interviews with security expert, Security webinars	Define SMART recommendations	Recommendations and implementation plan

Table 2 Research question list

This table lists the sub-questions and the corresponding data collection and analysis, along with the sub-products that result from this. This gives the reader an idea of what the research will yield and how these products have been developed.

3.3 Research process

The research process comprises various stages aimed at gaining insights into project aspects. Initially, a literature review is conducted to identify relevant pieces and theoretical frameworks applicable to the research, such as data collection, classification, risks, and potential mitigation strategies.

After establishing the theoretical foundation, actual data collection is performed. All relevant data sources used within the project are identified and categorized into different types, such as application program interfaces (APIs), open external sources, and closed external sources. The data flow is also mapped to provide a comprehensive overview. A crucial step in the research process is linking these data sources to their associated risks. This leads to the development of a detailed risk register and matrix for a better understanding of the risk landscape. After identifying the risks, potential mitigation strategies to reduce the potential negative impact are explored.

The findings are then evaluated in collaboration with experts, collecting and incorporating feedback. Based on this input, final adjustments are made to the risk analysis. Subsequently, a security requirement list is compiled directly from the risk analysis. If necessary, the platform's development process and data flow are redesigned, depending on identified risks and potential areas for improvement. Finally, an advisory report is prepared, consolidating all improvement points and adjustments.

There is minimal stakeholder involvement because of the nature of the project. The project is focussed around innovation and thus is making a product out of opportunitie. The stakeholders will be inquired after the research has been done. This will be interviews with experts about the outcome of the research. They were asked about their opinion on the data sources and the risks that belong to these sources we also discussed potential mitigations. These opinions are later compared to the outcome of the risk analysis and mitigations from this study.

3.4 Validity and reliability

The validity of this research is ensured as the applied theories address similar cases to those examined in the research. The focus within these theoretical frameworks is on the risks associated with open web pages and APIs, as well as various mitigation options relevant to these risks. The reliability of the research is guaranteed by applying the Verschuren and Doorewaard method. This methodology closely aligns with the project and emphasizes an iterative research process, providing flexibility to adjust earlier stages if necessary.

4. Results

This research is composed of five subquestions, each contributing to the ultimate advice and outcome. This chapter presents the results per subquestion and highlights the most important results.

4.1 Results question 1

Question: Which theoretical perspectives and concepts can help understand how to assess data risks?

To answer this question, literature research was conducted on theories involving a similar case as in this study according to Subquestion 1. From this, various theories emerged that can be used during the research. See Theoretical framework. First, it was established what theories were needed for data profiling, threat modelling, data modelling, data security legislation, and best practices. These theories contribute to understanding and managing data risks in various contexts. They are briefly explained below along with how they are used in the project.

4.1.1 Data profiling

For the data profiling component, the approach outlined by Abedjan in 2017 was implemented. (Abedjan et al., 2017) This methodology is designed to profile large-scale datasets, preparing them for applications in data mining and machine learning processes. This is relevant for the Narcoview project because they are using data mining and machine learning to try and predict furture drug dumping sites. The framework is pertinent to the objectives of the NarcoView project. The research predominantly adhered to Abedjan's practices for data collection and profiling, underscoring the importance of systematically classifying data and navigating the inherent challenges of computational demands and the management and interpretation of profiling outputs. See 2.1 Data profiling. These considerations are integrated into the risk analysis by reevaluating the data sources, ensuring a thorough examination of the diverse data types employed within the project's scope.

4.1.2 Threat modelling

After careful consideration it was decided that for threat modelling, a combination of OWASP and STRIDE would be used. See 2.2 Threat modelling, OWASP and STRIDE. OWASP is an international nonprofit organization dedicated to improving the security of software. It provides free, openly available resources for developers, security professionals, and organizations to understand, develop, and maintain software that is secure from potential threats and attacks. (*Threat Modeling - OWASP Cheat Sheet Series*, z.d.) OWASP provides a good framework for threat modelling; it includes not only the threat modelling but also a better understanding of the data that is used and collected in the project. However, for this research more detail was needed in the threat modelling. This is where STRIDE comes in. (jegeib, 2023). The STRIDE method is a threat modelling framework used to identify and categorize security threats in software applications. Developed by Microsoft, STRIDE stands for six types of security threats: spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege. These six aspects also correspond to the risks that OWASP warns about. These techniques together ensure a detailed and comprehensive threat analysis for the project.

4.1.3 Security legislation and best practices

Finally, it was necessary to investigate whether there are laws and regulations attached to NarcoView's data collection method. A thorough search was conducted, but to date there are no laws or regulations surrounding this form of data collection. The reason for this is that NarcoView does not work with sensitive personal data. The existing laws regarding data collection address how to handle personal data and what requirements a company must meet to process this data. Examples include the European Union's (EU's) Artificial Intelligence (AI) Act(*EU Artificial Intelligence Act | Up-to-Date Developments and Analyses of the EU AI Act*, z.d.), General Data Protection Regulation (GDPR) (*General Data Protection Regulation (GDPR) – Official Legal Text*, z.d.), and the DPA(Koninkrijksrelaties, 2017). None of these are applicable to the application that NarcoView is developing. However, there are best practices and quality certificates that the project can employ to ensure that the data they use is still treated securely. These include multifactor authentication (MFA), ISO 27001, and physical measures that NarcoView can take to keep their processes safe. OWASP also provides tools that can be used in an advisory capacity.

These three frameworks and methodologies—data profiling, threat modelling, and security legislation and best practices—form the foundation of this research study. They contribute to the research by ensuring that the advice and conclusions drawn are well-supported by a thorough understanding of data risk assessment and management.

4.2 Results question 2

Question: What is the origin of the data necessary in the use of the NarcoView platform?

To answer this question the method of Subquestion 2 was used. Discussions were held with various team members of the project to understand their work and the data utilized see appendix E and F. These members were selected because of their involvement and overall view of the data used within the project. They work daily with the data and are well aware of the changes made in the sources. These conversations yielded insights into the diverse sources of data. Additionally, a file detailing all the different sources was shared, although it was initially outdated. Further interviews with the team members led to the identification of changes in the sources used for their work, resulting in the creation of the new version of the data source list, as seen in table 1.

In summary, the data necessary for the NarcoView platform is derived from a combination of external and internal sources. External sources encompass data from entities outside the stakeholder parties. In contrast, internal sources are contributions from stakeholder parties, including the Dutch and Belgian police and the Netherlands Food and Consumer Product Safety Authority (NVWA). The internal data specifically includes information on former dumping sites and areas previously affected by pollution. For further elaboration see appendix A.

No	Data type	Data source
1	Risk map	Manually made
2	Road network	https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-
3	Railway network	https://www.pdok.nl/introductie/-/article/spoorwegen
4	Neighbourhoods	https://www.pdok.nl/downloads/-/article/cbs-wijken-en-buurten
5	Provinces	https://www.pdok.nl/introductie/-/article/bestuurlijke-gebieden
6	Land use	https://service.pdok.nl/cbs/bestandbodemgebruik/2017/wfs/v1_0?request=getCapabilities&service=WFS
7	Waterways	https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-
8	Key register crop parcels (BRP)	https://www.pdok.nl/introductie/-/article/basisregistratie-gewaspercelen-brp-
9	TOP10 NL (Basic Topography Registry)	https://www.pdok.nl/introductie/-/article/basisregistratie-topografie-brt-topnl
10	Soil moisture	https://nsidc.org/data/smap/data
11	Sentinel 2 MSI imagery	https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/APIHubDescription?TWIKISID=2c4ba407f55151742a94fbd577a81572
12	NDVI layer	Manually made
13	Satellite or aerial imagery for object detection	https://opendata.beeldmateriaal.nl/pages/bekijken-1
14	Dump site locations	Police, NL, and BE
15	Polluted fields (ground truth)	NVWA

Table 3 New data list

This table contains the new data list that was developed during the research. This data list provides a complete overview of the data sources used in the project.

4.3 Results question 3

Question: What is the intention of the data that will be used in the NarcoView platform?
To answer this question, the data profiling theory was used with the methodology of Subquestion 3. First, I identified the different data sources to make up a list. After drafting this list, each data source was profiled. The data list comprises information from diverse sources, categorized into two main types: data sources included in the API and data sources not included in the API. The sources excluded from the API are used in making the risk map; these sources are not directly connected by the API but are rather used by the team members to make these risk maps. Data sources included in the API are directly linked to the API and are used to obtain specific information by putting in a request via the API. For the final result, see Table 4 and Table 5.

There is a difference between internal data and external data. Internal data emanates from project stakeholders such as the NVWA and the police, who actively participate in the project and oversee the management of these internal data sources. On the other hand, external data is sourced from locations like the Public Services on the Map (PDOK) platform. While profiling the data, different intended users were assigned to the different data sources in order to split up the access levels later in the project. See appendix G for the interview in which these results are dicussed.

There are different users for the platform analysts and inspectors. Analyst users will mostly be users from the NVWA. Inspector users will mostly be from the police. Users will be assigned to an account by the administrator. The administrator will be situated within the police. Both users will have the same rights but different intentions for using the data. The users also do not have direct access to all the data; they need to request the data that they need.

The intended usage for each of the data sources is also assigned. Analysts will use the data for monitoring and further predictions; this will mostly be the users from the NVWA and the NarcoView project team. Inspectors will use the data for detecting crime and current dumping sites; this information will be used by the police.

No.	Data type	Data source	Intended users	Intended use
1	Risk map	Manually made	Analysts, Inspectors	Reporting
2	Road network	https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-	Analysts, Inspectors	Analytics
3	Railway network	https://www.pdok.nl/introductie/-/article/spoorwegen	Analysts, Inspectors	Analytics
4	Neighbourhoods	https://www.pdok.nl/downloads/-/article/cbs-wijken-en-buurten	Analysts, Inspectors	Analytics
5	Provinces	https://www.pdok.nl/introductie/-/article/bestuurlijke-gebieden	Analysts, Inspectors	Analytics
6	Land use	https://service.pdok.nl/cbs/bestandbodemgebruik/2017/wfs/v1_0?request=getCapabilities&service=WFS	Analysts, Inspectors	Reporting
7	Waterways	https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-	Analysts, Inspectors	Analytics
8	Key register crop parcels (BRP)	https://www.pdok.nl/introductie/-/article/basisregistratie-gewaspercelen-brp-	Analysts, Inspectors	Reporting
9	TOP10 NL (Basic Topography Registry)	https://www.pdok.nl/introductie/-/article/basisregistratie-topografie-brt-topnl	Analysts, Inspectors	Analytics
10	Soil moisture	https://nsidc.org/data/smap/data	Analysts, Inspectors	Analytics

Table 4 Data sources included in API

No.	Data type	Data source	Intended users	Intended use
11	Sentinel 2 MSI imagery	https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/APIHubDescription?TWIKISID=2c4ba407f55151742a94fbd577a81572	Analysts, Inspectors	Machine learning
12	NDVI layer	Manually made	Analysts, Inspectors	Analytics
13	Satellite or aerial imagery for object detection	https://opendata.beeldmateriaal.nl/pages/bekijken-1	Analysts, Inspectors	Machine learning
14	Dumpsite locations	Police, NL, and BE	Analysts, Inspectors	Machine learning
15	Polluted fields (ground truth)	NVWA	Analysts, Inspectors	Analytics

Table 5 Data sources excluded in the API

These tables provide an overview of the data sources and whether or not they have been included in the API. Also, the sources are linked to the target users and intended use. This gives an understanding of the purpose of the sources.

The primary objective behind collecting and integrating the data is to leverage the combined dataset for training an AI model. This model is designed to identify potential risk areas where illegal dumping might occur, including those in remote and challenging-to-access locations, utilizing specialized imaging techniques. The platform is equipped to generate comprehensive reports that hold significant utility in forthcoming investigations. This functionality is predominantly intended for use by law enforcement agencies and prosecutors. The reports can be instrumental in analysing the impact of a dumping site on surrounding fields, allowing the NVWA analysts to assess the potential contamination of fields earmarked for production. Notably, the platform extends equal rights to both user groups, the police and prosecutors, although the results obtained can be applied divergently based on the unique needs and inquiries directed to the API by each user.

The unified modelling language (UML) class diagram below Figure 3 illustrates the process of retrieving land data. The API directs the inquiry to the extractor class, which then requests the data from PDOK. After PDOK responds, OWSLib is used to extract the specific data, and the API sends it back to the user. For further details see appendix A and B.

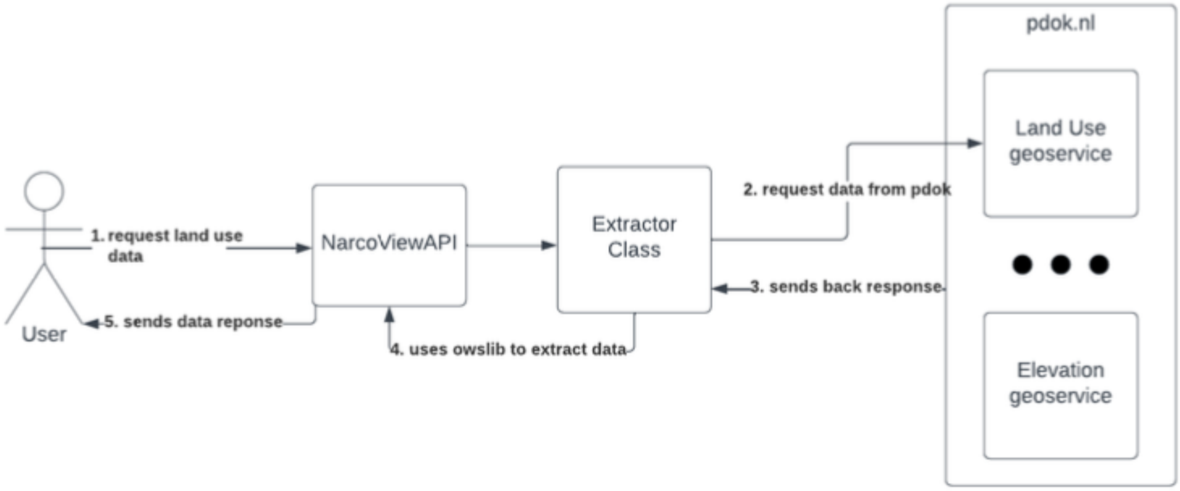


Figure 3 UML class diagram API

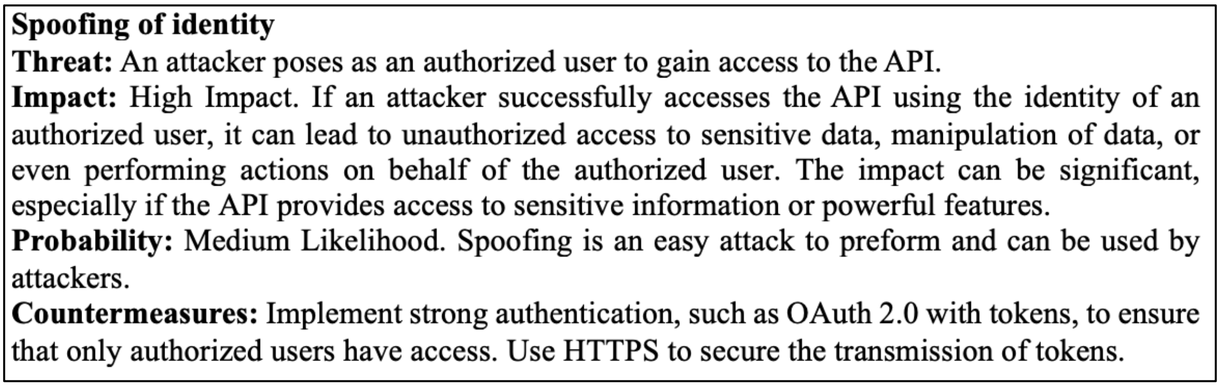
This figure illustrates how a request progresses through the entire process, highlighting the various components involved and their corresponding actions. The aim is to give a better understanding of the process through this visual representation.

4.4 Results question 4

Question: Which specific risks and threats are associated with the collection, storage, and processing of the data used in the NarcoView platform?

For this question methodology Subquestion 4 was used. Various data sources pose distinct threats, and an early-stage risk analysis guided by relevant literature see appendix X was conducted for each data type. Despite identifying numerous risks associated with the data, it is crucial to note that not all risks are within the control of the research group. This limitation arises due to the presence of sensitive information managed directly by stakeholders. The research group can exert influence over the security of the data incorporated into the project but lacks control over data stored by external stakeholders. Currently, data is transferred from stakeholders to the research group in small batches for program development. When the final product is launched, it is up to the users to implement it and connect it to their own databases to streamline data transfers and enhance security measures.

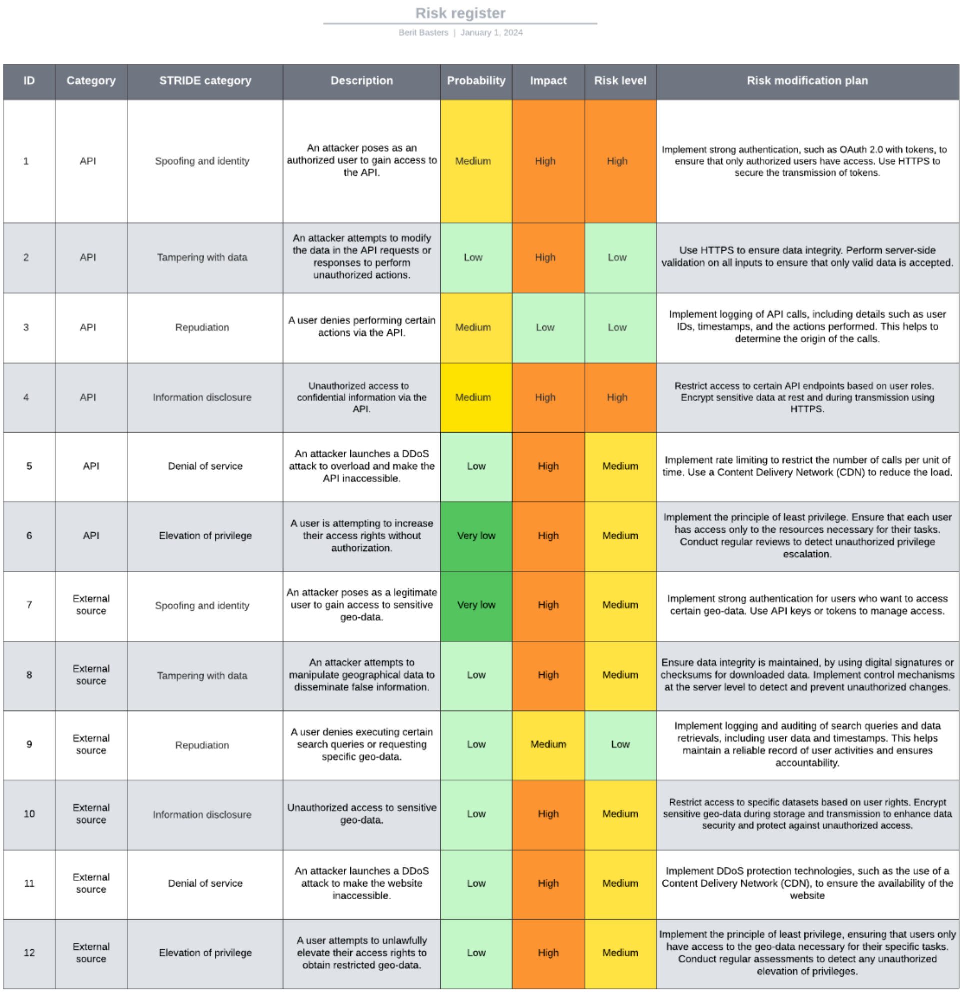
For the risk assessment, a mix of OWASP and STRIDE was used for the API and external sources. For the full assessment, see appendix B. To analyse the risks, a distinction between API threats and threats using external data sources was maintained. For these two categories, each STRIDE step is used below as an example Figure 4.



Spoofing of identity
Threat: An attacker poses as an authorized user to gain access to the API.
Impact: High Impact. If an attacker successfully accesses the API using the identity of an authorized user, it can lead to unauthorized access to sensitive data, manipulation of data, or even performing actions on behalf of the authorized user. The impact can be significant, especially if the API provides access to sensitive information or powerful features.
Probability: Medium Likelihood. Spoofing is an easy attack to preform and can be used by attackers.
Countermeasures: Implement strong authentication, such as OAuth 2.0 with tokens, to ensure that only authorized users have access. Use HTTPS to secure the transmission of tokens.

Figure 4 Example STRIDE API threat

In this example it can be seen that the S in STRIDE—“spoofing of identity”—was used. Through talking to team members and research, the possible threat was decided. The impact describes the impact that the threat would have on the application. The probability is the likelihood of a threat really happening, and the countermeasures describe possible countermeasures to eliminate or minimize the impact and probability. All the threats and their impact and probability are discussed with professionals and team members. A risk register was created to summarize the different risks Figure 5.



Risk register

Berit Basters | January 1, 2024

ID	Category	STRIDE category	Description	Probability	Impact	Risk level	Risk modification plan
1	API	Spoofing and identity	An attacker poses as an authorized user to gain access to the API.	Medium	High	High	Implement strong authentication, such as OAuth 2.0 with tokens, to ensure that only authorized users have access. Use HTTPS to secure the transmission of tokens.
2	API	Tampering with data	An attacker attempts to modify the data in the API requests or responses to perform unauthorized actions.	Low	High	Low	Use HTTPS to ensure data integrity. Perform server-side validation on all inputs to ensure that only valid data is accepted.
3	API	Repudiation	A user denies performing certain actions via the API.	Medium	Low	Low	Implement logging of API calls, including details such as user IDs, timestamps, and the actions performed. This helps to determine the origin of the calls.
4	API	Information disclosure	Unauthorized access to confidential information via the API.	Medium	High	High	Restrict access to certain API endpoints based on user roles. Encrypt sensitive data at rest and during transmission using HTTPS.
5	API	Denial of service	An attacker launches a DDoS attack to overload and make the API inaccessible.	Low	High	Medium	Implement rate limiting to restrict the number of calls per unit of time. Use a Content Delivery Network (CDN) to reduce the load.
6	API	Elevation of privilege	A user is attempting to increase their access rights without authorization.	Very low	High	Medium	Implement the principle of least privilege. Ensure that each user has access only to the resources necessary for their tasks. Conduct regular reviews to detect unauthorized privilege escalation.
7	External source	Spoofing and identity	An attacker poses as a legitimate user to gain access to sensitive geo-data.	Very low	High	Medium	Implement strong authentication for users who want to access certain geo-data. Use API keys or tokens to manage access.
8	External source	Tampering with data	An attacker attempts to manipulate geographical data to disseminate false information.	Low	High	Medium	Ensure data integrity is maintained, by using digital signatures or checksums for downloaded data. Implement control mechanisms at the server level to detect and prevent unauthorized changes.
9	External source	Repudiation	A user denies executing certain search queries or requesting specific geo-data.	Low	Medium	Low	Implement logging and auditing of search queries and data retrievals, including user data and timestamps. This helps maintain a reliable record of user activities and ensures accountability.
10	External source	Information disclosure	Unauthorized access to sensitive geo-data.	Low	High	Medium	Restrict access to specific datasets based on user rights. Encrypt sensitive geo-data during storage and transmission to enhance data security and protect against unauthorized access.
11	External source	Denial of service	An attacker launches a DDoS attack to make the website inaccessible.	Low	High	Medium	Implement DDoS protection technologies, such as the use of a Content Delivery Network (CDN), to ensure the availability of the website
12	External source	Elevation of privilege	A user attempts to unlawfully elevate their access rights to obtain restricted geo-data.	Low	High	Medium	Implement the principle of least privilege, ensuring that users only have access to the geo-data necessary for their specific tasks. Conduct regular assessments to detect any unauthorized elevation of privileges.

Figure 5 Risk register

When this is placed in a risk matrix, as can been seen in Figure 6, it can be seen that risks 1 and 4 come out as the highest risks; these lie within the API.

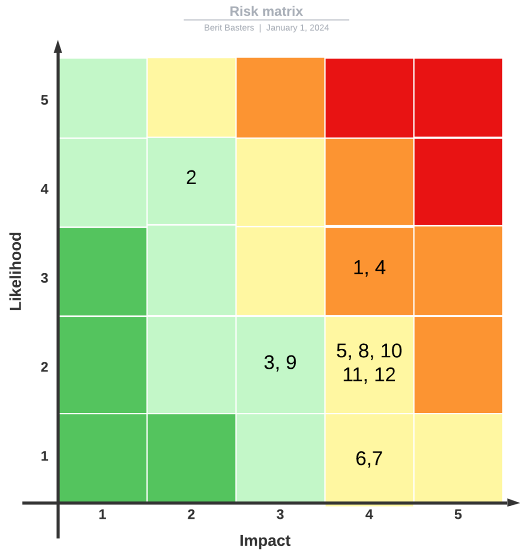


Figure 6 Risk matrix

4.5 Results question 5

Question: What are the security requirements and priorities of the stakeholders?
For this question methodology Subquestion 5 was used. Unfortunately, direct engagement with stakeholders was hampered due to their schedules. However, conversations were held with two experts to gather their perspectives on the research and proposed mitigations. A notable suggestion from one expert, Mr. Knotter, was to consider the implications of the newly agreed-upon AI act by the EU. This act could have significant relevance and impact on the research outcomes. The EUs AI Act, agreed upon in December 2023, represents a significant step in regulating AI technologies. It is considered the world's first comprehensive legal framework on AI, setting a precedent for future global AI legislation. In the research on the new AI Act, it was found that as long as the project does not utilize personal data and remains inaccessible to the general public, it will not be impacted by this regulation. The model used must be classified according to risk level; however, it is expected to be categorized as “limited risk” owing to the minimal user data involved. Importantly, the model does not require any personal data from users for its intended functions, aligning with the Act's compliance requirements.

Additionally, expert opinions were incorporated to enhance the mitigation strategies identified in the research. This not only supplements our findings but also provides validation for the work completed thus far. For a comprehensive view of these expert insights, please refer to appendices C and D.

4.6 Results question 6

Question: What are possible solutions to eliminating or minimizing these risks that can be implemented into the NarcoView platform?
To answer this question, the risk matrix in Figure 6 must be examined. This helps identify and prioritize the highest risks for immediate improvement. To provide an estimate for the time and costs of implementing these mitigations along with the order in which they should be addressed, the complexity, resources, and potential dependencies need to be considered.

Further using the methodology from Subquestion 6. The risks categorized as high (1, 4, and 6) include implementing hypertext transfer protocol secure (HTTPS) and two-factor authentication (2FA), restricting physical access to the API by using a private server, and managing user-level access within the application. Implementing HTTPS and 2FA would take about one-to-two weeks each. Restricting access can take two-to-four weeks or longer depending on the place and building management. Managing user-level access can take three-to-five weeks of development. Costs include purchasing SSL certificates and integrating 2FA services. An organization-validated certificate should be enough protection for the use of the NarcoView application. These certificates range from €20 to €250 per year. Buying a server for the project could also cost up to €50,000. In addition would be work costs for all the development time.

For medium risks (5, 8, 10, 11, and 12), strategies include implementing rate limiting (‘What Is Rate Limiting | Types & Algorithms | Imperva’, z.d.) to guard against denial-of-service (DOS) attacks using a content delivery network (CDN) (*What is a content delivery network (CDN)? | How do CDNs work? | Cloudflare*, z.d.) to minimize load and applying digital signatures (*Digital Signature Service - DSS*, z.d.) for report downloads. Implementing rate limiting would take around one-to-two weeks, implementing a CDN around one week, and making digital signatures for reports around two-to-three weeks. The costs for rate limiting and digital signatures would only be the development time. The costs for a CDN are generally around a few hundred euros per month.

The lower risks (2, 3, and 9) involve establishing logging protocols for API calls, including timestamps and user IDs. Implementing logging for the API would take around two-to-four weeks and would mainly cost development time.

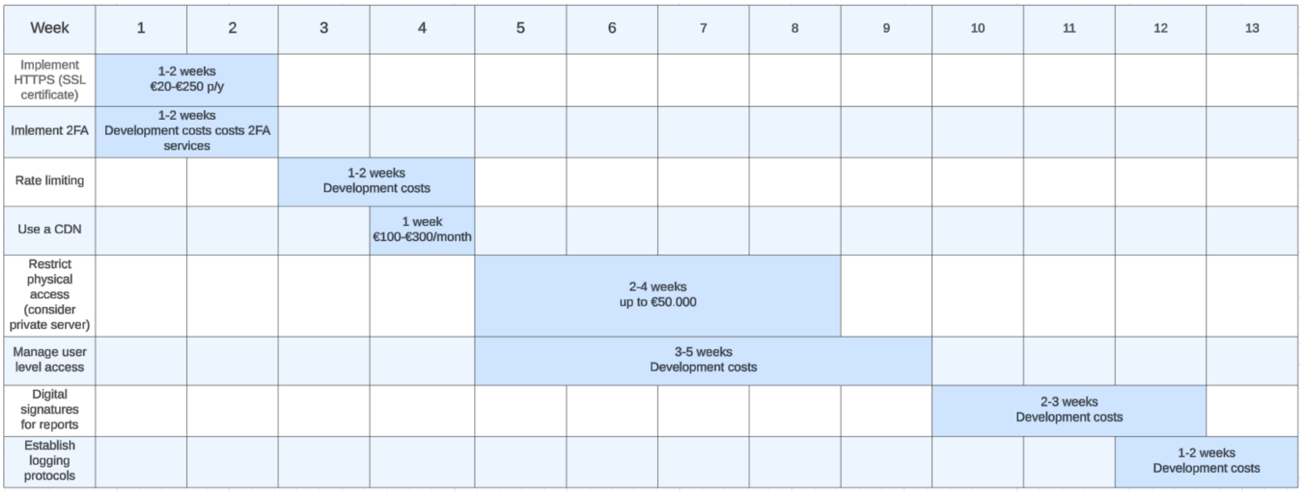
There are some general considerations for these mitigations. Some depend on the completion of others (e.g., user-level access management might depend on implementing 2FA). The resources also need to be considered: How many people are available, and what kind of expertise do they have? Lastly, the costs of ongoing maintenance and updates must be considered.

The recommended order for these implementations would be as follows:

1. Implement HTTPS
2. Implement 2FA
3. Rate limiting
4. Use a CDN
5. Restrict physical access
6. Manage user-level access
7. Digital signatures for reports
8. Establish logging protocols

To visualize this, I made a roadmap to share within the team and create awareness for these risks (see Figure 8). The roapmap was made with the research and results of this thesis.

Implement HTTPS (SSL certificate)



Week	1	2	3	4	5	6	7	8	9	10	11	12	13
Implement HTTPS (SSL certificate)	1-2 weeks €20-€250 p/y												
Imlement 2FA	1-2 weeks Development costs costs 2FA services												
Rate limiting			1-2 weeks Development costs										
Use a CDN				1 week €100-€300/month									
Restrict physical access (consider private server)					2-4 weeks up to €50.000								
Manage user level access					3-5 weeks Development costs								
Digital signatures for reports										2-3 weeks Development costs			
Establish logging protocols												1-2 weeks Development costs	

Figure 7 Roadmap

In the roadmap, it is shown how the various recommendations can be implemented, including time and cost estimates. Some costs are difficult to predict and require further research to determine the actual expenses, particularly regarding development time. These costs can vary significantly with each implementation and over time. This also means that the implementation is only a guideline and can be influenced in practice by various factors such as priority, budget, and staffing.

5. Conclusion

In this chapter is discussed the answer to the main question of this thesis: How does the NarcoView project ensure secure handling of data used on the platform and customize it to meet the data security requirements of the stakeholders? This will be answered using the results of the subquestions.

5.1 Conclusion question 1

Which theoretical perspectives and concepts can help understand how to assess data risks?
The literature review resulted in theoretical perspectives and concepts such as data profiling, threat modelling, and security legislation. The methodologies derived from Abedjan's data profiling and the OWASP and STRIDE models provide a comprehensive framework for identifying and managing potential risks.

5.2 Conclusion question 2

What is the origin of the data necessary in using the NarcoView platform?
Discussions with team members and subsequent data profiling revealed that the NarcoView platform relies on a mix of internal and external data sources. The integration of data types, ranging from risk maps to satellite imagery, underscores the platform's broad approach to data collection and utilization.

5.3 Conclusion question 3

What is the intention of the data that will be used in the NarcoView platform?
The intention behind the data used in the NarcoView platform is primarily for monitoring, predictions, and crime detection. This is facilitated by profiling the data and assigning different intended users and usage purposes, ensuring that the data serves its intended function.

5.4 Conclusion question 4

Which specific risks and threats are associated with the collection, storage, and processing of the data used in the NarcoView platform?
A detailed risk analysis using OWASP and STRIDE methodologies revealed various risks, particularly related to API and external data sources. However, not all risks are within the control of the research group due to sensitive information managed by stakeholders. The research suggests that careful consideration of these risks is crucial for the platform's security.

5.5 Conclusion question 5

What are the security requirements and priorities of the stakeholders?
While direct engagement with all stakeholders was not possible, insights from experts highlighted the importance of considering the implications of regulatory frameworks like the EU's AI Act. The stakeholders prioritize a secure platform that aligns with legal standards and effectively mitigates identified risks.

5.6 Conclusion question 6

What are possible solutions to eliminate or minimize these risks that can be implemented into the NarcoView platform?

The proposed solutions include a range of technical and procedural measures, such as implementing HTTPS, two-factor authentication, and managing user-level access. These solutions are prioritized based on the severity and impact of the identified risks.

5.7 Final conclusion

The NarcoView project should ensure secure handling of data by integrating a mix of theoretical approaches and practical solutions, aligning with both stakeholder expectations and regulatory requirements. It involves data profiling, risk analysis, and stakeholder engagement to mitigate risks and prioritize security.

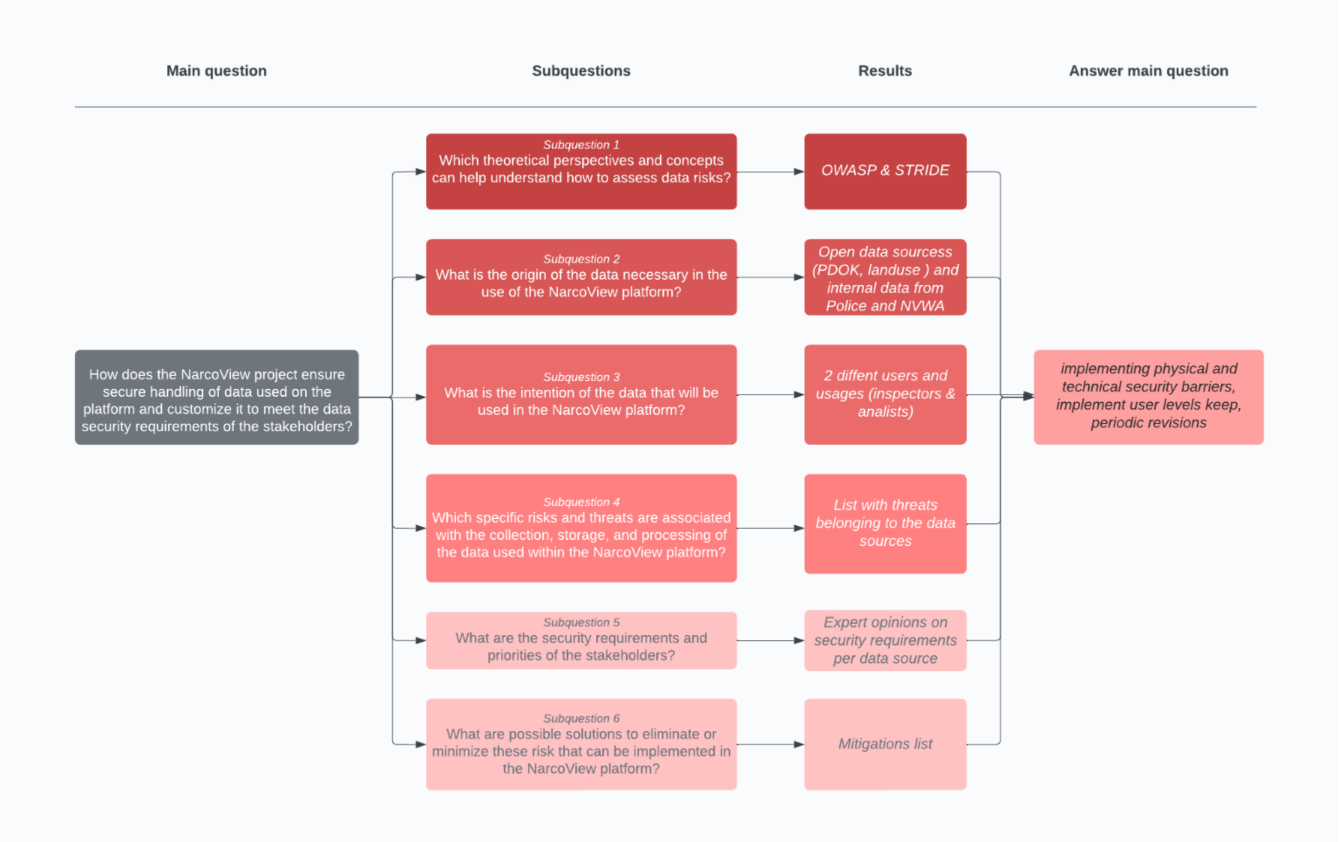

Main question
Subquestions
Results
Answer main question
How does the NarcoView project ensure secure handling of data used on the platform and customize it to meet the data security requirements of the stakeholders?
Subquestion 1
Which theoretical perspectives and concepts can help understand how to assess data risks?
OWASP & STRIDE
Subquestion 2
What is the origin of the data necessary in the use of the NarcoView platform?
Open data sourcess (PDOK, landuse) and internal data from Police and NVWA
Subquestion 3
What is the intention of the data that will be used in the NarcoView platform?
2 diffent users and usages (inspectors & analists)
Subquestion 4
Which specific risks and threats are associated with the collection, storage, and processing of the data used within the NarcoView platform?
List with threats belonging to the data sources
Subquestion 5
What are the security requirements and priorities of the stakeholders?
Expert opinions on security requirements per data source
Subquestion 6
What are possible solutions to eliminate or minimize these risk that can be implemented in the NarcoView platform?
Mitigations list
implementing physical and technical security barriers, implement user levels keep, periodic revisions

Figure 8 Visual Summary Thesis

In the figure above Figure 8 Visual Summary Thesis shows a visual summary of the core of the thesis. This figure shows how the main question leads to subquestions and how these answers lead back to answer the main question.

6. Discussion

In examining the validity and reliability of this research, it is important to recognize the systematic approach employed, grounded in the Verschuren and Doorewaard method. This methodology, aligned with iterative processes, ensures a clear framework allowing for adjustments based on emerging insights and feedback. Based on this, it can be stated that if this research were to be repeated, the results would be the same, and thereby the results of this research are valid.

The results of this research, fundamentally based on data profiling, threat modelling, and security legislation, provide an understanding of data security in innovation projects. It is evident that the outcomes did not entirely correspond with the initial expectations, particularly regarding the specific risks and threats associated with data handling. A possible explanation for this result could be the evolving nature of cyber threats and the ongoing progress in the project affecting data security. Furthermore, the current research complements the existing literature on data security in innovation projects.

This research contributes new insights by emphasizing the importance of regulatory frameworks like the EU's AI Act and the application of STRIDE and OWASP methodologies in a context-specific setting. It underscores the necessity of considering a broad spectrum of data sources and associated risks, thereby enriching the literature on data security in technological innovation.

While the research is comprehensive, certain limitations must be acknowledged. The reliance on the research group's control over data security measures, the limited engagement with all stakeholders, and the specific focus on nonpersonal data in the NarcoView platform might have influenced the results. The implication of these limitations is a call for broader stakeholder engagement and a more diversified data approach in future research.

Considering the findings and limitations, the recommendation for further research is to explore the integration of advanced AI and machine learning techniques for predictive analytics in data security. Further research might also focus on the human aspect of data security, exploring how user behaviour and organizational culture impact the effectiveness of technical safeguards.

In conclusion, this research provides a comprehensive understanding of data security challenges and opportunities within the NarcoView project. It lays a foundation for future studies and practical applications, aiming to enhance the security and efficiency of data handling in innovative technological solutions.

7. Recommendations

This chapter outlines a series of structured recommendations derived from the conclusions of the research conducted within the NarcoView project. These recommendations are designed to enhance data handling, physical security, and digital security, catering to the needs of law enforcement and other stakeholders involved in the project. The recommendations are presented in a sequence that reflects their priority and relevance to the core objectives of enhancing data security and efficiency.

7.1 Data handling

1. Continuously update the list of data sources.
Regular updates to the data source list ensure that all information remains current and relevant, reducing the risk of relying on outdated or inaccurate data. This is vital for maintaining the integrity and reliability of the system, as emphasized in the project's focus on data profiling and integration feasibility.

2. Assess each new data source for risks before integration.
Evaluating every new data source for potential risks ensures that vulnerabilities are identified and addressed early. This can be achieved by using the threat modelling methodologies OWASP and STRIDE. This is essential in preventing data breaches and ensuring data quality and security.

3. Periodically review new developments in laws and regulations.
Staying informed about changes in legislation and regulations enables the organization to remain compliant and avoid legal repercussions. It ensures that the company's data handling practices are always aligned with the latest legal standards, particularly considering the evolving nature of regulations like the EU's AI Act.

4. Implement user access levels.
By creating different user access levels, the organization can control who has access to what information. This minimizes the risk of sensitive data being exposed to unauthorized individuals and ensures that users only have access to the data necessary for their roles, thereby enhancing security and operational efficiency.

7.2 Server security

1. Secure the server both online and physically.
Protecting the server from both digital and physical threats is crucial. Implementing HTTPS secures data transmission online, while ensuring the server is in a locked and controlled access room protects it from physical tampering or theft. These measures are vital in safeguarding sensitive data and the platform's overall integrity.

7.3 Digital security

1. Refrain from storing or processing personal data.
Avoiding the storage and processing of personal data minimizes the risk of privacy breaches and reduces the organization's liability. It also aligns with privacy regulations and builds trust with stakeholders concerned about their personal information. This approach is particularly pertinent given the project's focus on nonpersonal data and regulatory considerations such as the EU's AI Act.

8. Reflection

My thesis focuses on innovating the detection of illegal drug dumping using machine learning and predictive models. The primary goal was to assess the safety of the data used and to provide measures to keep this data secure now and in the future.

I collected information for my thesis through interviews and literature research. I conducted risk analyses using STRIDE and OWASP to identify risks and provided mitigation advice based on expert opinions. A major challenge was writing and communicating in English, dealing with changes in an innovation project, and setbacks in contacting stakeholders.

The thesis has brought to light crucial data security threats and has shown what still needs to be improved or implemented to keep data safe. A notable finding was the lack of specific regulations for a project like Narcoview. These findings have led to a greater awareness of the risks associated with data and the need for thorough assessment before using it.

Writing this thesis has significantly improved my proficiency in English, as well as my ability to think ahead in research. I have learned to read critically, and my research skills have significantly improved.

The most significant feedback came on the setup of the research, as my method is less common in HBO (university of applied sciences) studies. The feedback from my supervisors was always useful and helped me improve my thesis, work through problems, and gain new insights.

An important lesson I learned is the importance of being present and asking questions as soon as they arise. These personal improvements will help me in my future job or studies. This journey was very educational and has prepared me for the challenges I will encounter in my professional life.

Bibiliography

Abedjan, Z., Golab, L., & Naumann, F. (2017). Data Profiling: A Tutorial. *Proceedings of the 2017 ACM International Conference on Management of Data*, 1747-1751. https://doi.org/10.1145/3035918.3054772

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In *Handbook of Practical Program Evaluation* (pp. 492-505). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch19

Digital Signature Service—DSS. (z.d.). Geraadpleegd 19 januari 2024, van https://ec.europa.eu/digital-building-blocks/sites/digital-building-blocks/sites/display/DIGITAL/Digital+Signature+Service+-++DSS

EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act. (z.d.). Geraadpleegd 19 januari 2024, van https://artificialintelligenceact.eu/

General Data Protection Regulation (GDPR) – Official Legal Text. (z.d.). General Data Protection Regulation (GDPR). Geraadpleegd 19 januari 2024, van https://gdpr-info.eu/

jegeib. (2023, juni 1). *Bedreigingen—Microsoft Threat Modeling Tool—Azure*. https://learn.microsoft.com/nl-nl/azure/security/develop/threat-modeling-tool-threats

Koninkrijksrelaties, M. van B. Z. en. (2017, oktober 19). *Data protection—Personal data—Government.nl* [Onderwerp]. Ministerie van Algemene Zaken. https://www.government.nl/topics/personal-data/data-protection

OWASP Top Ten | OWASP Foundation. (z.d.). Geraadpleegd 21 september 2023, van https://owasp.org/www-project-top-ten/

Ree, H. de. (2023, juni 2). *Explosieve toename dumpingen chemisch afval in West-Brabant, zorgen over waterverontreiniging: 'Drugsproductie draait weer op volle toeren'*. bndestem.nl. https://www.bndestem.nl/breda/br-explosieve-toename-dumpingen-chemisch-afval-in-west-brabant-zorgen-over-waterverontreiniging-drugsproductie-draait-weer-op-volle-toeren~ac92fe722/

Threat Modeling—OWASP Cheat Sheet Series. (z.d.). Geraadpleegd 21 september 2023, van https://cheatsheetseries.owasp.org/cheatsheets/Threat_Modeling_Cheat_Sheet.html

UML Class Diagram Tutorial. (z.d.). Lucidchart. Geraadpleegd 1 januari 2024, van https://www.lucidchart.com/pages/uml-class-diagram

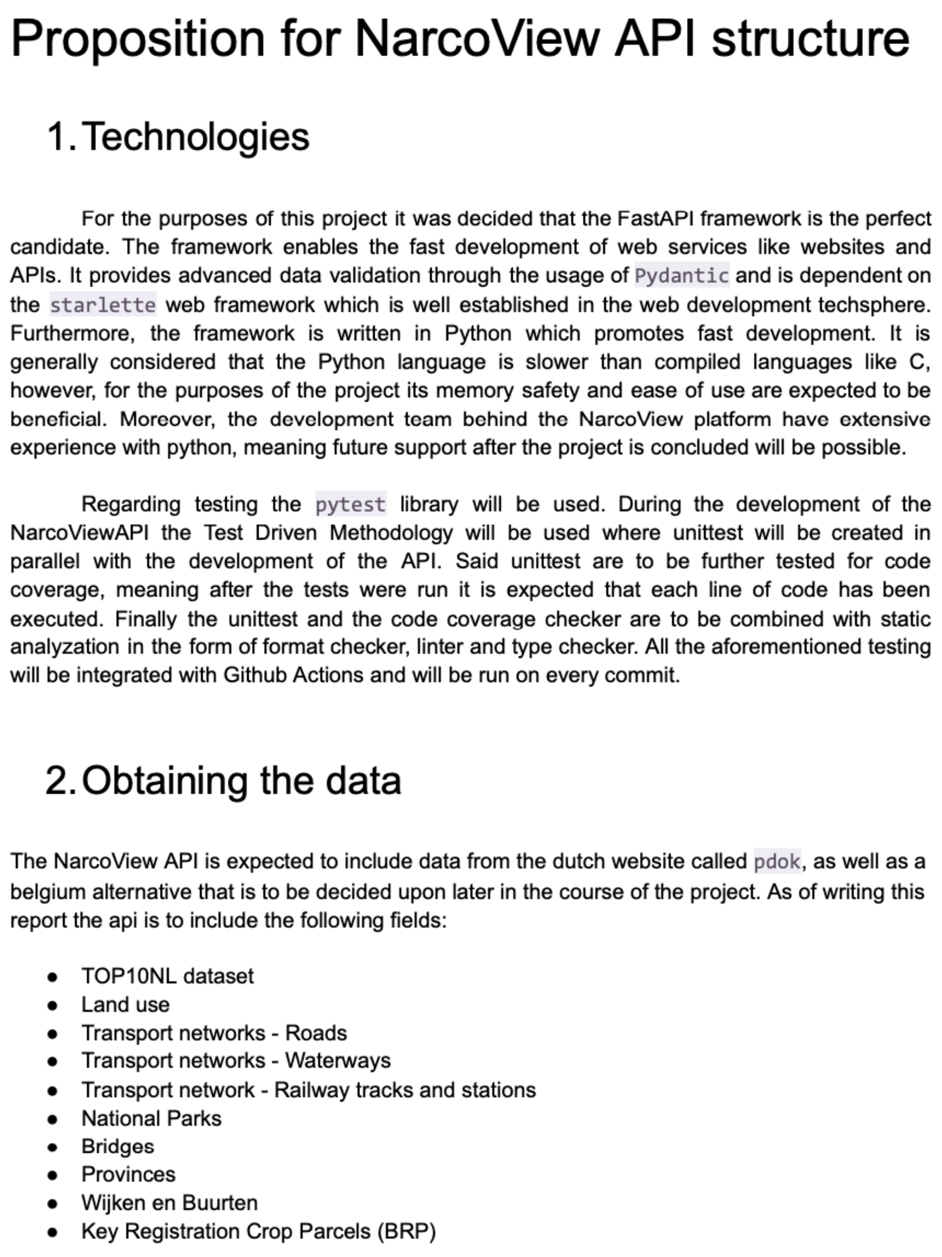
Verschuren, P., & Doorewaard, H. (z.d.). *Designing a Research Project*.

What is a content delivery network (CDN)? | How do CDNs work? | Cloudflare. (z.d.). Geraadpleegd 19 januari 2024, van https://www.cloudflare.com/learning/cdn/what-is-a-cdn/#

What is Rate Limiting | Types & Algorithms | Imperva. (z.d.). *Learning Center*. Geraadpleegd 19 januari 2024, van https://www.imperva.com/learn/application-security/rate-limiting/

Appendix

Appendix A
Proposition for Narcoview API structure



Proposition for NarcoView API structure

1. Technologies

For the purposes of this project it was decided that the FastAPI framework is the perfect candidate. The framework enables the fast development of web services like websites and APIs. It provides advanced data validation through the usage of `Pydantic` and is dependent on the `starlette` web framework which is well established in the web development techsphere. Furthermore, the framework is written in Python which promotes fast development. It is generally considered that the Python language is slower than compiled languages like C, however, for the purposes of the project its memory safety and ease of use are expected to be beneficial. Moreover, the development team behind the NarcoView platform have extensive experience with python, meaning future support after the project is concluded will be possible.

Regarding testing the `pytest` library will be used. During the development of the NarcoViewAPI the Test Driven Methodology will be used where unittest will be created in parallel with the development of the API. Said unittest are to be further tested for code coverage, meaning after the tests were run it is expected that each line of code has been executed. Finally the unittest and the code coverage checker are to be combined with static analyzation in the form of format checker, linter and type checker. All the aforementioned testing will be integrated with Github Actions and will be run on every commit.

2. Obtaining the data

The NarcoView API is expected to include data from the dutch website called `pdok`, as well as a belgium alternative that is to be decided upon later in the course of the project. As of writing this report the api is to include the following fields:

- TOP10NL dataset
- Land use
- Transport networks - Roads
- Transport networks - Waterways
- Transport network - Railway tracks and stations
- National Parks
- Bridges
- Provinces
- Wijken en Buurten
- Key Registration Crop Parcels (BRP)

n

- NDVI Satellite data (It must be noted that said data will not be extracted from pdok and may required different type of processing)

All of these fields have appropriate web services in pdok.nl from which are WMS (Web Map services), WFS (Web Feature Services), WCS (Web Coverage Services) and WMTS (Web Map TIle Services). It is important to note that if a field, for instance Land use, may have multiple web services, thus all will be supported. Said services are generally used by GIS software, however, they can be utilized using a Python library called `owslib`. This library is created for the purposes of extracting data from OGC (Open Geospatial Consortium) web service, like the aforementioned.

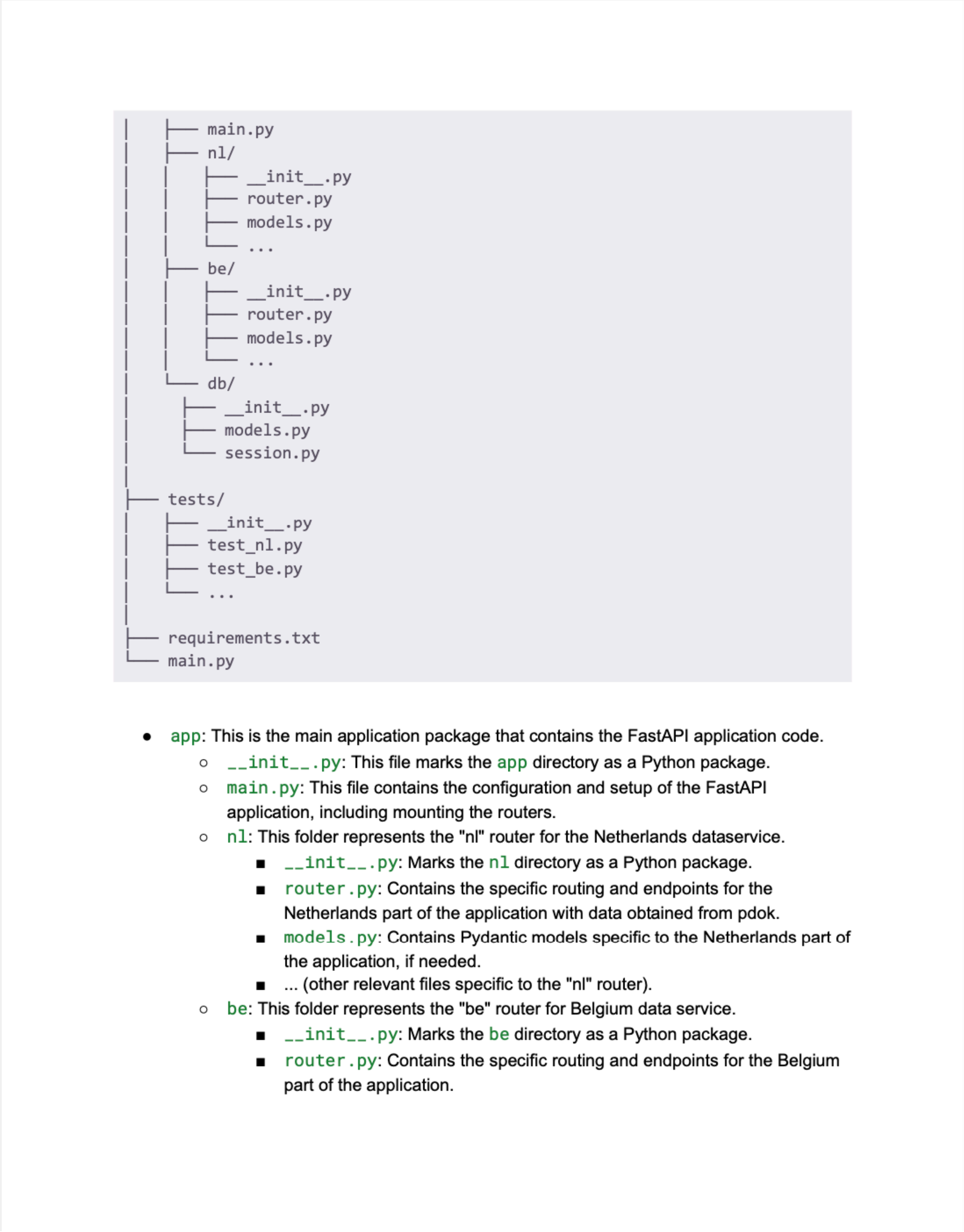
For ease of understanding a simple example will be given on how data can be obtained from the "Land use" geoservice:

```
from owslib.wms import WebMapService
wms_url =
'https://service.pdok.nl/cbs/bestandbodemgebruik/2010/wms/v1_0?request=getc
apabilities&service=wms'
wms = WebMapService(wms_url)
layers = list(wms.contents)


layer_name = layers[0]
bbox = (xmin, ymin, xmax, ymax)  width = 800
height = 600
image_format = 'image/png'

img = wms.getmap(layers=[layer_name], srs='EPSG:4326', bbox=bbox,
size=(width, height), format=image_format)
```

3. Project structure

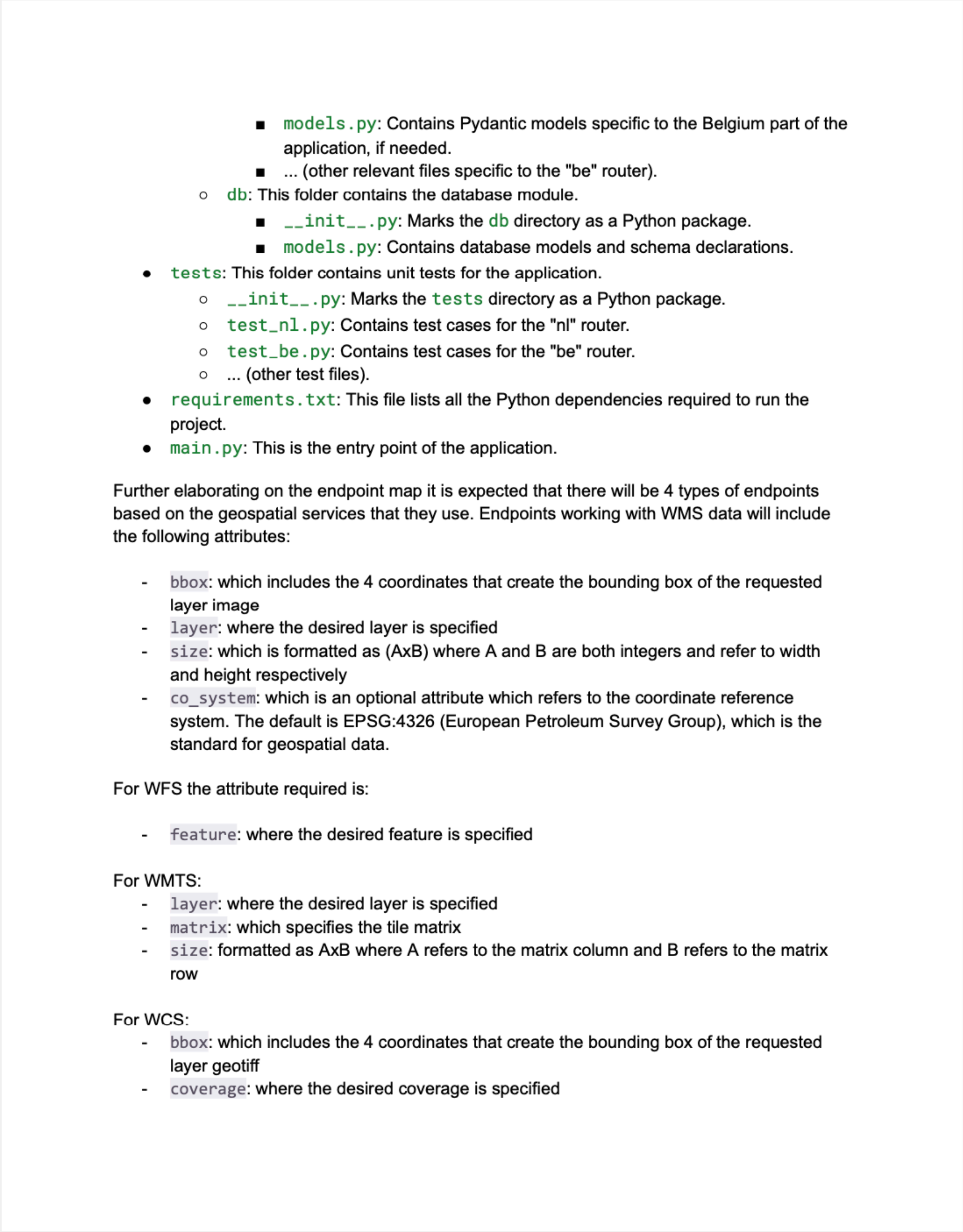
For the purposes of this project the following folder structure will be developed:

```
my_fastapi_project/
│
├── app/
│   ├── __init__.py
```



```
│   ├── main.py
│   ├── nl/
│   │   ├── __init__.py
│   │   ├── router.py
│   │   ├── models.py
│   │   └── ...
│   ├── be/
│   │   ├── __init__.py
│   │   ├── router.py
│   │   ├── models.py
│   │   └── ...
│   └── db/
│     ├── __init__.py
│     ├── models.py
│     └── session.py
│
├── tests/
│   ├── __init__.py
│   ├── test_nl.py
│   ├── test_be.py
│   └── ...
│
├── requirements.txt
└── main.py
```

- `app`: This is the main application package that contains the FastAPI application code.
  - `__init__.py`: This file marks the `app` directory as a Python package.
  - `main.py`: This file contains the configuration and setup of the FastAPI application, including mounting the routers.
  - `nl`: This folder represents the "nl" router for the Netherlands dataservice.
    - `__init__.py`: Marks the `nl` directory as a Python package.
    - `router.py`: Contains the specific routing and endpoints for the Netherlands part of the application with data obtained from pdok.
    - `models.py`: Contains Pydantic models specific to the Netherlands part of the application, if needed.
    - ... (other relevant files specific to the "nl" router).
  - `be`: This folder represents the "be" router for Belgium data service.
    - `__init__.py`: Marks the `be` directory as a Python package.
    - `router.py`: Contains the specific routing and endpoints for the Belgium part of the application.



        - `models.py`: Contains Pydantic models specific to the Belgium part of the application, if needed.
        - ... (other relevant files specific to the "be" router).
    - `db`: This folder contains the database module.
        - `__init__.py`: Marks the `db` directory as a Python package.
        - `models.py`: Contains database models and schema declarations.
- `tests`: This folder contains unit tests for the application.
    - `__init__.py`: Marks the `tests` directory as a Python package.
    - `test_nl.py`: Contains test cases for the "nl" router.
    - `test_be.py`: Contains test cases for the "be" router.
    - ... (other test files).
- `requirements.txt`: This file lists all the Python dependencies required to run the project.
- `main.py`: This is the entry point of the application.

Further elaborating on the endpoint map it is expected that there will be 4 types of endpoints based on the geospatial services that they use. Endpoints working with WMS data will include the following attributes:

- `bbox`: which includes the 4 coordinates that create the bounding box of the requested layer image
- `layer`: where the desired layer is specified
- `size`: which is formatted as (AxB) where A and B are both integers and refer to width and height respectively
- `co_system`: which is an optional attribute which refers to the coordinate reference system. The default is EPSG:4326 (European Petroleum Survey Group), which is the standard for geospatial data.

For WFS the attribute required is:

- `feature`: where the desired feature is specified

For WMTS:

- `layer`: where the desired layer is specified
- `matrix`: which specifies the tile matrix
- `size`: formatted as AxB where A refers to the matrix column and B refers to the matrix row

For WCS:

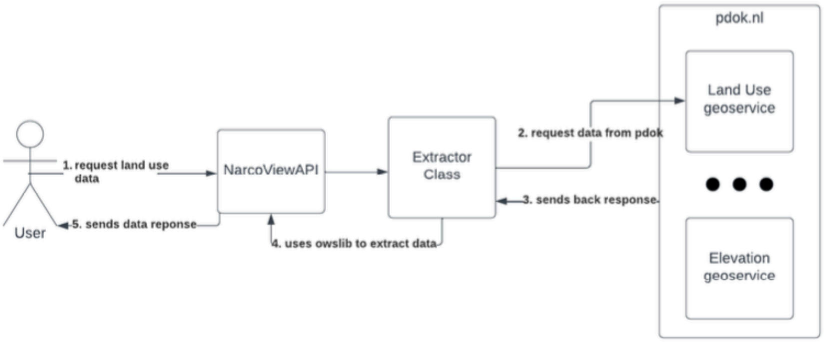
- `bbox`: which includes the 4 coordinates that create the bounding box of the requested layer geotiff
- `coverage`: where the desired coverage is specified

As for the endpoint link it is expected to look the following way:

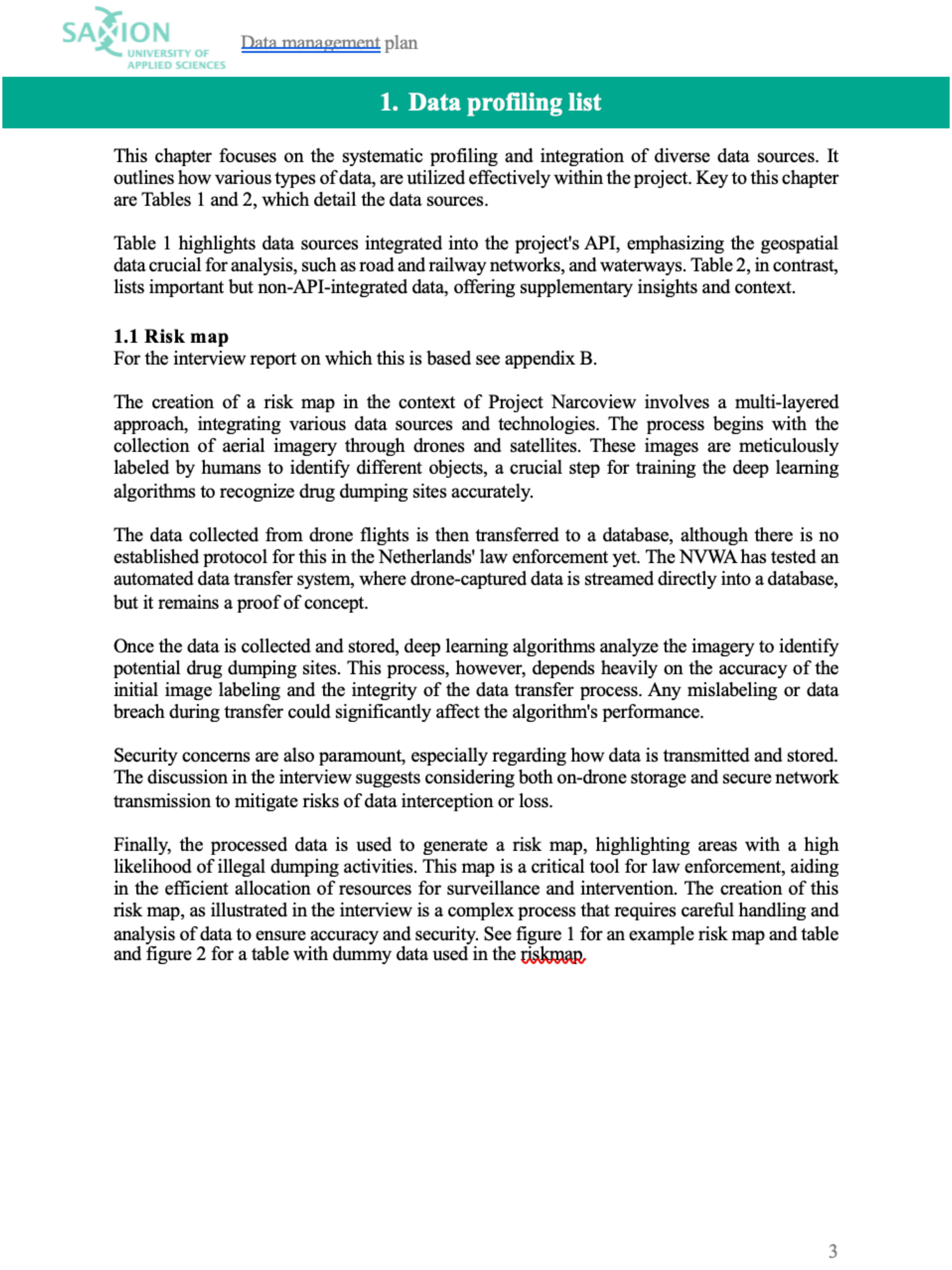
```
<ip>/<router>/<service>/<year>/<attributes>
```

Where the router can be `nl`, `be` or `db`, the service represents one of the aforementioned fields formatted in snake case (`land_use` for instance), the year represents the requested year and the attributes are the attributes related to the webservice (`bbox`, `layer`, etc.) . It must be noted that attributes can be substituted with `` `get_<fields>` `` in case of not having prior knowledge about a services `'layers'`, `'matrix'` and/or `'coverage'`.

For simplification a block diagram was created:



Appendix B
Data management plan



Data management plan

1. Data profiling list

This chapter focuses on the systematic profiling and integration of diverse data sources. It outlines how various types of data, are utilized effectively within the project. Key to this chapter are Tables 1 and 2, which detail the data sources.

Table 1 highlights data sources integrated into the project's API, emphasizing the geospatial data crucial for analysis, such as road and railway networks, and waterways. Table 2, in contrast, lists important but non-API-integrated data, offering supplementary insights and context.

1.1 Risk map

For the interview report on which this is based see appendix B.

The creation of a risk map in the context of Project Narcoview involves a multi-layered approach, integrating various data sources and technologies. The process begins with the collection of aerial imagery through drones and satellites. These images are meticulously labeled by humans to identify different objects, a crucial step for training the deep learning algorithms to recognize drug dumping sites accurately.

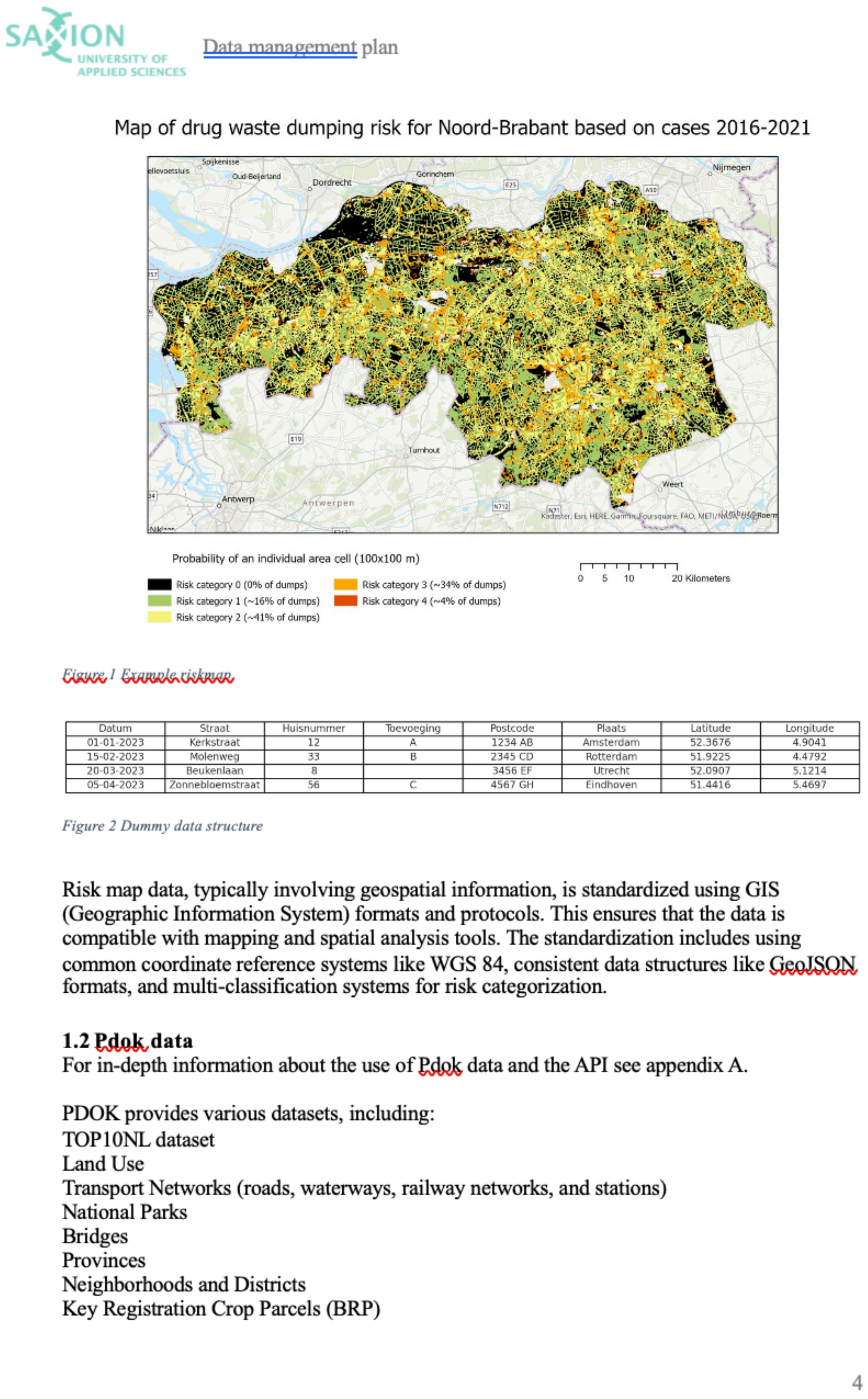
The data collected from drone flights is then transferred to a database, although there is no established protocol for this in the Netherlands' law enforcement yet. The NVWA has tested an automated data transfer system, where drone-captured data is streamed directly into a database, but it remains a proof of concept.

Once the data is collected and stored, deep learning algorithms analyze the imagery to identify potential drug dumping sites. This process, however, depends heavily on the accuracy of the initial image labeling and the integrity of the data transfer process. Any mislabeling or data breach during transfer could significantly affect the algorithm's performance.

Security concerns are also paramount, especially regarding how data is transmitted and stored. The discussion in the interview suggests considering both on-drone storage and secure network transmission to mitigate risks of data interception or loss.

Finally, the processed data is used to generate a risk map, highlighting areas with a high likelihood of illegal dumping activities. This map is a critical tool for law enforcement, aiding in the efficient allocation of resources for surveillance and intervention. The creation of this risk map, as illustrated in the interview is a complex process that requires careful handling and analysis of data to ensure accuracy and security. See figure 1 for an example risk map and table and figure 2 for a table with dummy data used in the riskmap.

3



Data management plan
Map of drug waste dumping risk for Noord-Brabant based on cases 2016-2021
Spijkenisse
Hellevoetsluis
Oud-Beijerland
Dordrecht
Gorinchem
Nijmegen
E25
A50
E19
Turnhout
Weert
Antwerp
Antwerpen
N712
N71
Kadaster, Esri, HERE, Garmin, Foursquare, FAO, METI/NASA, USGS
Probability of an individual area cell (100x100 m)
Risk category 0 (0% of dumps)
Risk category 1 (~16% of dumps)
Risk category 2 (~41% of dumps)
Risk category 3 (~34% of dumps)
Risk category 4 (~4% of dumps)
0 5 10 20 Kilometers

Figure 1 Example riskmap

Datum	Straat	Huisnummer	Toevoeging	Postcode	Plaats	Latitude	Longitude
01-01-2023	Kerkstraat	12	A	1234 AB	Amsterdam	52.3676	4.9041
15-02-2023	Molenweg	33	B	2345 CD	Rotterdam	51.9225	4.4792
20-03-2023	Beukenlaan	8		3456 EF	Utrecht	52.0907	5.1214
05-04-2023	Zonnebloemstraat	56	C	4567 GH	Eindhoven	51.4416	5.4697

Figure 2 Dummy data structure

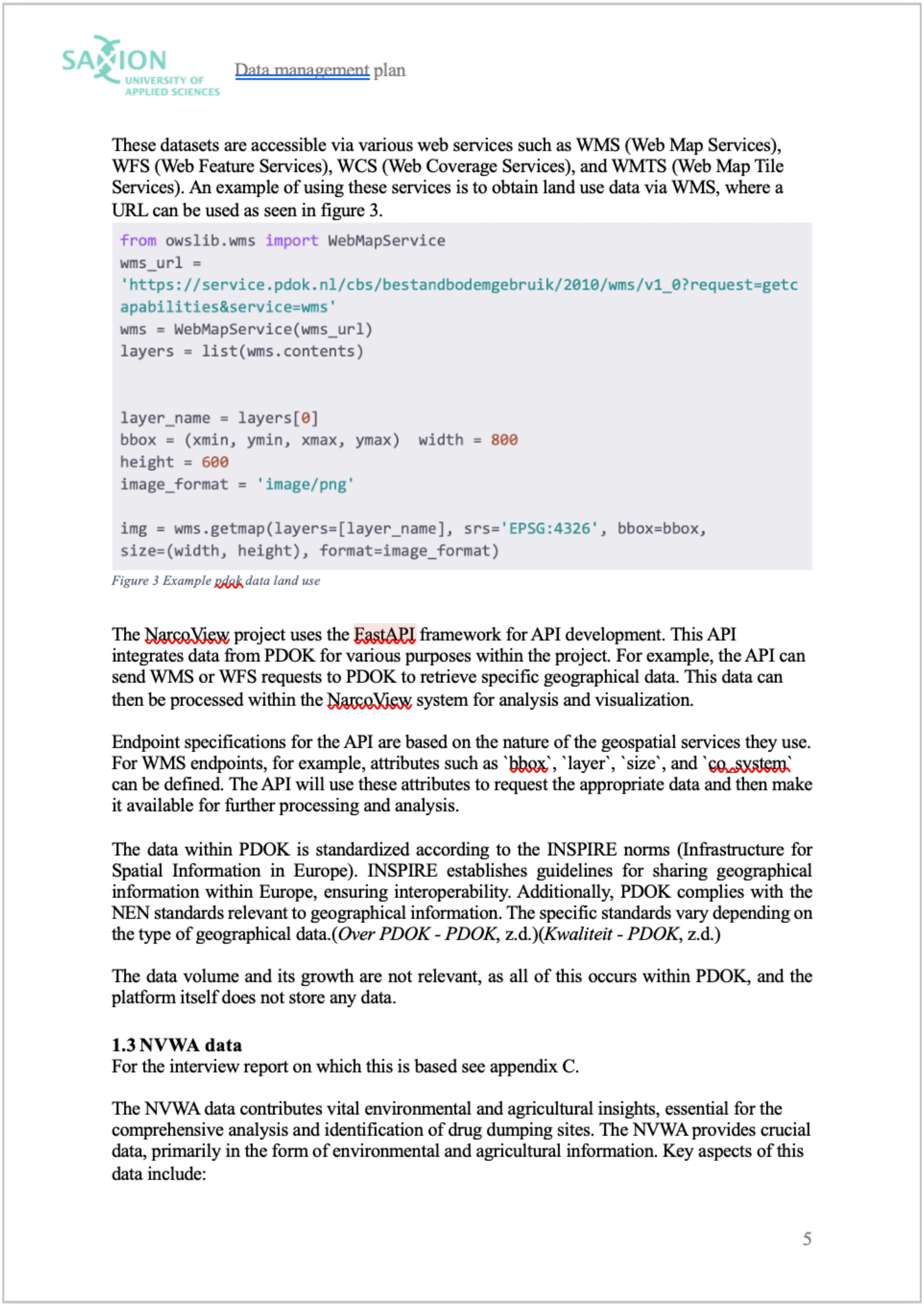
Risk map data, typically involving geospatial information, is standardized using GIS (Geographic Information System) formats and protocols. This ensures that the data is compatible with mapping and spatial analysis tools. The standardization includes using common coordinate reference systems like WGS 84, consistent data structures like GeoJSON formats, and multi-classification systems for risk categorization.

1.2 Pdok data

For in-depth information about the use of Pdok data and the API see appendix A.

PDOK provides various datasets, including:
TOP10NL dataset
Land Use
Transport Networks (roads, waterways, railway networks, and stations)
National Parks
Bridges
Provinces
Neighborhoods and Districts
Key Registration Crop Parcels (BRP)

4



Data management plan

These datasets are accessible via various web services such as WMS (Web Map Services), WFS (Web Feature Services), WCS (Web Coverage Services), and WMTS (Web Map Tile Services). An example of using these services is to obtain land use data via WMS, where a URL can be used as seen in figure 3.

```
from owslib.wms import WebMapService
wms_url =
'https://service.pdok.nl/cbs/bestandbodemgebruik/2010/wms/v1_0?request=getc
apabilities&service=wms'
wms = WebMapService(wms_url)
layers = list(wms.contents)


layer_name = layers[0]
bbox = (xmin, ymin, xmax, ymax)  width = 800
height = 600
image_format = 'image/png'

img = wms.getmap(layers=[layer_name], srs='EPSG:4326', bbox=bbox,
size=(width, height), format=image_format)
```

Figure 3 Example pdok data land use

The NarcoView project uses the FastAPI framework for API development. This API integrates data from PDOK for various purposes within the project. For example, the API can send WMS or WFS requests to PDOK to retrieve specific geographical data. This data can then be processed within the NarcoView system for analysis and visualization.

Endpoint specifications for the API are based on the nature of the geospatial services they use. For WMS endpoints, for example, attributes such as `bbox`, `layer`, `size`, and `co_system` can be defined. The API will use these attributes to request the appropriate data and then make it available for further processing and analysis.

The data within PDOK is standardized according to the INSPIRE norms (Infrastructure for Spatial Information in Europe). INSPIRE establishes guidelines for sharing geographical information within Europe, ensuring interoperability. Additionally, PDOK complies with the NEN standards relevant to geographical information. The specific standards vary depending on the type of geographical data.(*Over PDOK - PDOK*, z.d.)(*Kwaliteit - PDOK*, z.d.)

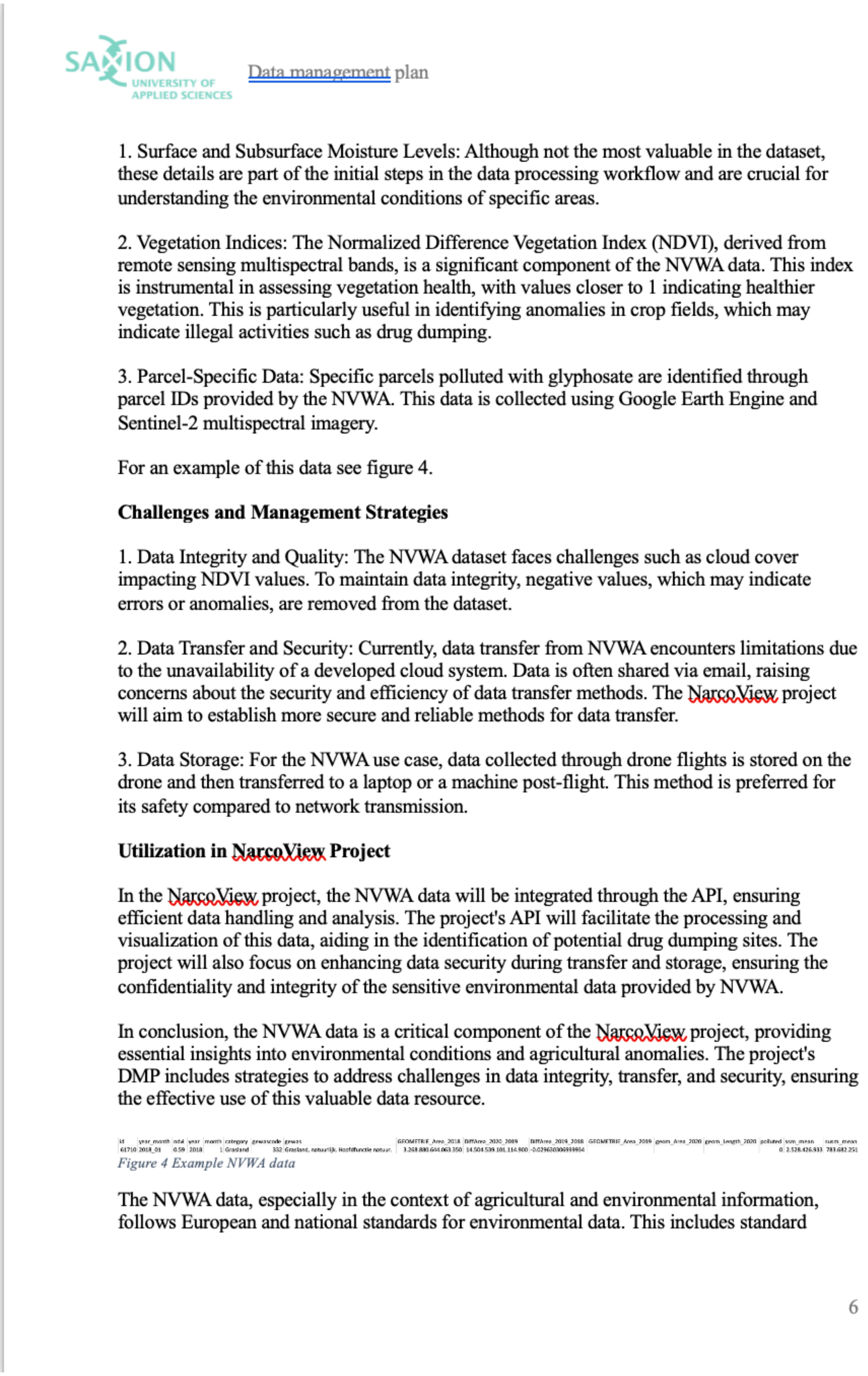
The data volume and its growth are not relevant, as all of this occurs within PDOK, and the platform itself does not store any data.

1.3 NVWA data

For the interview report on which this is based see appendix C.

The NVWA data contributes vital environmental and agricultural insights, essential for the comprehensive analysis and identification of drug dumping sites. The NVWA provides crucial data, primarily in the form of environmental and agricultural information. Key aspects of this data include:

5



1. Surface and Subsurface Moisture Levels: Although not the most valuable in the dataset, these details are part of the initial steps in the data processing workflow and are crucial for understanding the environmental conditions of specific areas.

2. Vegetation Indices: The Normalized Difference Vegetation Index (NDVI), derived from remote sensing multispectral bands, is a significant component of the NVWA data. This index is instrumental in assessing vegetation health, with values closer to 1 indicating healthier vegetation. This is particularly useful in identifying anomalies in crop fields, which may indicate illegal activities such as drug dumping.

3. Parcel-Specific Data: Specific parcels polluted with glyphosate are identified through parcel IDs provided by the NVWA. This data is collected using Google Earth Engine and Sentinel-2 multispectral imagery.

For an example of this data see figure 4.

Challenges and Management Strategies

1. Data Integrity and Quality: The NVWA dataset faces challenges such as cloud cover impacting NDVI values. To maintain data integrity, negative values, which may indicate errors or anomalies, are removed from the dataset.

2. Data Transfer and Security: Currently, data transfer from NVWA encounters limitations due to the unavailability of a developed cloud system. Data is often shared via email, raising concerns about the security and efficiency of data transfer methods. The NarcoView project will aim to establish more secure and reliable methods for data transfer.

3. Data Storage: For the NVWA use case, data collected through drone flights is stored on the drone and then transferred to a laptop or a machine post-flight. This method is preferred for its safety compared to network transmission.

Utilization in NarcoView Project

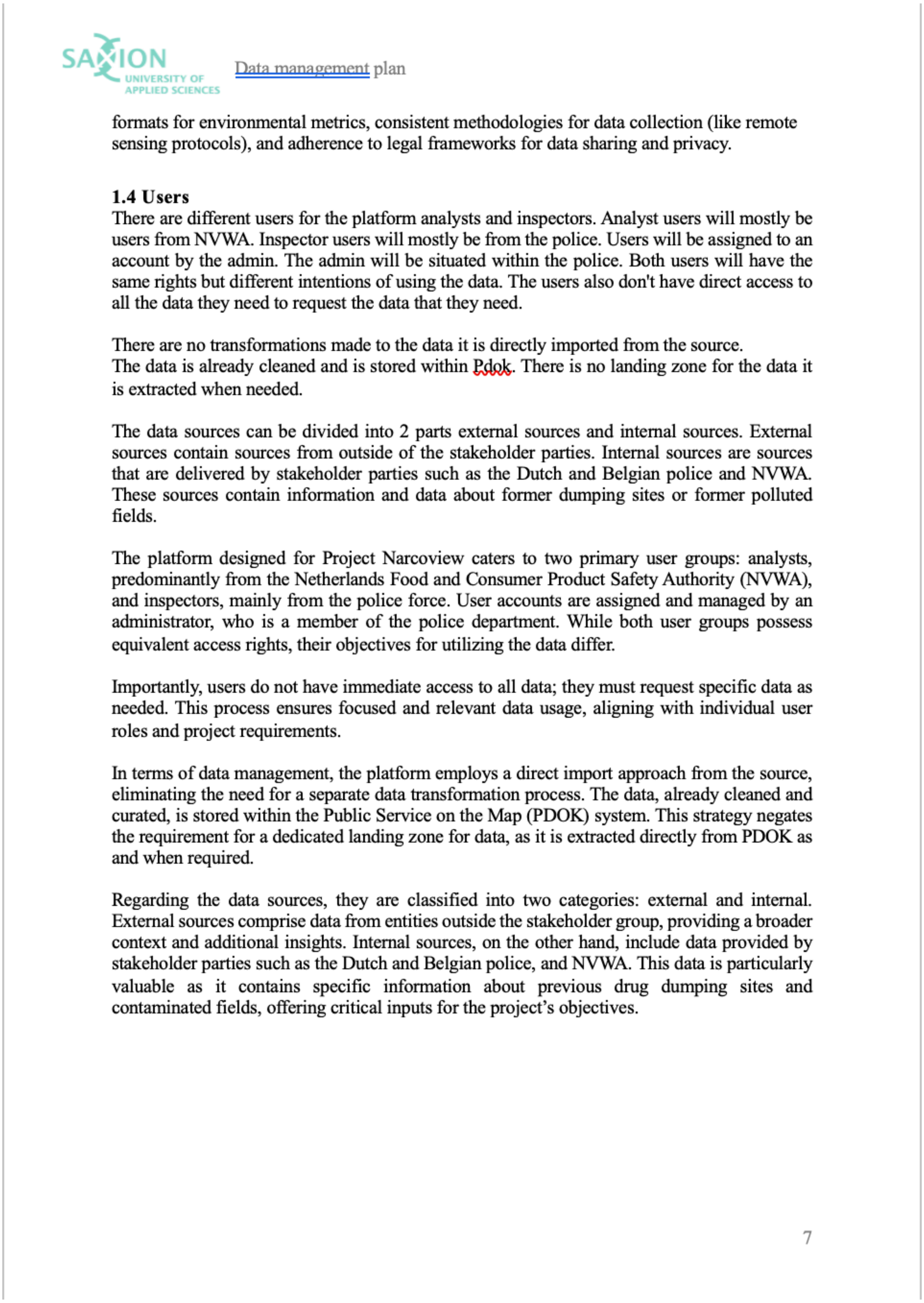
In the NarcoView project, the NVWA data will be integrated through the API, ensuring efficient data handling and analysis. The project's API will facilitate the processing and visualization of this data, aiding in the identification of potential drug dumping sites. The project will also focus on enhancing data security during transfer and storage, ensuring the confidentiality and integrity of the sensitive environmental data provided by NVWA.

In conclusion, the NVWA data is a critical component of the NarcoView project, providing essential insights into environmental conditions and agricultural anomalies. The project's DMP includes strategies to address challenges in data integrity, transfer, and security, ensuring the effective use of this valuable data resource.

id	year_month	ndvi	year	month	category	gewascode	gewas	GEOMETRIE_Area_2018	DiffArea_2020_2019	DiffArea_2019_2018	GEOMETRIE_Area_2019	geom_Area_2020	geom_Length_2020	polluted	ssm_mean	susm_mean
61710	2018_01	0.59	2018	1	Grasland	332	Grasland, natuurlijk. Hoofdfunctie natuur.	3.268.880.644.063.350	14.504.539.101.114.900	-0.0296303069999964				0	2.528.426.933	783.682.251

Figure 4 Example NVWA data

The NVWA data, especially in the context of agricultural and environmental information, follows European and national standards for environmental data. This includes standard



formats for environmental metrics, consistent methodologies for data collection (like remote sensing protocols), and adherence to legal frameworks for data sharing and privacy.

1.4 Users

There are different users for the platform analysts and inspectors. Analyst users will mostly be users from NVWA. Inspector users will mostly be from the police. Users will be assigned to an account by the admin. The admin will be situated within the police. Both users will have the same rights but different intentions of using the data. The users also don't have direct access to all the data they need to request the data that they need.

There are no transformations made to the data it is directly imported from the source.
The data is already cleaned and is stored within Pdok. There is no landing zone for the data it is extracted when needed.

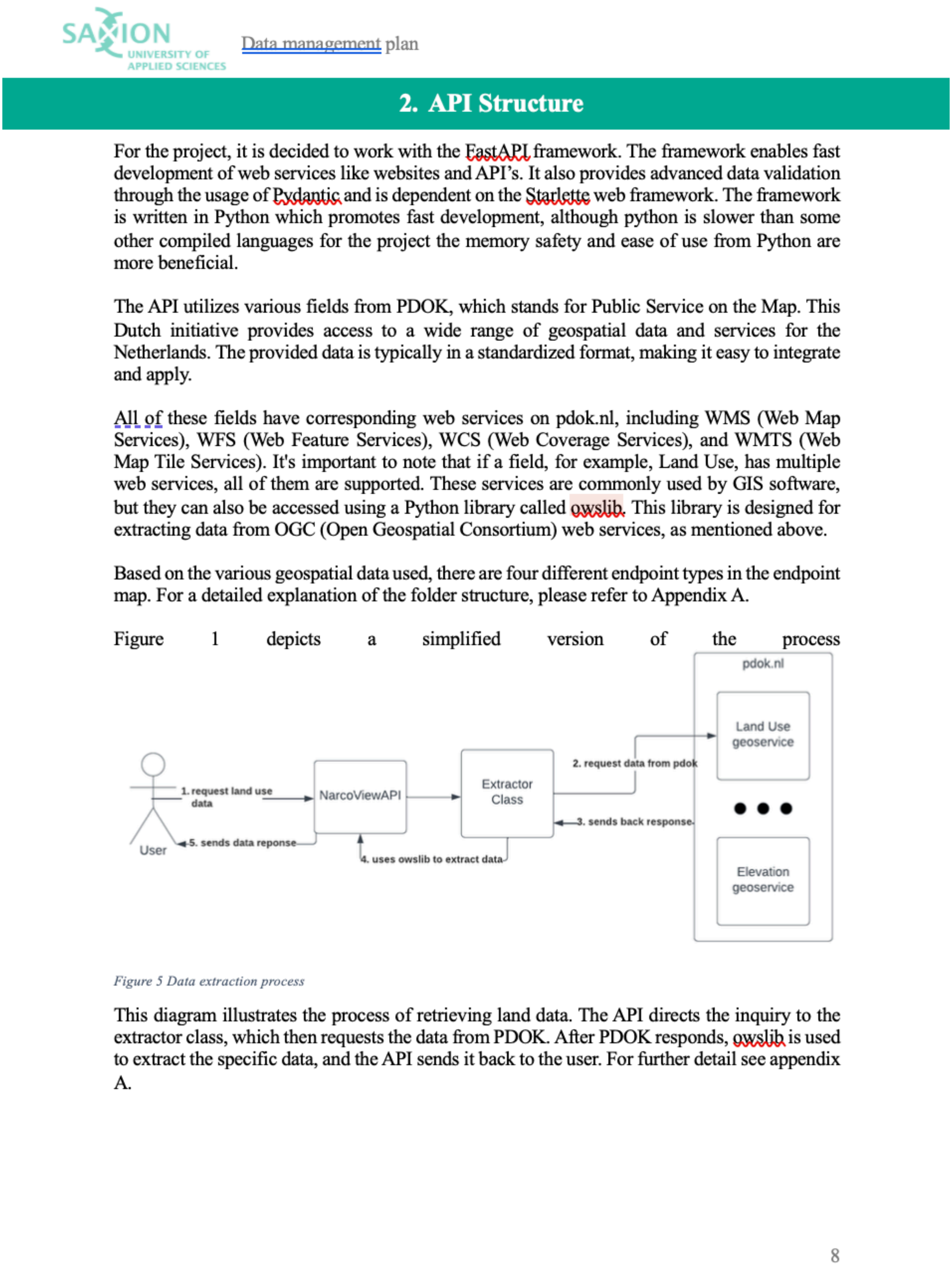
The data sources can be divided into 2 parts external sources and internal sources. External sources contain sources from outside of the stakeholder parties. Internal sources are sources that are delivered by stakeholder parties such as the Dutch and Belgian police and NVWA. These sources contain information and data about former dumping sites or former polluted fields.

The platform designed for Project Narcoview caters to two primary user groups: analysts, predominantly from the Netherlands Food and Consumer Product Safety Authority (NVWA), and inspectors, mainly from the police force. User accounts are assigned and managed by an administrator, who is a member of the police department. While both user groups possess equivalent access rights, their objectives for utilizing the data differ.

Importantly, users do not have immediate access to all data; they must request specific data as needed. This process ensures focused and relevant data usage, aligning with individual user roles and project requirements.

In terms of data management, the platform employs a direct import approach from the source, eliminating the need for a separate data transformation process. The data, already cleaned and curated, is stored within the Public Service on the Map (PDOK) system. This strategy negates the requirement for a dedicated landing zone for data, as it is extracted directly from PDOK as and when required.

Regarding the data sources, they are classified into two categories: external and internal. External sources comprise data from entities outside the stakeholder group, providing a broader context and additional insights. Internal sources, on the other hand, include data provided by stakeholder parties such as the Dutch and Belgian police, and NVWA. This data is particularly valuable as it contains specific information about previous drug dumping sites and contaminated fields, offering critical inputs for the project's objectives.



2. API Structure

For the project, it is decided to work with the FastAPI framework. The framework enables fast development of web services like websites and API's. It also provides advanced data validation through the usage of Pydantic and is dependent on the Starlette web framework. The framework is written in Python which promotes fast development, although python is slower than some other compiled languages for the project the memory safety and ease of use from Python are more beneficial.

The API utilizes various fields from PDOK, which stands for Public Service on the Map. This Dutch initiative provides access to a wide range of geospatial data and services for the Netherlands. The provided data is typically in a standardized format, making it easy to integrate and apply.

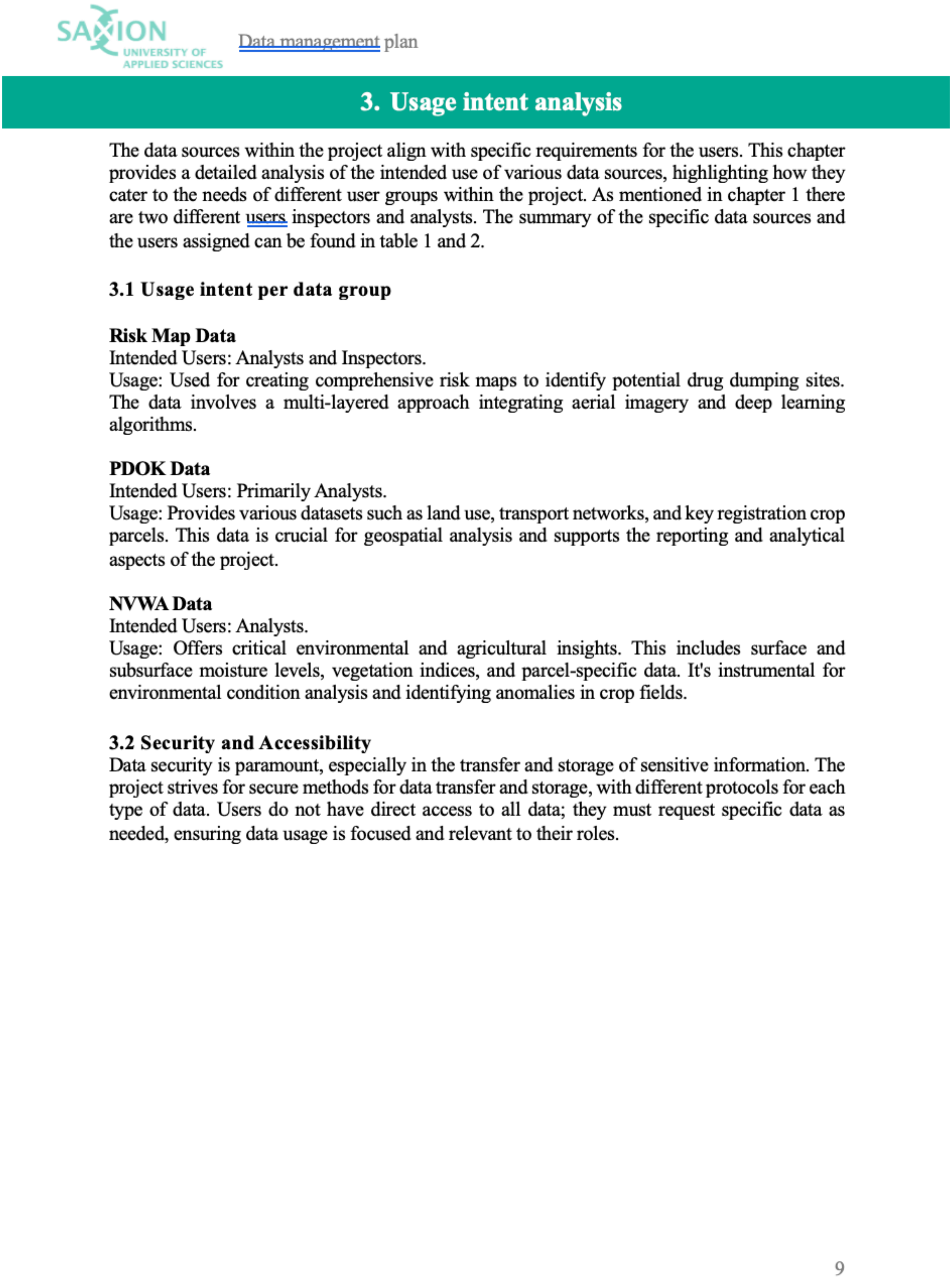
All of these fields have corresponding web services on pdok.nl, including WMS (Web Map Services), WFS (Web Feature Services), WCS (Web Coverage Services), and WMTS (Web Map Tile Services). It's important to note that if a field, for example, Land Use, has multiple web services, all of them are supported. These services are commonly used by GIS software, but they can also be accessed using a Python library called owslib. This library is designed for extracting data from OGC (Open Geospatial Consortium) web services, as mentioned above.

Based on the various geospatial data used, there are four different endpoint types in the endpoint map. For a detailed explanation of the folder structure, please refer to Appendix A.

Figure 1 depicts a simplified version of the process

Figure 5 Data extraction process

This diagram illustrates the process of retrieving land data. The API directs the inquiry to the extractor class, which then requests the data from PDOK. After PDOK responds, owslib is used to extract the specific data, and the API sends it back to the user. For further detail see appendix A.



3. Usage intent analysis

The data sources within the project align with specific requirements for the users. This chapter provides a detailed analysis of the intended use of various data sources, highlighting how they cater to the needs of different user groups within the project. As mentioned in chapter 1 there are two different users inspectors and analysts. The summary of the specific data sources and the users assigned can be found in table 1 and 2.

3.1 Usage intent per data group

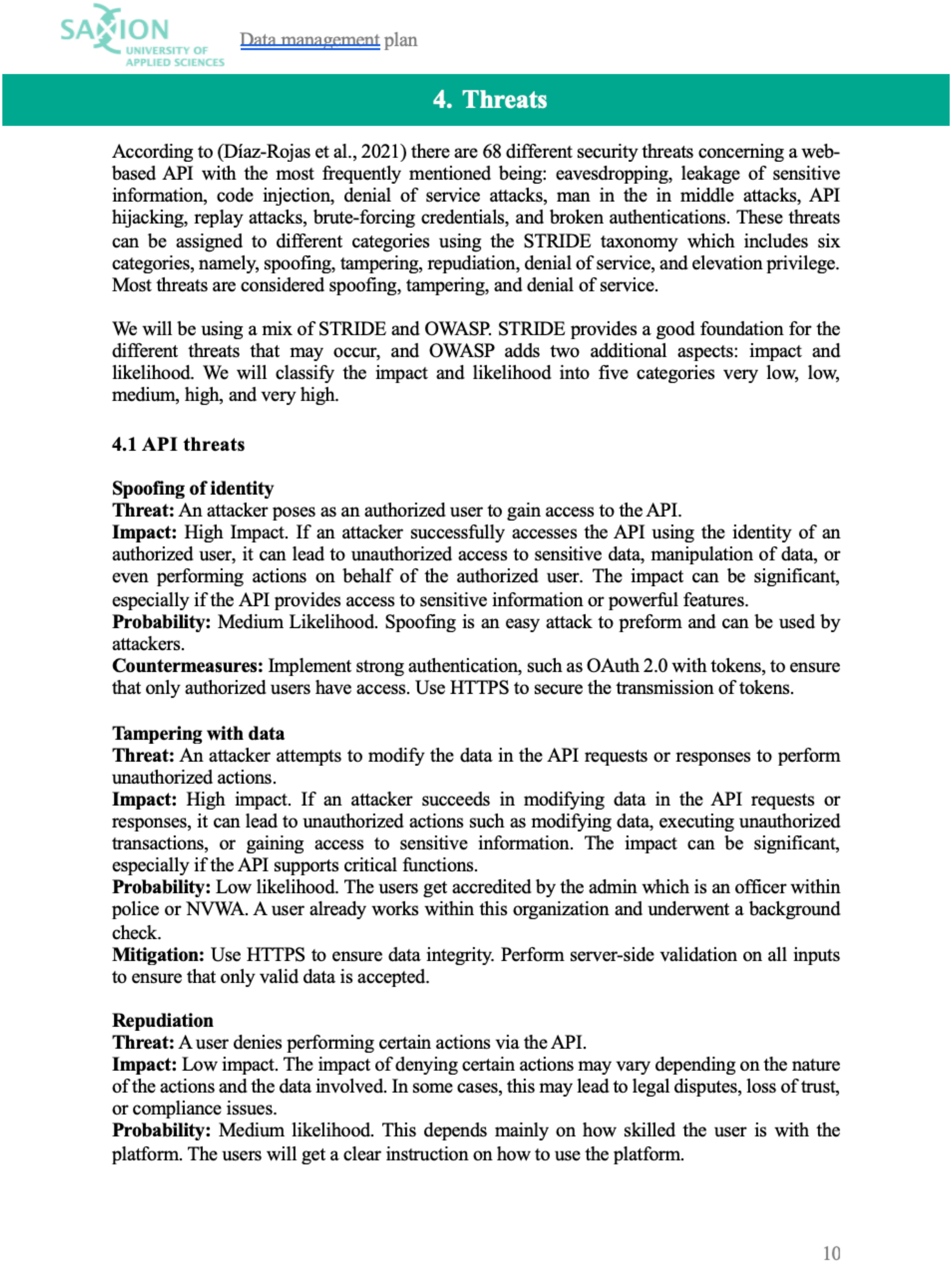
Risk Map Data
Intended Users: Analysts and Inspectors.
Usage: Used for creating comprehensive risk maps to identify potential drug dumping sites. The data involves a multi-layered approach integrating aerial imagery and deep learning algorithms.

PDOK Data
Intended Users: Primarily Analysts.
Usage: Provides various datasets such as land use, transport networks, and key registration crop parcels. This data is crucial for geospatial analysis and supports the reporting and analytical aspects of the project.

NVWA Data
Intended Users: Analysts.
Usage: Offers critical environmental and agricultural insights. This includes surface and subsurface moisture levels, vegetation indices, and parcel-specific data. It's instrumental for environmental condition analysis and identifying anomalies in crop fields.

3.2 Security and Accessibility

Data security is paramount, especially in the transfer and storage of sensitive information. The project strives for secure methods for data transfer and storage, with different protocols for each type of data. Users do not have direct access to all data; they must request specific data as needed, ensuring data usage is focused and relevant to their roles.



4. Threats

According to (Díaz-Rojas et al., 2021) there are 68 different security threats concerning a web-based API with the most frequently mentioned being: eavesdropping, leakage of sensitive information, code injection, denial of service attacks, man in the in middle attacks, API hijacking, replay attacks, brute-forcing credentials, and broken authentications. These threats can be assigned to different categories using the STRIDE taxonomy which includes six categories, namely, spoofing, tampering, repudiation, denial of service, and elevation privilege. Most threats are considered spoofing, tampering, and denial of service.

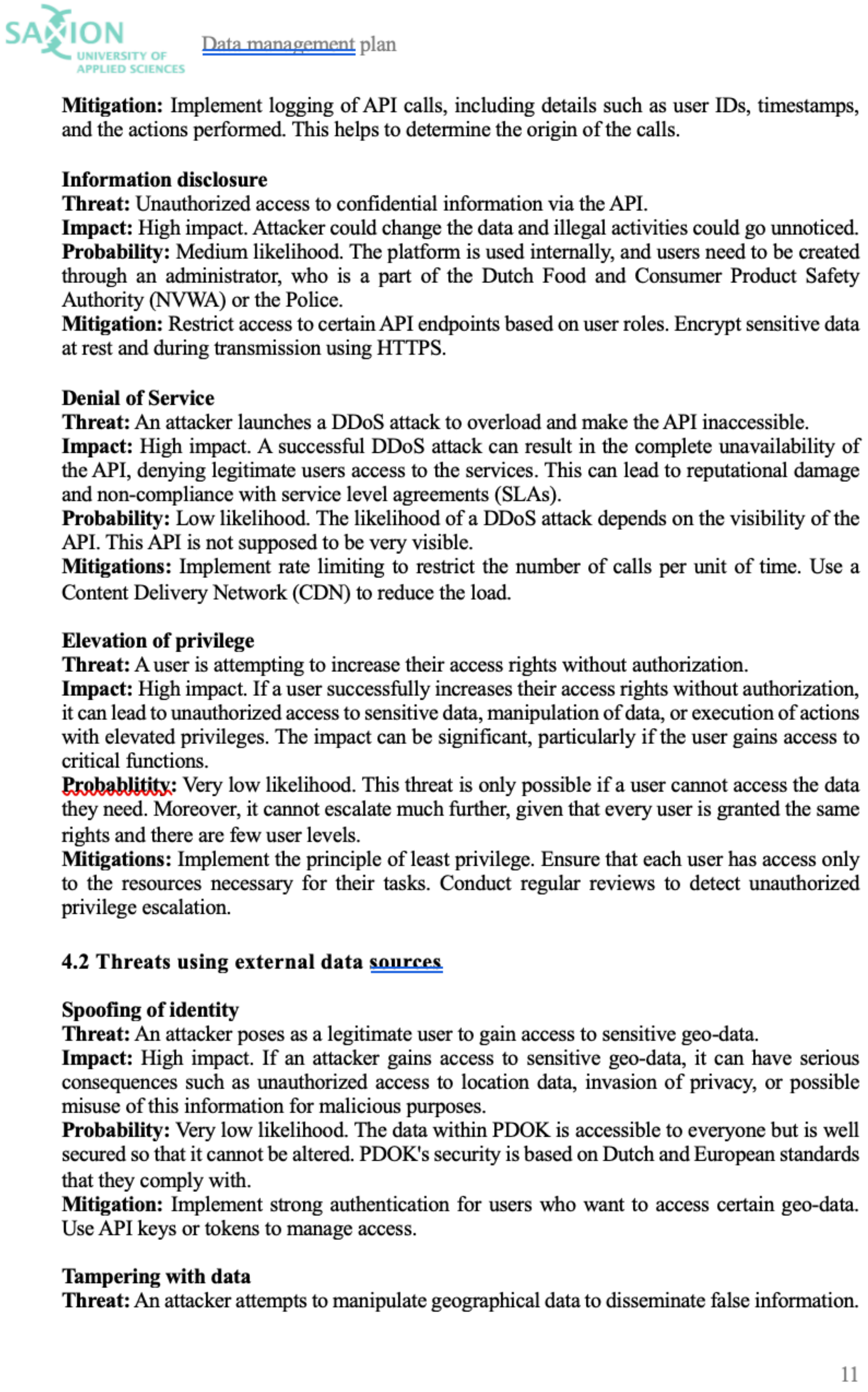
We will be using a mix of STRIDE and OWASP. STRIDE provides a good foundation for the different threats that may occur, and OWASP adds two additional aspects: impact and likelihood. We will classify the impact and likelihood into five categories very low, low, medium, high, and very high.

4.1 API threats

Spoofing of identity
Threat: An attacker poses as an authorized user to gain access to the API.
Impact: High Impact. If an attacker successfully accesses the API using the identity of an authorized user, it can lead to unauthorized access to sensitive data, manipulation of data, or even performing actions on behalf of the authorized user. The impact can be significant, especially if the API provides access to sensitive information or powerful features.
Probability: Medium Likelihood. Spoofing is an easy attack to preform and can be used by attackers.
Countermeasures: Implement strong authentication, such as OAuth 2.0 with tokens, to ensure that only authorized users have access. Use HTTPS to secure the transmission of tokens.

Tampering with data
Threat: An attacker attempts to modify the data in the API requests or responses to perform unauthorized actions.
Impact: High impact. If an attacker succeeds in modifying data in the API requests or responses, it can lead to unauthorized actions such as modifying data, executing unauthorized transactions, or gaining access to sensitive information. The impact can be significant, especially if the API supports critical functions.
Probability: Low likelihood. The users get accredited by the admin which is an officer within police or NVWA. A user already works within this organization and underwent a background check.
Mitigation: Use HTTPS to ensure data integrity. Perform server-side validation on all inputs to ensure that only valid data is accepted.

Repudiation
Threat: A user denies performing certain actions via the API.
Impact: Low impact. The impact of denying certain actions may vary depending on the nature of the actions and the data involved. In some cases, this may lead to legal disputes, loss of trust, or compliance issues.
Probability: Medium likelihood. This depends mainly on how skilled the user is with the platform. The users will get a clear instruction on how to use the platform.



Mitigation: Implement logging of API calls, including details such as user IDs, timestamps, and the actions performed. This helps to determine the origin of the calls.

Information disclosure
Threat: Unauthorized access to confidential information via the API.
Impact: High impact. Attacker could change the data and illegal activities could go unnoticed.
Probability: Medium likelihood. The platform is used internally, and users need to be created through an administrator, who is a part of the Dutch Food and Consumer Product Safety Authority (NVWA) or the Police.
Mitigation: Restrict access to certain API endpoints based on user roles. Encrypt sensitive data at rest and during transmission using HTTPS.

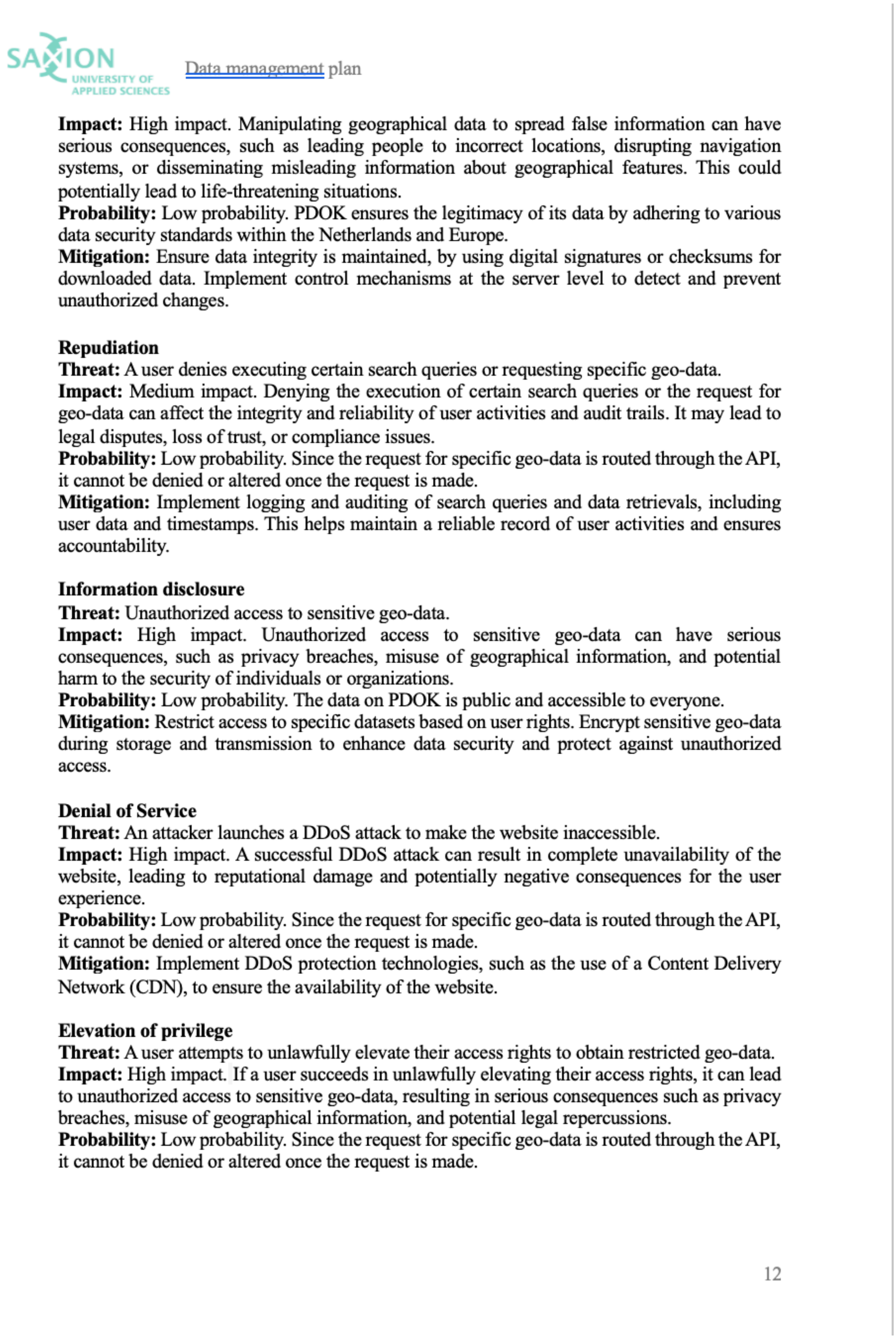
Denial of Service
Threat: An attacker launches a DDoS attack to overload and make the API inaccessible.
Impact: High impact. A successful DDoS attack can result in the complete unavailability of the API, denying legitimate users access to the services. This can lead to reputational damage and non-compliance with service level agreements (SLAs).
Probability: Low likelihood. The likelihood of a DDoS attack depends on the visibility of the API. This API is not supposed to be very visible.
Mitigations: Implement rate limiting to restrict the number of calls per unit of time. Use a Content Delivery Network (CDN) to reduce the load.

Elevation of privilege
Threat: A user is attempting to increase their access rights without authorization.
Impact: High impact. If a user successfully increases their access rights without authorization, it can lead to unauthorized access to sensitive data, manipulation of data, or execution of actions with elevated privileges. The impact can be significant, particularly if the user gains access to critical functions.
Probablitity: Very low likelihood. This threat is only possible if a user cannot access the data they need. Moreover, it cannot escalate much further, given that every user is granted the same rights and there are few user levels.
Mitigations: Implement the principle of least privilege. Ensure that each user has access only to the resources necessary for their tasks. Conduct regular reviews to detect unauthorized privilege escalation.

4.2 Threats using external data sources

Spoofing of identity
Threat: An attacker poses as a legitimate user to gain access to sensitive geo-data.
Impact: High impact. If an attacker gains access to sensitive geo-data, it can have serious consequences such as unauthorized access to location data, invasion of privacy, or possible misuse of this information for malicious purposes.
Probability: Very low likelihood. The data within PDOK is accessible to everyone but is well secured so that it cannot be altered. PDOK's security is based on Dutch and European standards that they comply with.
Mitigation: Implement strong authentication for users who want to access certain geo-data. Use API keys or tokens to manage access.

Tampering with data
Threat: An attacker attempts to manipulate geographical data to disseminate false information.



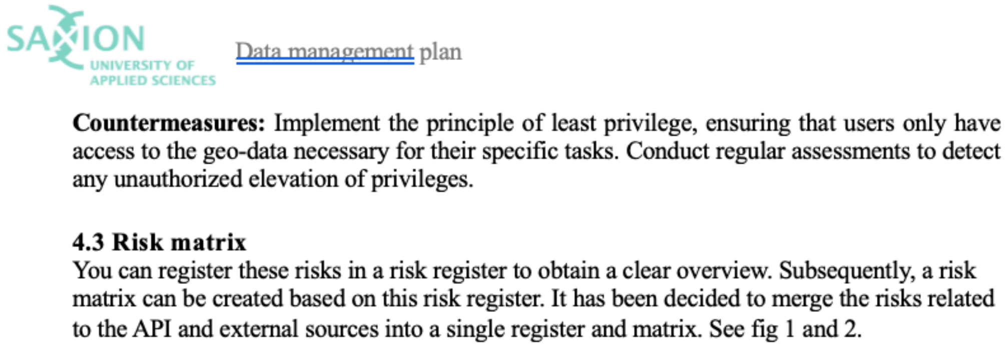
Impact: High impact. Manipulating geographical data to spread false information can have serious consequences, such as leading people to incorrect locations, disrupting navigation systems, or disseminating misleading information about geographical features. This could potentially lead to life-threatening situations.
Probability: Low probability. PDOK ensures the legitimacy of its data by adhering to various data security standards within the Netherlands and Europe.
Mitigation: Ensure data integrity is maintained, by using digital signatures or checksums for downloaded data. Implement control mechanisms at the server level to detect and prevent unauthorized changes.

Repudiation
Threat: A user denies executing certain search queries or requesting specific geo-data.
Impact: Medium impact. Denying the execution of certain search queries or the request for geo-data can affect the integrity and reliability of user activities and audit trails. It may lead to legal disputes, loss of trust, or compliance issues.
Probability: Low probability. Since the request for specific geo-data is routed through the API, it cannot be denied or altered once the request is made.
Mitigation: Implement logging and auditing of search queries and data retrievals, including user data and timestamps. This helps maintain a reliable record of user activities and ensures accountability.

Information disclosure
Threat: Unauthorized access to sensitive geo-data.
Impact: High impact. Unauthorized access to sensitive geo-data can have serious consequences, such as privacy breaches, misuse of geographical information, and potential harm to the security of individuals or organizations.
Probability: Low probability. The data on PDOK is public and accessible to everyone.
Mitigation: Restrict access to specific datasets based on user rights. Encrypt sensitive geo-data during storage and transmission to enhance data security and protect against unauthorized access.

Denial of Service
Threat: An attacker launches a DDoS attack to make the website inaccessible.
Impact: High impact. A successful DDoS attack can result in complete unavailability of the website, leading to reputational damage and potentially negative consequences for the user experience.
Probability: Low probability. Since the request for specific geo-data is routed through the API, it cannot be denied or altered once the request is made.
Mitigation: Implement DDoS protection technologies, such as the use of a Content Delivery Network (CDN), to ensure the availability of the website.

Elevation of privilege
Threat: A user attempts to unlawfully elevate their access rights to obtain restricted geo-data.
Impact: High impact. If a user succeeds in unlawfully elevating their access rights, it can lead to unauthorized access to sensitive geo-data, resulting in serious consequences such as privacy breaches, misuse of geographical information, and potential legal repercussions.
Probability: Low probability. Since the request for specific geo-data is routed through the API, it cannot be denied or altered once the request is made.



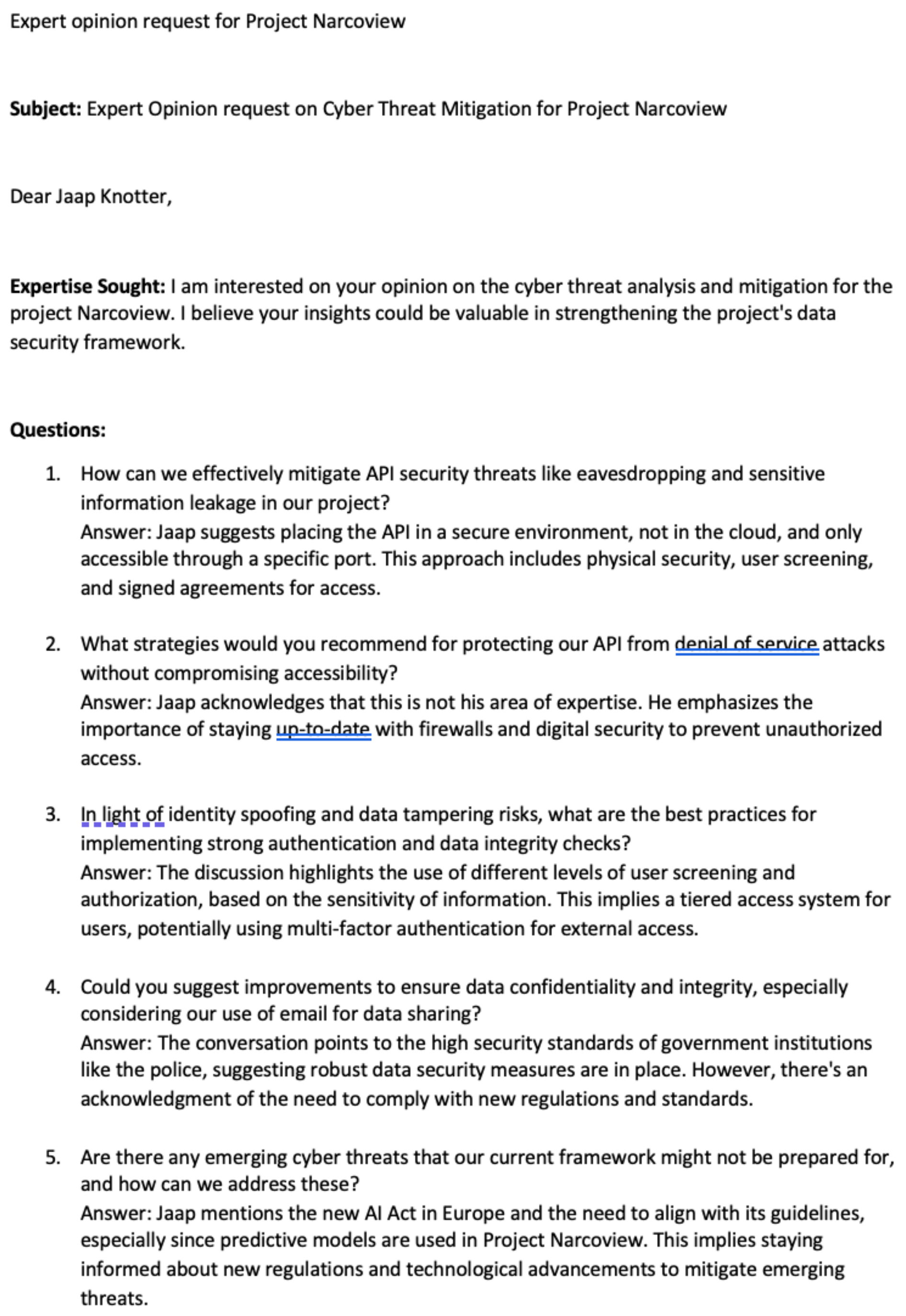
Data management plan

Countermeasures: Implement the principle of least privilege, ensuring that users only have access to the geo-data necessary for their specific tasks. Conduct regular assessments to detect any unauthorized elevation of privileges.

4.3 Risk matrix

You can register these risks in a risk register to obtain a clear overview. Subsequently, a risk matrix can be created based on this risk register. It has been decided to merge the risks related to the API and external sources into a single register and matrix. See fig 1 and 2.

Appendix C
Expert Opinion request on Cyber Threat Mitigation for Project NarcoView



Expert opinion request for Project Narcoview

Subject: Expert Opinion request on Cyber Threat Mitigation for Project Narcoview

Dear Jaap Knotter,

Expertise Sought: I am interested on your opinion on the cyber threat analysis and mitigation for the project Narcoview. I believe your insights could be valuable in strengthening the project's data security framework.

Questions:

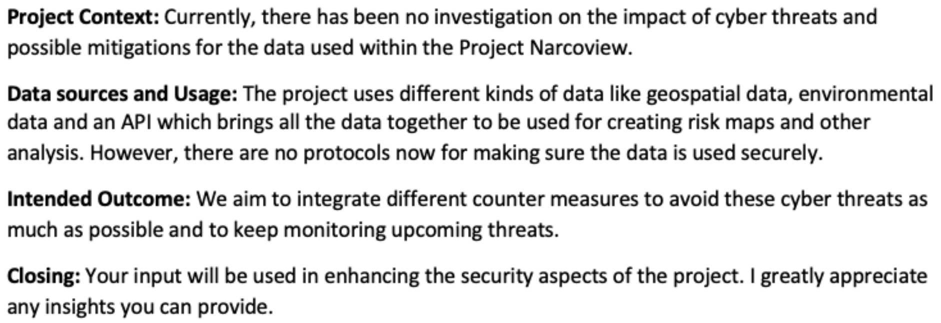
1. How can we effectively mitigate API security threats like eavesdropping and sensitive information leakage in our project?
   Answer: Jaap suggests placing the API in a secure environment, not in the cloud, and only accessible through a specific port. This approach includes physical security, user screening, and signed agreements for access.

2. What strategies would you recommend for protecting our API from denial of service attacks without compromising accessibility?
   Answer: Jaap acknowledges that this is not his area of expertise. He emphasizes the importance of staying up-to-date with firewalls and digital security to prevent unauthorized access.

3. In light of identity spoofing and data tampering risks, what are the best practices for implementing strong authentication and data integrity checks?
   Answer: The discussion highlights the use of different levels of user screening and authorization, based on the sensitivity of information. This implies a tiered access system for users, potentially using multi-factor authentication for external access.

4. Could you suggest improvements to ensure data confidentiality and integrity, especially considering our use of email for data sharing?
   Answer: The conversation points to the high security standards of government institutions like the police, suggesting robust data security measures are in place. However, there's an acknowledgment of the need to comply with new regulations and standards.

5. Are there any emerging cyber threats that our current framework might not be prepared for, and how can we address these?
   Answer: Jaap mentions the new AI Act in Europe and the need to align with its guidelines, especially since predictive models are used in Project Narcoview. This implies staying informed about new regulations and technological advancements to mitigate emerging threats.



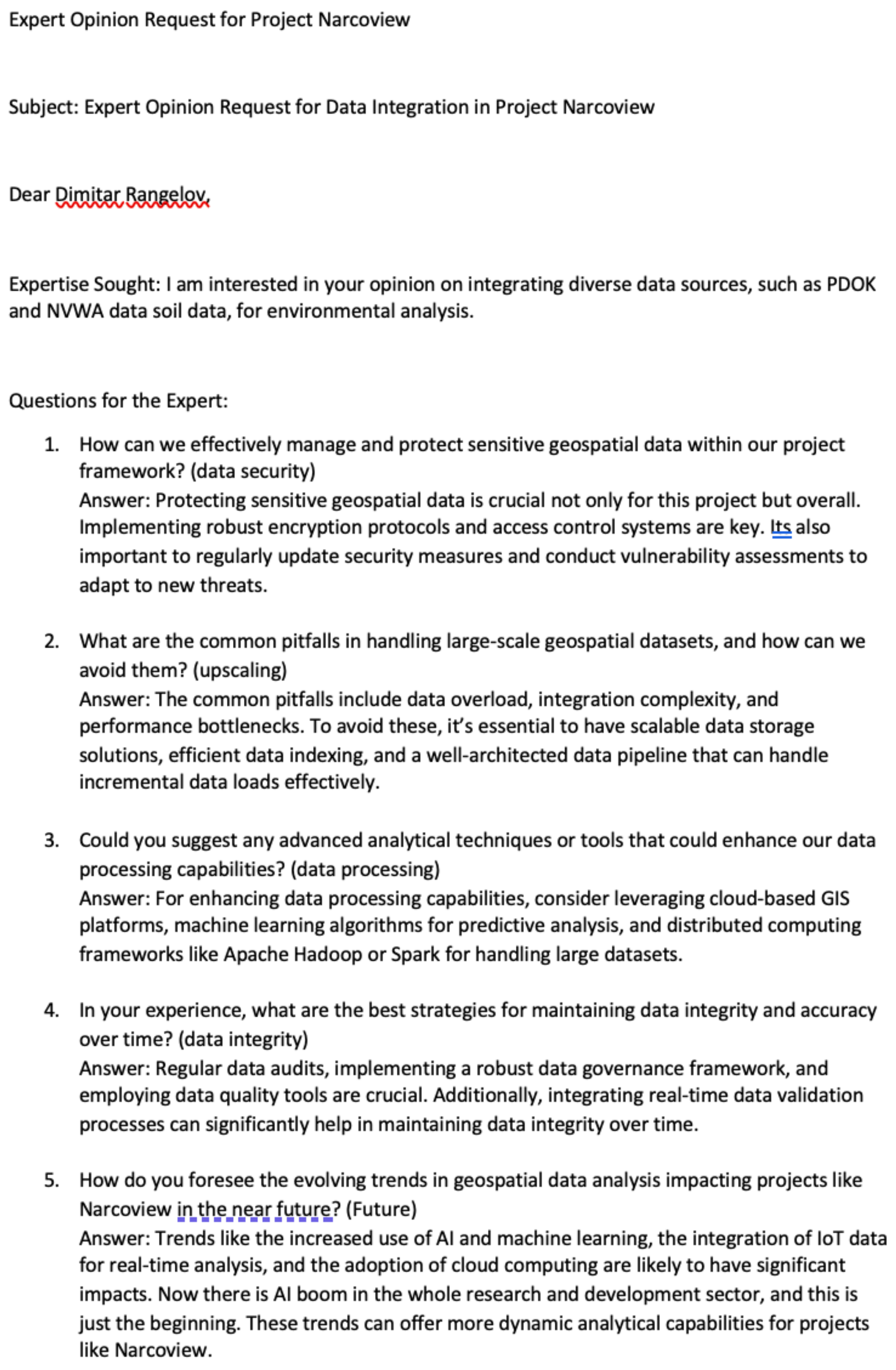
Project Context: Currently, there has been no investigation on the impact of cyber threats and possible mitigations for the data used within the Project Narcoview.

Data sources and Usage: The project uses different kinds of data like geospatial data, environmental data and an API which brings all the data together to be used for creating risk maps and other analysis. However, there are no protocols now for making sure the data is used securely.

Intended Outcome: We aim to integrate different counter measures to avoid these cyber threats as much as possible and to keep monitoring upcoming threats.

Closing: Your input will be used in enhancing the security aspects of the project. I greatly appreciate any insights you can provide.

Appendix D
Expert Opinion Request for Data Integration in Project NarcoView



Expert Opinion Request for Project Narcoview

Subject: Expert Opinion Request for Data Integration in Project Narcoview

Dear Dimitar Rangelov,

Expertise Sought: I am interested in your opinion on integrating diverse data sources, such as PDOK and NVWA data soil data, for environmental analysis.

Questions for the Expert:

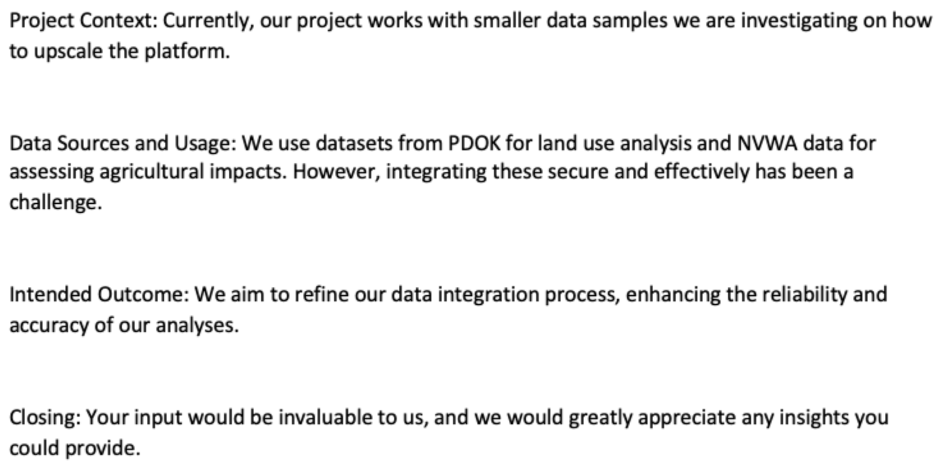
1. How can we effectively manage and protect sensitive geospatial data within our project framework? (data security)
   Answer: Protecting sensitive geospatial data is crucial not only for this project but overall. Implementing robust encryption protocols and access control systems are key. Its also important to regularly update security measures and conduct vulnerability assessments to adapt to new threats.

2. What are the common pitfalls in handling large-scale geospatial datasets, and how can we avoid them? (upscaling)
   Answer: The common pitfalls include data overload, integration complexity, and performance bottlenecks. To avoid these, it's essential to have scalable data storage solutions, efficient data indexing, and a well-architected data pipeline that can handle incremental data loads effectively.

3. Could you suggest any advanced analytical techniques or tools that could enhance our data processing capabilities? (data processing)
   Answer: For enhancing data processing capabilities, consider leveraging cloud-based GIS platforms, machine learning algorithms for predictive analysis, and distributed computing frameworks like Apache Hadoop or Spark for handling large datasets.

4. In your experience, what are the best strategies for maintaining data integrity and accuracy over time? (data integrity)
   Answer: Regular data audits, implementing a robust data governance framework, and employing data quality tools are crucial. Additionally, integrating real-time data validation processes can significantly help in maintaining data integrity over time.

5. How do you foresee the evolving trends in geospatial data analysis impacting projects like Narcoview in the near future? (Future)
   Answer: Trends like the increased use of AI and machine learning, the integration of IoT data for real-time analysis, and the adoption of cloud computing are likely to have significant impacts. Now there is AI boom in the whole research and development sector, and this is just the beginning. These trends can offer more dynamic analytical capabilities for projects like Narcoview.



Project Context: Currently, our project works with smaller data samples we are investigating on how to upscale the platform.

Data Sources and Usage: We use datasets from PDOK for land use analysis and NVWA data for assessing agricultural impacts. However, integrating these secure and effectively has been a challenge.

Intended Outcome: We aim to refine our data integration process, enhancing the reliability and accuracy of our analyses.

Closing: Your input would be invaluable to us, and we would greatly appreciate any insights you could provide.

Appendix E
Interview Report: Data Management and Security in Drone-Based Surveillance

Interview Participants:
Tatjana Kuznecova
Berit Basters

Date of Interview:
November 27, 2023

Key Discussion Points:

1. Data Labelling and Algorithm Performance:
Tatjana Kuznecova highlighted the importance of accurate labelling of containers in images to guide algorithms in identifying different types of objects. She noted that incorrect labelling or image replacement could compromise algorithm performance.

2. Data Transfer Protocols from Drones:
Berit Basters inquired about the protocols for transferring data from drones to the police database. Tatjana Kuznecova mentioned that there is no clear protocol currently in place. However, the NVWA (Netherlands Food and Consumer Product Safety Authority) has tested a proof-of-concept for automated data transfer to their database, involving semi-automated drone flights with data streaming into a database.

3. Security of Data Transfer:
The conversation touched upon the security aspects of data transfer. Berit Basters questioned if the data streaming involved a secure connection, to which Tatjana Kuznecova admitted uncertainty, suggesting this as a point of inquiry with NVWA.

4. Data Sharing and Exchange Infrastructure:
Tatjana Kuznecova commented on the current state of data sharing and exchange infrastructures in the Netherlands, particularly in law enforcement. She noted that these infrastructures are not well established yet but are planned for future development.

5. Data Storage and Security Concerns in Different Contexts:
Berit Basters referenced scenarios in war zones where data security from drone scans is critical, emphasizing the dilemma between network transmission and on-drone storage. Tatjana Kuznecova acknowledged that military protocols tend to be stricter and are still under evaluation for optimal data security. Berit Basters suggested that for their context, storing data on the drone might be more secure than transferring it over a network.

6. Legal Framework and International Practices:
The discussion briefly touched on legal aspects, with Berit Basters inquiring about laws in Belgium regarding data sharing and exchange. Tatjana Kuznecova directed to consult with data officers in Belgium and referenced a document on the EU legal framework.

7. Data Access and Privacy Concerns:

Tatjana Kuznecova advised Berit Basters to request anonymized data sets from Rowena for NVWA crop field data. She stressed the importance of not sharing the data publicly and suggested creating 'fake' examples for illustration in reports.

Summary:
The interview provided insights into the challenges and considerations in managing and securing data in drone-based surveillance, especially in law enforcement contexts. Key topics included the importance of accurate data labelling, the need for secure and efficient data transfer protocols, and the evolving nature of data sharing infrastructures. The conversation also underscored the importance of adhering to legal frameworks and maintaining data privacy in such initiatives.

Appendix F
Interview Report: NVWA Use Case Data Processing Workflow

Participants:
Rowena Emaus
Berit Basters
Nilay Swarge

Date of Interview: November 29, 2023

Summary of Discussion:

1. Data Processing and Analysis:
Rowena Emaus discussed the initial steps in the data processing workflow, including ID assignment and analysis of surface and subsurface moisture levels. Though not highly valuable, this data is included in the workflow.
Nilay Swarge provided insights into the use of vegetation indices like the NDVI (Normalized Difference Vegetation Index), explaining their derivation from remote sensing multispectral bands. The NDVI is used to assess vegetation health, with values close to 1 indicating healthy vegetation.

2. Data Collection Methods:
Nilay Swarge detailed the data collection process using Google Earth Engine and Sentinel-2 multispectral imagery. The focus was on parcels polluted with glyphosate, identified through parcel IDs provided by the NVWA.

3. Data Integrity and Challenges:
The discussion touched upon issues like cloud cover impacting NDVI values, necessitating the removal of negative values from the dataset.
Berit Basters raised concerns about data corruption, to which Rowena Emaus responded that negative NDVI values (possible errors) are removed to maintain data integrity.

4. Data Transfer and Security:
Nilay Swarge mentioned difficulties in accessing a developed cloud system for data transfer, resulting in reliance on email for data sharing. This raised concerns about the security and efficiency of data transfer methods.

5. Drone Data Storage and Transfer:
There was a discussion about the methods of storing data collected via drone flights. For the NVWA use case, data is stored on the drone and then transferred to a laptop or machine post-flight. This method is considered safer compared to network transmission, especially in sensitive environments.

6. External Data Access and Cooperation:
Berit Basters and Rowena Emaus discussed challenges in coordinating with external parties like the police and NVWA for data access and sharing. The conversation highlighted the difficulties in establishing effective communication and data exchange protocols with these entities.

7. Future Concerns and Steps:
The interview concluded with considerations for data security in the API development and the importance of securing data transfer methods. The discussion also touched upon the need for further coordination with external parties for comprehensive data access and integration into the project workflow.

Concluding Remarks:
The interview provided valuable insights into the complexities of data management, particularly in relation to environmental data collection, processing, and security. The challenges highlighted, such as securing data transfer and ensuring data integrity, underscore the importance of robust data management strategies in the NVWA use case.

Appendix G
Interview report: data users and usage

Participants:
Berit Basters
Dimitar Rangelov

Date of Interview: November 26, 2023

Summary of Discussion:

1. User Categories and Future Plans:
The platform currently lacks distinct user categories but plans to include two types of users in the future: inspectors and analysts. These users primarily come from the NVWA and police force.

2. Roles and Access:
User accounts, managed by an administrator from the police department, offer equal access rights to both user groups. However, access to data is controlled and requires specific requests, ensuring data usage aligns with individual roles and project needs.

3. Data Management:
The platform utilizes a direct import approach for data management, sourcing data from the Public Service on the Map (PDOK) system. This strategy avoids the need for a separate data transformation process, ensuring data is readily available and curated.

4. Data Sources:
The data sources are bifurcated into external and internal categories. External sources provide broader context and insights from outside the stakeholder group, while internal sources include specific information from stakeholder parties like the Dutch and Belgian police and NVWA. This data is crucial for the project, offering details about drug dumping sites and contaminated fields.

This interview sheds light on the operational aspects and future plans for the Project Narcoview platform, highlighting its user structure, data management, and sources.