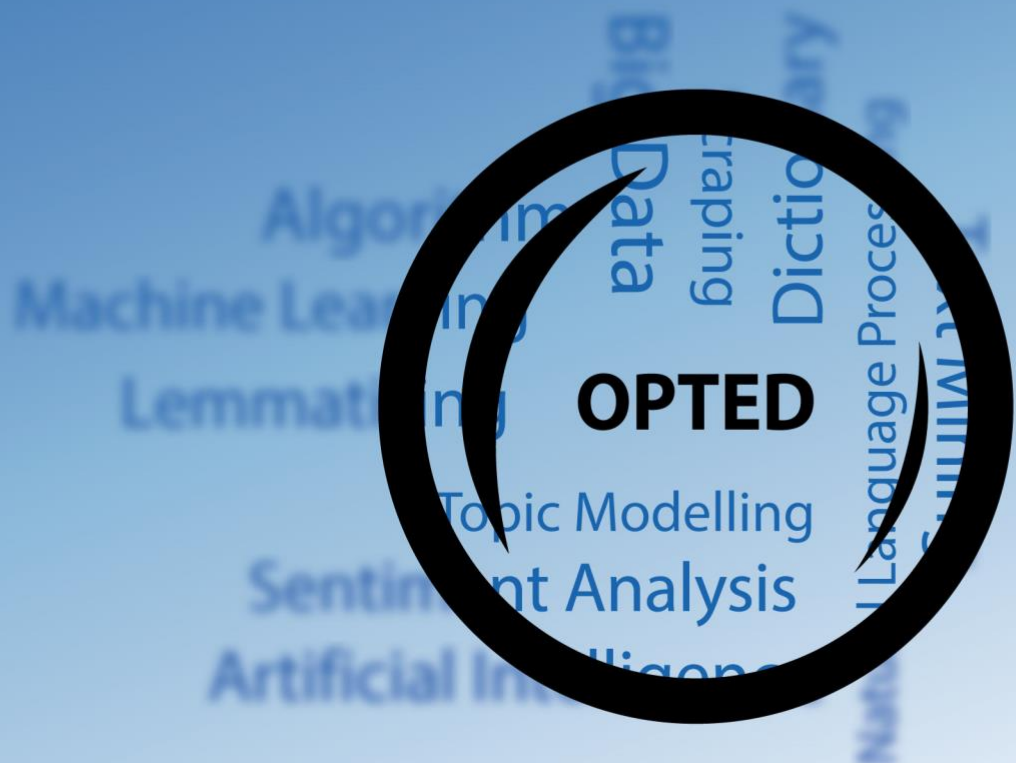


OPTED

ELSA by Design Questionnaire

Hripsime Asatryan, Bianca-loana Marcu, Rosamunde Elise van Brakel, & Paul de Hert



Disclaimer

This project has received funding from the European Union's Horizon 2020 research & innovation programme under grant agreement No 951832. The document reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained herein.

Dissemination level

Public

Type

Report



OPTED

Observatory for Political Texts in European Democracies:
A European research infrastructure

ELSA by Design Questionnaire

Deliverable D10.3

Authors: Hripsime Asatryan¹, Bianca-Ioana Marcu¹, Rosamunde Elise van Brakel¹, & Paul de Hert¹

¹ Vrije Universiteit Brussel

Due date: July 2021



1 Executive Summary

10.3 of the OPTED project is part of the Ethical and Legal Framework led by Work Package 10. The deliverable gathers input from the OPTED consortium members on how the legal, societal and ethical requirements laid out in the ELSA Framework which was presented in Deliverable 10.1, will be met throughout the different stages of the project. The deliverable will demonstrate the extent of awareness to potential social, ethical and legal concerns that may arise during the project. The results of this questionnaire will be used to provide a scope for the next deliverable of Work Package 10, Deliverable 10.4 ELSA Report on the validation of the ELSA requirements.

Consortium members are required to describe briefly the technology, infrastructure or techniques they are developing. Some of these techniques may have an impact on the right to personal data protection and privacy of the users. The questionnaire is designed to analyse how deep this impact is and which measures can be implemented to mitigate potential risks. Consortium members are requested to analyse the impact of the technological solutions embedded in their Work Package; how their adoption will change or improve their work; and which social, ethical or legal issues might arise from their perspective.

2 Introduction

This deliverable is part of Work Package 10 of the OPTED project. It is embedded in Task 10.2 ELSA by Design. ELSA by Design provides the ethical and legal framework to guide the development of new tools for political text analysis under the auspices of the OPTED project. It addresses the relevant societal, ethical and legal concerns stemming from the project through an impact assessment and, it is an expected continuation of the assessment process, by implementing the relevant safeguards in the system architecture.

2.1 Purpose of this deliverable

The purpose of this deliverable is to address the relevant societal concerns (relating, *inter alia*, to the right to the protection of personal data, right to privacy, ethics and societal acceptance – ELSA considerations) affected by the OPTED infrastructure. To ensure a wide range of commitment to non-technical requirements (i.e. privacy by design, ethics by design) relevant safeguards will be implemented in the conceptual design phase of the OPTED infrastructure. Privacy by design

Throughout this deliverable the ELSA-compliant design of the system will rely on the views and expertise of project partners.

The objective is to support the OPTED consortium partners to integrate considerations on data protection and ethics during project development and operational processes with its social impact, monitor the research activities, and ensuring that they comply with social, ethical and legal requirements.

The goal of this deliverable is to adopt measures that are oriented to prevent any deviation from the societal and ethical principles to which the project is bound, and compliant by design with the regulation. The General Data Protection Regulation¹ defines the approach of “privacy by design and by default” as part of Recital 78:

“When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations.”

2.2 Structure of the document

As for the structure of the questionnaire, each question is made up of three elements:

- The first part contains the question that consortium members shall answer;

¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)

- The second part contextualises the question by providing the reason behind asking the question;
- The third part is the box where consortium members fill in their answer.

In the left column, the questionnaire will specify whether ‘All Partners’ are expected to answer that question, or whether the question is intended for a specific Work Package. This is because some questions are more relevant for specific Work Packages, particularly those who engage in personal data processing activities.

3 Instructions to fill in questionnaire

Each partner needs to fill in one questionnaire.

Read all questions first to estimate the time you need to answer and identify the ones you do not clearly understand. Please, contact Work Package 10 in case of doubts.

Please send the filled in questionnaires by 30 June 2021.

We will review the answers and contact you in case there is any additional information or clarification to be provided.

Please respond as accurately and completely as possible. This questionnaire is an important part of the evaluation that we will perform over the technologies and infrastructure used throughout OPTED.

4 Question addressed to the Consortium Members

4.1 Brief Technical Description

| Relevant Partner | Required Input |
|---------------------|--|
| All partners | <i>Question:</i> Please provide a brief description of the element(s) of the OPTED infrastructure that you are developing. |
| | <i>Reason:</i> A brief description ensures that the answers further in the questionnaire are contextualised and precise. |
| | <i>Answer:</i> |

| Relevant Partner | Required Input |
|---------------------|--|
| All partners | <i>Question:</i> What is the need/relevance of this element within the OPTED project? |
| | <i>Reason:</i> A brief description of the relevance helps to clarify its application within the project. |
| | <i>Answer:</i> |

| Relevant Partner | Required Input |
|---------------------|---|
| All partners | <i>Question:</i> What is the intended outcome from the design and development of this particular element? |
| | <i>Reason:</i> A brief description of the objective and expected outcome of the element helps to assess its impact. |
| | <i>Answer:</i> |

4.2 Questions related to personal data protection

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> Does your WP require the processing of personal data? |
| | <i>Reason:</i> Personal data relates to any information that may directly or indirectly identify an individual. The processing of personal data in the EU requires compliance with the GDPR and the identification of a valid legal basis. |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> If the answer to the question above is YES, can you specify the types of personal data your WP processes? (For example: name, email address, location, gender, job title). |
| | <i>Reason:</i> It is necessary to identify which categories of personal data are processed as entirely as possible. |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> Does your WP require the processing of sensitive personal data? (For example: political opinions, philosophical beliefs, trade union memberships). |
| | <i>Reason:</i> Collecting and processing sensitive personal data requires additional safeguards as the risk for violating fundamental rights is higher. |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> For which purpose(s) do you process personal data? |
| | <i>Reason:</i> The processing of personal data requires the identification of a specific purpose and legal basis. |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> If you require consent from data subjects for personal data processing, how will you ensure that it is informed, specific and freely given? |
| | <i>Reason:</i> These are specific requirements under the GDPR to obtain valid consent. |

| | |
|--|----------------|
| | |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question: Where will the processed data be stored? Please specify if it requires transfer to a third country (non-EU) and specify which one.</i> |
| | <i>Reason: Under the GDPR, it is more secure for personal data to be stored in the EU and not to be transferred to third countries. If a transfer is necessary, the relevant compliance measures need to be taken, such as concluding data processing agreements.</i> |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question: Does the collected data meet the requirements of relevancy and accuracy?</i> |
| | <i>Reason: Under the GDPR, it is a requirement that only the personal data necessary for a specific purpose are collected, and that it is accurate. This point also relates to ethical requirements of the OPTED infrastructure, by ensuring that sources are accurately representative.</i> |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question: Is the end date of the processing set? For how long is personal data retained? What will happen to the personal data after project completion?</i> |
| | <i>Reason: The processing of personal data shall have an end date, taking into consideration rules on data minimisation, accuracy and relevance.</i> |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question: How is the processed data kept secure? Please explain as completely as possible.</i> |
| | <i>Reason: Data security is an important element of ensuring effective legal compliance with the GDPR. It also minimises potential risks to data subjects.</i> |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question: Will you employ data anonymisation or pseudonymisation techniques?</i> |

| | |
|--|---|
| | |
| | <i>Reason: Such techniques help to guarantee data subject rights and confidentiality. They also help keep any personal data secure by minimising the risks of personal data breaches and helps ensure the privacy of data subjects.</i> |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question: Do you require the use of third-party solutions or processors in order to process personal data? (e.g. survey platforms). If yes, where are they based and are they GDPR compliant?</i> |
| | <i>Reason: This allows us to identify whether personal data transfers take place, and whether additional safeguards are required. If third-party processors are required, it is necessary to review their own privacy and data protection practices to ensure they are satisfactory.</i> |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question: How will transparency with regards to the OPTED infrastructure be achieved?</i> |
| | <i>Reason: Transparency ensures that all data processing is clearly understandable to users and data subjects in a clear and simple format. This requirement is not only in line with the GDPR, but also with project goals.</i> |
| | Answer: |

4.2.1 Additional questions specific to the OPTED website

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Does the OPTED website provide a privacy policy? If yes,</i> |
| | <i>Reason: The privacy policy ensures that visitors/users of the website know whether any of their data is being processed and why.</i> |
| | Answer: |

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Can data subjects exercise their data subject rights? (Right to information, right to access, right to rectification, right to object, right to deletion).</i> |

| | |
|--|--|
| | <i>Reason: Under the GDPR, individuals must be informed of how they may exercise their rights if their personal data is being collected.</i> |
| | Answer: |

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Does the OPTED website collect email addresses for the purposes of a newsletter to keep interested parties updated on the project?</i> |
| | <i>Reason: This helps to identify which legal basis is necessary and which categories of personal data are processed.</i> |
| | Answer: |

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Does the OPTED website process metadata?</i> |
| | <i>Reason: This helps to identify the types of data processed when users visit the OPTED website and ensure an accurate privacy policy.</i> |
| | Answer: |

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Does the OPTED website use cookies?</i> |
| | <i>Reason: This helps to identify the types of data processed when users visit the OPTED website and ensure an accurate privacy policy.</i> |
| | Answer: |

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Will the OPTED website also be available on mobile devices?</i> |
| | <i>Reason: This helps to identify the types of data and metadata processed when users visit the OPTED website and thus ensure an accurate privacy policy.</i> |
| | Answer: |

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP1 | <i>Question: Will the OPTED website be developed with privacy by design and by default principles? How?</i> |

| | |
|--|--|
| | <i>Reason: This requires that due consideration is given to data protection during the development phase of the website. For example, the website should not process or collect more personal data than what is strictly necessary and must be secure.</i> |
| | Answer: |

4.2.2 Additional questions specific to the OPTED infrastructure

| Relevant Partner | Required Input |
|------------------|---|
| WP 9 | <i>Question: How will the OPTED infrastructure ensure data security?</i> |
| | <i>Reason: Overall data security measures are especially important within the OPTED infrastructure. Data security measures minimise the potential for data breaches and help to build overall trust in the project. Please be as specific and thorough as possible.</i> |
| | Answer: |

| Relevant Partner | Required Input |
|------------------|--|
| WP 9 | <i>Question: Will privacy by design and by default principles be embedded into the design of the OPTED infrastructure?</i> |
| | <i>Reason: Privacy by design and by default are important accountability measures which require baking data protection into the very structure of an infrastructure or technology.</i> |
| | Answer: |

| Relevant Partner | Required Input |
|------------------|--|
| WP9 | <i>Question: Where, geographically, will the OPTED infrastructure be developed and stored?</i> |
| | <i>Reason: This helps to identify whether data transfers are necessary as part of the development or data storage plans of the OPTED infrastructure.</i> |
| | Answer: |

| Relevant Partner | Required Input |
|------------------|--|
| WP9 | <i>Question: How will transparency with regards to the OPTED infrastructure be achieved?</i> |
| | <i>Reason: Transparency ensures that all data processing is clearly understandable to users and data subjects in a clear and simple format. This requirement is not only in line with the GDPR, but also with project goals.</i> |

| | |
|--|----------------|
| | <i>Answer:</i> |
|--|----------------|

| | |
|------------------|---|
| Relevant Partner | Required Input |
| WP9 | <i>Question:</i> Does the development of the OPTED infrastructure require the use of third-party developers or processors? If yes, where are they located? |
| | <i>Reason:</i> This helps to identify whether any of the data held in the OPTED infrastructure will be available to parties outside of the Consortium or requires transfers to third countries. |
| | <i>Answer:</i> |

| | |
|------------------|--|
| Relevant Partner | Required Input |
| WP9 | <i>Question:</i> If the answer to the question above is YES, are third-party processors compliant with the GDPR? |
| | <i>Reason:</i> Under the accountability principle, data controllers are required to ensure that processors they engage respect the GDPR. |
| | <i>Answer:</i> |

4.3 Questions related to ethics, bias and discrimination

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question:</i> Do elements of your WP require the design of your own algorithms? |
| | <i>Reason:</i> Algorithms used for Big Data processing are to a certain extent biased and can lead to risks of discrimination. |
| | <i>Answer:</i> |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question:</i> Do elements of your WP require the use of algorithms? |
| | <i>Reason:</i> Algorithms used for Big Data processing are to a certain extent biased and can lead to risks of discrimination. Types of algorithms could include linear regression algorithms, K-Means clustering algorithm, dynamic programming algorithm, greedy algorithm, brute force algorithm. |
| | <i>Answer:</i> |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question:</i> If the answer to the question above is YES, please specify how these were chosen, taking into consideration principles of representation, fairness, and possible risks of bias and discrimination. And more importantly, what steps are you going to undertake to mitigate the risks of bias? |

| | |
|--|--|
| | <i>Reason: Algorithms used for Big Data processing are to a certain extent biased and can lead to risks of discrimination. It is important to identify which have been chosen and why.</i> |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All Partners | <i>Question: Do you use machine learning / artificial intelligence? If so, can you explain the criteria behind the selection of specific models? How have these been selected and why?</i> |
| | <i>Reason: The objective of this question is to assess whether the algorithms behind the technology that the OPTED consortium partners use can result biased implications on the data provided by the users. So, while answering this question, consider if potentially risks of ethical implications on data subjects or users by the technology you are going to create/use within the project.</i> |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All Partners | <i>Question: Following on from the question above, how are you going to ensure 'explainability' of automated decisions or complex techniques and results? How can their functioning be as clear as possible for stakeholders?</i> |
| | <i>Reason: If complex models or automated decisions are used, it is necessary to be transparent with regards to how they work and how they are employed. This is not only an ethical requirement, but will also inform the wider academic community that wishes to engage with the OPTED project and its results. This questions applies both to techniques built by OPTED partners and ready-made techniques that OPTED partners will use.</i> |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All Partners | <i>Question: How are you going to ensure that data sources are representative of the broad spectrum of political opinions, dialogues, legislation, discourses, speeches, etc.?</i> |
| | <i>Reason: Representation within the political sciences is an important ethical requirement. Political text analysis may influence political decision-making and research, with broad consequences for society, thus requiring adequate representation of a variety of voices and stakeholders.</i> |
| | Answer: |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All Partners | <i>Question: Will certain political texts, opinions, legislation or research be excluded from the scope? If yes, why?</i> |

| | |
|--|--|
| | <i>Reason: Being able to justify why certain sources are excluded (or included) adds an additional element of transparency to political text analysis, as well as accountability for its contribution(s) to future political decision-making and research.</i> |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question: How are you going to ensure that data sources are sufficiently representative of a diverse research community?</i> |
| | <i>Reason: For example, who will be invited to participate in the surveys which are part of the OPTED project, and how can diversity be ensured?</i> |
| | Answer: |

| | |
|------------------|--|
| Relevant Partner | Required Input |
| WP2 | <i>Question: Which scientific journals will be chosen to create the inventories? What is the criteria for their selection?</i> |
| | <i>Reason: This helps clarify whether the inventories will represent a diverse range of opinions, research and contributions, and thus minimise potential risks of bias or discrimination.</i> |
| | Answer: |

4.4 Questions related to Copyright

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question: Will your WP employ text and data mining techniques? If yes, for which purpose(s)?</i> |
| | <i>Reason: Text and data mining (TDM) techniques hold specific implications for the EU Copyright Directive. TDM is not always freely allowed in the EU due to copyright laws. However, the EU Copyright Directive provides an exception for TDM used for scientific research purposes, which facilitates research using TDM techniques.</i> |
| | Answer: |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question: Will the accessed journals and articles be open access or behind a paywall?</i> |
| | <i>Reason: This helps identify whether and how certain texts can be made available through the created inventories. (This relates both to the journals where OPTED results are published and the prototypes that OPTED partners will build).</i> |

| | |
|--|---------------|
| | |
| | <i>Answer</i> |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> Will any of the accessed journals and articles be under the protection of author rights? If yes, how will permission be ensured? Does the journal in question have a specific policy with regards to scientific research use? |
| | <i>Reason: Compliance with the policy of the journal(s) is required.</i> |
| | <i>Answer</i> |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question:</i> For journals and articles that are behind a paywall, how will their inclusion in an inventory be justified? |
| | <i>Reason: This helps identify whether and how certain texts can be made available through the curated inventories. This may depend on the policy of the journal(s) in question.</i> |
| | <i>Answer</i> |

| | |
|---------------------|--|
| Relevant Partner | Required Input |
| All Partners | <i>Question:</i> Are extracted texts (e.g. journal articles, academic and media political texts, legislative speeches, newspaper articles) already included in a database? |
| | <i>Reason: This helps identify whether and how certain texts can be made available through the curated inventories. This may depend on the policy of the journal(s) in question.</i> |
| | <i>Answer</i> |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All Partners | <i>Question:</i> How will an appropriate level of security be ensured for extracted texts (e.g. journal articles, academic and media political texts, legislative speeches, newspaper articles) on which text and data mining techniques are applied? |
| | <i>Reason: Ensuring that works are stored with a high level of security is an integral component of the 'research exemption' for text and data mining under the EU Copyright Directive.</i> |
| | <i>Answer</i> |

| | |
|---------------------|---|
| Relevant Partner | Required Input |
| All partners | <i>Question:</i> For how long will extracted works (e.g. journal articles, academic and media political texts, legislative speeches, newspaper articles) be retained? |

| | |
|--|---|
| | <i>Reason: The retention of these works may be required for the verification of research results, for which a retention period needs to be set.</i> |
| | Answer |

| Relevant Partner | Required Input |
|---------------------|---|
| All partners | <i>Question: How will the integrity of the extracted works (e.g. journal articles, academic and media political texts, legislative speeches, newspaper articles) be ensured?</i> |
| | <i>Reason: Although there are not yet any unified guidelines on ensuring integrity and security of the works, scientific researchers shall begin to create standards which may be useful and shared with the wider community.</i> |
| | Answer: |

| Relevant Partner | Required Input |
|---------------------|--|
| All partners | <i>Question: Will your WP employ text and data mining techniques on social media data (e.g., Twitter, Reddit) or messenger data (e.g., WhatsApp, Telegram). If yes, how will you follow the policies of the respective websites for text and data mining for scientific research purposes?</i> |
| | <i>Reason: Social media platforms often allow text and data mining for scientific research, but only in accordance with their policy. Compliance with both their policy and with the EU Copyright Directive is therefore necessary.</i> |
| | Answer |

| Relevant Partner | Required Input |
|---------------------|--|
| All partners | <i>Question: Will text and data mining on social media data (e.g. Twitter, Reddit) or messenger data (e.g., WhatsApp, Telegram) include the processing of personal data?</i> |
| | <i>Reason: This helps determine whether the GDPR will also apply in addition to the EU Copyright Directive.</i> |
| | Answer |