Proposal Acronym:
DATIS

Proposal Title:

Data for Inclusive Societies: Foes and Friends of Inclusiveness in contemporary Greece

D1.1: Data Management Plan First Version

June 2024

Introduction

This report provides an overview of the data management strategy of the project. Focusing, therefore, on the data management policy (data collection procedures (i) the methodology that will be applied in terms of data collection and quality assurance procedures, (ii) the usefulness of the data, and (iii) the procedures to be followed sharing, protection and retention of the data after the end of the project the project. This is the first version of the data management strategy of the Datis, with the final version to be submitted in the last month of the project (M24).

Data Collection and Quality Assurance Procedures

DATIS project includes data from the survey for the European Parliament 2024 (candidates' and voters' surveys) as well as data from the ISSP 2024 (International Social Survey Programme) related to citizens' attitudes towards Digital Societies and data from the pilot survey for inclusive societies, with the aim of implementing the main DATIS survey in 2025.

The data collected in the context of this project are related to the inclusion, exclusion of minority groups from society but also to the European Union and the degree of inequality due to the rise of the Internet and digital technologies. This data falls into two categories: a) online survey data (candidates, voters, experts), b) data from social media (META).

The surveys (European Parliament 2024, ISSP 2024, DATIS pilot survey) are conducted through "LimeSurvey", an open access software for online surveys. Survey data is also extracted from the same software. From the initial design of the surveys, each participant is given a specific "token", so as to ensure that the survey is completed by only one participant. Then, once the data has been collected, data cleaning methods are applied based on response quality indicators such as item-nonresponse, time spent on each response, etc.

The use of data

DATIS data will be useful, both to scientists and to society and business. More specifically, regarding candidate data from online surveys and data from social media, the combination of these two types of data can help scientists as well as companies engaged in questionnaire surveys to reduce the size of the questionnaires without lacking any information they want to collect. In fact, data from social media can be associated with online survey questions in some cases. Thus, by combining the two methods we can obtain all the necessary information while simultaneously minimizing the size of the questionnaires and, as a consequence, the time it will take to answer them.

Minimizing, therefore, the size of questionnaires and their response time we can get more and better quality data as the time it takes to answer a questionnaire is quite an important factor affecting the quality of data collected. Overall, all project data and its analysis will contribute to making it available valuable scientific knowledge about the two political phenomena under consideration in society.

The dissemination of scientific knowledge to society and in particular information related to contemporary political phenomena can contribute to a better representation of citizens, who, knowing better the positions of political parties and candidates on various political issues, can choose the party or the candidate whose perspectives' are converging their own. In addition, since the data will be freely accessible to scientists, it is also strengthened scientific research is encouraged. More specifically, DATIS data can be used by social scientists, by data scientists, by those -scientists and non-scientists- engaged in online research, etc.

General Data Protection Regulation (GDPR)

Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 establishes the new General Data Protection Regulation in the European Union. The new regulation establishes strict rules regarding the collection, processing and storage of personal data of European Union citizens. These include individuals, companies, public or non-governmental organizations. The General Data Protection Regulation (GDPR) was approved on 27 April 2016 and came into force on 25 May 2018 after a two-year transition period.

However, the legal interpretation of the term "personal data" is quite broad as it is data related to the individual and concerning information, both about their personal, professional and public life, which, if collected together, can lead to their identification. For example, such information is: name, home address, email address, Internet Protocol (IP) address etc.

According to the General Data Protection Regulation, data collection can only start after the respondents have clearly expressed their consent. The respondents must also be clearly informed about the data being collected and for what purpose. After being informed about the type and purpose of the data, the time of their retention and the way they can exercise their rights, they must expressly agree to the collection of the data. This is ensured through the consent the participants give at the beginning of each survey. Only after they give their consent, they can start answering the survey. In case they do not agree, the survey is terminated immediately and they are informed that they cannot complete the survey, due to their objection to the collection of their data.

The respondent has the right to withdraw consent at any time. Withdrawal of consent does not affect the lawfulness of processing that was based on consent prior to its withdrawal. In any case and at any time, the respondent has the right to access, correct and delete (right to be forgotten) personal data, to limit their processing, to oppose their processing and to their portability.

Furthermore, according to the provisions of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (Article 8, Paragraph 1) the processing of a child's personal data is lawful only if the child is at least 16 years old. If the child is under the age of 16, such processing is lawful only if and to the extent that such consent is given or approved by the person having parental responsibility.

When personal data is collected, the controller provides the data subject with the

following information: his/her identity and contact details of the data controller (and the personal data protection officer, if any), the purposes of the processing for which the personal data is intended and the way of exercising his rights.

Data Protection

The data that will be collected during the project takes two forms: a) publicly available data and b) personal data (data of a personal nature). Regarding the collection of personal data there will be full compliance with the General Data Protection Regulation (GDPR). More specifically, in all surveys that will be conducted by the project's research team - both online and over the phone – at the beginning of the surveys there will be a consent form, which will include information about the purpose of the research, ensuring anonymity and the rights of the participants, while informing them that their participation in the research is optional and they can interrupt it at any stage of conducting the investigation.

In addition, in surveys addressed to voters, there is a question that appears after the respondents have given their consent for recording, storing and statistical processing of their responses, asking respondents if they are under 16 years old. If the respondents answer that they are under 16 years old, they are automatically excluded from the survey as according to Article 8(1) of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 he cannot participate in investigations without the consent of a guardian.

In terms of ensuring anonymity, voter survey data cannot easily lead to the identification of participants, as the technique of generalization is adopted. Additionally, since the invitations to participate in the voter survey are sent to mobile phone numbers, it is necessary to clarify that the mobile phone numbers are randomly generated by a computer algorithm (RDD – Random Digit Dialing) and therefore it is not possible to associate a number with its owner. Invitations to participate in the survey to randomly selected numbers are sent using the infrastructure of a private company specialized in sending text messages. It is worth noting that there will be no connection between the mobile phone numbers and participants' responses. Mobile phone numbers are kept separately from survey responses so that there is no link between survey responses and mobile phone numbers. The mobile phone numbers of the participants together with the link included in the invitation, will be on the server of the company that will be used to send the text messages and they will be deleted after the sending process is over. These are highly secure servers and the only way to extract information that could link a mobile phone number with a specific response in the survey is if both servers were hacked during the period of data collection. However, if any problem occurs on one of the two servers, the other will be automatically informed and stronger measures will be taken.

In addition, the voter survey meets the requirements of the current AUTH Code of Ethics and all participants in the surveys are informed about surveys' purposes. Sample members are invited to a website, where they can find information about the personal data protection policy, the rights of the participants and how to exercise them, as well as the contact information of the data controller and the scientific manager of the project.

Ethical Issues META and Twitter (X)

In the context of this project, all the necessary measures have been taken to ensure the anonymity of the research participants and the protection, in accordance with the GDPR (General Data Protection Regulation), of their sensitive personal data. In particular, with regard to ensuring anonymity in the specific research that includes the collection of data from social media, which concern publications (posts) of European Elections 2024 candidates and parties on pages or groups on Facebook, Instagram or/and Twitter (X), is detailed in article 28 of Law No. 4624/2019 regarding the reconciliation of the right to the protection of personal data with the right to freedom of expression and information, including processing for journalistic purposes and for purposes of academic, artistic or literary expression.

According to this law, the processing of personal data is permitted, when the right to freedom of expression and the right to information prevail over the right to protect the subject's personal data, in particular for matters of general interest or when it concerns the personal data of public figures. However, in cases where it is easy to identify the candidate for European Parliament, it is possible to hide answers that would help to identify him. After all, the data from social media, which are posts and information about the posts (e.g. number of likes, etc.) are open to the general public and therefore cannot be considered personal data, as long as they are published by the candidates.

Data Storage and Retention

All the data of this project will be stored in an encrypted form on the server of the scientific manager and will be kept for at least five years after the end of this projects. At the same time, the data will be published on the project website (http://datis.gr/) where they will remain available for the same period of time. Survey data will be stored in online open access repositories such as Zenodo.

Data Sharing

All DATIS data is open access data. Online survey data will be made available in data repositories (e.g. Zenodo: https://zenodo.org/). These are open access platforms, where registration is required, in order to download the available data. In particular, with regard to Zenodo, the person interested in downloading the data is asked to accept the conditions set by the platform.