GDPoweR – Recovering workers' data to negotiate and monitor collective agreements in the platform economy

Research Design Addendum II - Data Management Plan

Leonard Geyer

April 5, 2024

Abstract: This Data Management Plan (DMP) describes what data will be collected in the research project GDPoweR and how the data must be handled to make it FAIR and to ensure data safety. The DMP complements the GDPoweR Research Design, the Data Recovery Protocol and the Research Design Addendum I.

April, 2024

For more information geyer@euro.centre.org

Information may be quoted provided the source is stated accurately and clearly.
This publication is also available via https://www.euro.centre.org/projects/detail/4722

Contents

1. Introduction

GDPoweR is a collaborative research project on industrial relations in the platform economy involving seven research and social partner organizations from five EU countries (Austria, Belgium, France, Poland, Spain). It focuses on two location-based platform sectors, ride-hailing and food delivery, and explores three thematic areas. The first area is the collection and use of worker data by digital labour platforms and the impact companies' collection and use of worker data has on worker well-being and their inclination to engage in collective action. The second area of interest is strategies employed by social partners to negotiate and implement collective and company-level agreements in the platform economy. The third and last area relates to the implementation, monitoring, and enforcement of such negotiated agreements.

The project foresees the collection of different types of primary data. The purpose of this Data Management Plan (DMP) is to describe what data will be collected, how the data should be handled in line with the FAIR (findable, accessible, interoperable, and re-usable) principle and how to ensure data safety. The plan is based on a template for Horizon 2020 projects.[1] It complements and should be read in conjunction with the GDPoweR Research Design, the Data Recovery Protocol and the Research Design Addendum I.

2. Data Summary

New data will be generated by the GDPoweR research through several methods.

Focus groups with platform workers as well as social partners and activists will be conducted. Conditional on the participants' consent, the focus groups will be recorded, and the audio recordings will be stored as MP3 files. The recordings will be transcribed, and the text documents will be saved electronically as .docx files.

Platform workers will be asked to donate personal data collected about them by digital labour platforms to project researchers after the workers received copies of this data from platform companies through Subject Access Requests under the European General Data Protection Regulation. It is unclear what data and in which formats will be donated. However, it is likely that much of the data will be in .CSV format.

Movement data of platform workers will be collected through tracking apps like the Vehicle Trip LogBook Tracker or the self- tracking app WeClock or through GPS-watches. The format of the collected data will depend on the technology used. Where possible, the .kml format will be used.

Technical audits of platforms' apps may be conducted. Such technical audits can reveal which cell-phone functions an app accessed and when. In addition, the audit can reveal with which servers the

---

[1] European Commission (2016). H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. V 3.0 Available at: https://Ec.Europa.Eu/Research/Participants/Data/Ref/H2020/Grants_Manual/Hi/Oa_Pilot/H2020-Hi-Oa-Data-Mgt_En.Pdf#Page=10

app communicates and what data is sent and received. The technical audit may produce data in various electronic formats.

Table 1 summarises the types of data the project will use and collect, the WP it will be used for and the format in which it will be stored.

#	Data type	Origin	WP#	Format
1	Recordings of focus groups	Primary data	3	.mp3
2	Transcripts of focus groups	Primary data	3	.docx
3	Copies of workers' personal data requested from digital labour platforms.	Primary data	3	Unknown, but likely .cvs
4	Movement data	Primary data	3	Preferably .kml
5	Technical audit results	Primary data	3	Various electronic formats

3. Data Manager

Each project partner nominates one person as Data Manager (see also GDPoweR Data Recovery Protocol). The Data Manager is responsible for all aspects regarding the use of data described in this DMP by individuals working for that project partner. In particular, the data manager is responsible for informing all members of his/her organisation about the proper use of data and the enforcement of the rules described in this DMP.

4. FAIR data

This section describes how, and to what extent, primary data collected within the GDPoweR project will be made findable, accessible, interoperable, and re-usable (FAIR).

4.1 Making data findable, including provisions for metadata

4.1.1 Focus group data

The transcripts of the focus groups will be stored as docx-files. The name of each file includes the abbreviated project name, the country code, an abbreviation of the event (Focus group trade unions – FCTU; focus group employer groups – FCEG; focus group riders – FCR; focus group drivers –

FCD), and the date of the event (e.g. "GDPoweR_AT_FCTU_14.03.2024.docx"). Audio files referring to the same workshop will be stored as an MP3-file with the same name (e.g. "GDPoweR_AT_FCTU _14.03.2024.mp3"). In addition to the time and the respective cases the focus group refers to, the text files will include the following metadata:

- The place the focus group took place.

- The name(s) of the person(s) conducting the workshop/moderating the focus group.

- When the participants of the focus groups consent to being named as source: the name of the participant(s) and the organisation they represent.

- When focus group participants do not consent to being named as source: a unique identifier to protect the anonymity of the participants of the focus groups. The personal identifier should include the country code, the abbreviation of the event and a running number (e.g. "AT_FCTU_1" for the first person in the Austrian focus group with trade unions).

- For the focus group with workers, summary statistics on

    o the number of participants,
    o the number of male and female participants,
    o the number of participants aged 35 years or older,
    o the number of participants working for each of the platform companies covered by the study,
    o the number of participants who work as employees, self-employed and with other employment statuses,
    o the number of participants who are citizens of the country in which the focus group was conducted,
    o the number of participants who are trade union members.

For individuals who do not consent to being named as source, the person's full name together with his/her unique identifier must be stored in separate file in a different, secure location.

Audio files may be cut to remove irrelevant sequences (beginning, breaks, end). The final versions of recordings and transcripts should be stored with the suffix "_final".

4.1.2   Copies of workers' personal data recovered from platforms

Personal data recovered from platforms by platform workers and subsequently donated to project researchers must be safely stored under a pseudonym as described in the GDPoweR Data Recovery Protocol. All data belonging to one person should be stored in a separate folder. The folder should be named PD (Personal Data), followed by the country code, the workers job description (rider or driver) and a sequential number. For example, the data donated by the first rider in the Austrian case should be stored in a folder named "PD_AT_Rider_1". The name(s) of the file(s) containing the personal data should start with the same name as the folder.

The worker's full name and the corresponding pseudonym must be stored in separate file in a different, secure location. The data, the company from which it was recovered, its place of storage and any sharing of the data with other project researchers must be catalogued by the responsible data manager. An example of how to catalogue the data is provide in table 2 below.

Table 2: Cataloguing worker data

Data donor (pseudonym)	Country case	Platform company	Date received	Number of files	Personal identifiers deleted?	Place of storage	Data shared with (name, duration)	To be deleted by (date)
Leonard	AT	Lieferando	21.12.2023	7	Yes	C:/GDPoweR/ password protected folder/PD_AT_Rider_1	Elif Naz Kayran (31.12.2023 -)	September 2026

4.1.3 Movement data

Movement data collected with platform workers should be stored in a separate folder which should be named MD (Movement Data), followed by the country code, the workers job description (rider or driver) and a sequential number. For example, the folder containing movement data collected about the first rider in Austria should be named "MD_AT_Rider_1". The name(s) of the file(s) containing movement data should start with the same name as the folder.

In addition to the movement data, the following meta-data on how it was collected should be stored in a text document in the same folder:

- The period of data collection.
- The company the platform worker worked for during the data collection.

The worker's full name together with information on where his/her movement data is located must be stored in separate file in a different, secure location.

4.1.4 Data from technical audits

Data collected through Technical Audits should be stored in a separate folder. All file names related to one person should start with TA (Technical Audit), followed by the country code, the workers job description (rider or driver) and a sequential number. For example, a file with data collected through a technical audit about the first rider in Austria should be named "TA_AT_Rider_1…". In addition to the collected data, the following meta-data on how it was collected should be stored in a text document in the same folder:

- The period of data collection.
- The app that was audited.
- The brand and model of the (virtual) phone on which the audit was conducted.
- The technology used for the audit.

The full name of the person whose account was used for the audit and the file name where the data from the technical audit is located must be stored in separate file in a different, secure location.

4.2   Making data openly accessible

Data that cannot be fully anonymised will NOT be shared with individuals outside of the GDPoweR consortium. This applies specifically to personal data donated by workers and audio recordings from the focus group which will <u>not</u> be shared with individuals outside the GDPoweR research consortium.

Subject to informed consent by the relevant parties, anonymised transcripts from the focus groups as well as anonymised (partial) movement data and (partial) results of technical audits may be shared with other researchers upon request.

4.3   Making data interoperable

The data formats for written data (.docx), audio data (.mp3) and movement data (.kml) are widely used and in compliance with the recommended formats of the UK Data Service.[2] The naming of files as described in 4.1 follows the recommendations of the UK Data Service for brief and informative names.

The audio files and transcripts from focus groups are stored in the language in which the focus group was conducted to retain a maximum of information. The metadata information is stored in the English language to make them accessible to a large audience of researchers.

Data donated by platform workers is stored in the formats and languages in which they are received. The format of the data collected through technical audits of platform apps will depend on the specific technology used which may differ between apps and the cell phones on which they are installed.

4.4   Increase data re-use

To ensure the quality of audio recordings, two different recording devices are used as a default in all workshops and focus groups (conditional on the informed consent of the participants). To ensure the quality of written summaries of the workshops, handwritten notes will be taken even when audio recordings are permitted. The handwritten notes and the recordings together will then be used to prepare the transcripts.

Subject to informed consent by the relevant parties, pseudonymized data collected through focus groups and technical audits as well as pseudonymized movement data may be reused by project researchers for other scientific research and possibly for teaching or academic lectures. Data donated by platform workers may only be used for research outside of the GDPoweR project if explicit consent is given by the data donors.

4.5   Allocation of resources

The costs for making data FAIR are expected to be negligible. The resources for the preservation of the data for the described duration will be covered by the project partner who collected the data.

---

[2] UK Data Service (nd). Research data management. Available at: https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx

5. Data security

Data security is of upmost importance to ensure confidentiality and protect the privacy of individuals the data relates too. To ensure the safe and secure handling of the most sensitive data - personal data donated by platform workers - a Data Recovery Protocol was developed which describes a step-by-step process of how this data should donated, catalogued, securely stored, handled, and eventually deleted (for details see GDPoweR Data Recovery Protocol).

With regards to other primary data collected as part of the GDPoweR project, the following measures should be taken.

Storage: All primary data collected for the project should be stored only in password protected folders to which only researchers within the GDPoweR project who work with this data (need-to-know principle) have access. If the files are encrypted as well, the passwords for accessing the files should be stored separately. Storing the data in various places should be avoided to make it easier to trace and later delete all files.

Removal of personal identifiers:

'Personal identifiers' are data which can be used to identify a person like names, home addresses, bank account numbers or social insurance numbers. The names of focus group participants in transcripts should be replaced by unique identifiers (see section 4.1.1), unless they gave explicit consent to be named as source. Movement data and data from technical audits should be stored under a pseudonym. The persons' full names should be stored in separate file in a different, secure location.

Data collected through technical audits should be checked for sensitive personal information and identifiers (e.g. bank account information, date of birth, social insurance number). All information not directly relevant for the research should be removed or, where possible, replaced by a description of the information with the suffix "retracted_GDPoweR". For example, an IBAN like "AT30 1234 5678 9000 12" should be replaced with "IBAN_retracted_GDPoweR".

Ensuring that published results do not allow for the identification of the data owner:
When publishing results based on the collected primary data, utmost care should be taken to ensure that none of the results can be used to identify the donors. To this end, any results should be double-checked to ensure they contain no personally identifying information such as names, addresses, id numbers, date of birth, etc.

Sensitive data such as location data should only be published at a higher level of aggregation, i.e. when the data from several individuals is combined to ensure that no one person can be identified. Furthermore, the first and last drive of every day should not be displayed as these trips may start or end at the worker's personal address.

Deletion: The rules for deleting the collected personal data are laid down in the consent forms and the data donation contract. Unless specified otherwise, data which cannot be anonymised such as audio recordings and personal data donated by platform workers should be deleted one year after the end of the project.

6. Ethical aspects

All data is collected and stored in compliance with the General Data Protection Regulation (GDPR) of the European Union[3] and all other applicable European and national legislation.

All individuals participating in data gathering activities will be informed about the activity prior to its start through information letters (see GDPoweR Research Design Addendum I). Focus group participants and individuals participating in the collection of movement data or technical audits will be asked to sign consent forms (see GDPoweR Research Design Addendum I). The consent forms specify what data is collected, how it is used and how and for which duration it is stored. The consent forms further inform participants about their rights with regards to collection, use and storage (including the duration of storage), of all data.

The process of platform workers requesting copies of their personal data from digital labour platforms and subsequently donating it to the GDPoweR project for analysis is described in the GDPoweR Data Recovery Protocol. Platform workers willing to donate their data to the research project are asked to sign a data donation contract outlining their rights and how the data will be used.

---

[3] 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) Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

COORDINATOR

European Centre for Social Welfare Policy and Research (AT)

PARTNERS

HIVA-KU Leuven (BE)
ACV-Innovatief (BE)
ThEMA CY Cergy (FR)
Universidad de Castilla-La Mancha (ES)
Fundación 1° de Mayo (ES)
IBS - Institute for Structural Research (PL)