

DELIVERABLE

BIGPROD Personal Data Protection Strategy

Deliverable Information

Deliverable number and name:	Personal Data Protection Strategy
Due date:	31 th March 2022
Deliverable:	D10
Work Package:	WP3
Lead Partner for the Deliverable:	PPMI
Author:	Lukas Pukelis
Reviewers:	Arho Suominen, Hugo Hollanders and Scott Cunningham
Approved by:	Arho Suominen
Dissemination level:	Public
Version	22 th Apr 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

Disclaimer

This document contains a description of the **BIGPROD** project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium coordinator for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.



This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of BIGPROD consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (<http://europa.eu.int/>)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

Personal Data Protection and GDPR

Data security was one of our key priorities during the implementation of the BIG-PROD project. This encompassed our efforts to safeguard the data collected during the project from unwarranted access or breaches as well as ensuring that our own data collection practices would be in line with the highest standards of personal data protection and in compliance with the General Data Protection Regulation (GDPR).

The project team understands that if done incorrectly, using Big Data methods can result in breaches of GDPR and Personal Data Protection regulations. As a result, during the proposal phase of the BIGPROD project, we developed a strategy that helped us to mitigate these risks and to implement the project with a high degree of data security.

Measures at the Data Collection Stage

By its nature, BIGPROD is a data collection project, i.e., its main idea is that new data collected from new, previously unavailable sources could be beneficial to answering scientifically and socially relevant questions. As such, large amounts of data had to be collected during the project; however, we took active measures to ensure that we only collected data relevant to the project and to avoid collecting personal or sensitive data unintentionally.

Our team of researchers has performed extensive preliminary analyses and identified a list of patterns in webpages' URLs, which are very likely to contain personal data (e.g., "/contacts," "/contact-us," "/meet-the-team," etc.). We then programmed our web scraper in a way that would avoid the URLs containing these patterns. In doing so, our crawler never visits such webpages in the first place.

Measures Embedded in the Design of the Data Platform

Though the above-described measures are without a doubt effective, they do not guarantee that personal information is not shared unintentionally. For this reason, we have implemented additional measures. First, the initial data processing, integration, and structuring takes place on the PPMI premises in the dedicated database and ETL servers. Only when the data are processed and aggregated to the company level do we push it to Area I, where it can be accessed by all partners. The initial raw unaggregated data are kept on the PPMI databases only for the duration of the data mining process and are deleted soon thereafter.

This is particularly important as it also ensures a great degree of data security. It is extremely unlikely that a data leak could occur from the project platform and even less likely that it could contain sensitive unaggregated data or traces of personal data.

By design, some of the data collected and produced during the project are shared with the public. This sample of data only contains a subset of aggregated data that



were previously checked and cleared by the project team to be shared with the public.

In other words, by its very design, the BIGPROD data platform is compartmentalized in a way that helps to mitigate data security risks and to ensure the highest standards of personal data protection and GDPR compliance are upheld. The segmentation of the data platform helps to ensure that no data leaks are present and only that the intended data are shared with consortium members. Special procedures that were implemented during the release of the project's public dataset helped to ensure that no sensitive or personal data were unintentionally shared with the public.



For more information, please contact

Dr. Arho Suominen (Consortium leader)
Tel. +358 50 5050 354
arho.suominen@vtt.fi

About BIGPROD

BIGPROD is a research project focusing on Big Data based analysis of productivity using webscraped data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

The project partners in the project are Quantitative Science and Technology Studies team, Foresight-driven Business Strategies, 1) VTT Technical Research Centre of Finland, Competence Center Innovation and Knowledge Economy (Coordinator), 2) Fraunhofer ISI, Economics of Knowledge and Innovation team, 3) UNU-MERIT, Maastricht University, 4) Public Policy and Management Institute, 5) Economics of Technology and Innovations, Faculty of Technology, Policy and Management, 6) Delft University of Technology, Economics of Technology and Innovations, 7) Faculty of Technology, Policy and Management, Delft University of Technology



PPMI

Fraunhofer
ISI

Maastricht University

TU Delft
Delft University of Technology

University of
Strathclyde
Glasgow

www.bigprod.eu

BIG
PROD