Ref. Ares(2026)8849347 - 18/09/2026

BRIDGE//GAP

BRIDGING THE GAPS
IN EVIDENCE, REGULATION
AND IMPACT
OF ANTICORRUPTION
POLICIES

D11.3 Data Management Plan and Research Ethics - Second version

D11.3 Data Management Plan and Research Ethics - second version

Project Title	BRIDGEGAP
Grant Agreement Number	101132483
WP No & Title	WP 11 Management
Deliverable Lead & Beneficiary	Francesco Molinari The Lisbon Council (LC)
Contractual Delivery Date	30 June 2025
Actual Delivery Date	30 April 2026
Delayed	Yes, as agreed during the 1st project review meeting on 6 June 2025 and during following talks with the new PO
Authors	Francesco Molinari, Marcello Verona & Marieke Willems, LC
Contributors	Authors of BridgeGap deliverables spanning the Data Hub (D1.2), European Transparency Index (D5.1), EUCAL (D8.1), EUROPAM extensions (D9.2), and sensitive WP3/WP6/WP7 datasets
Acknowledgements (not grant participants)	
Reviewers	See list of Contributors
Dissemination Level	PUBLIC
Document description	Second edition of the document describing the consortium's overall data management lifecycle process for the data anticipated to be collected, processed, generated and further exchanged throughout the project duration. Takes benefit from the received recommendations after the 1st review meeting of the project, and from the contents of the deliverables uploaded until end December 2025.

Funded by the European Union

CC BY

Disclaimer

BRIDGEGAP is funded by the European Commission's Horizon Europe Programme for Research and Innovation, under the Grant Agreement number 101132483. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, neither can the European Union be held responsible for them

Document overview

This deliverable is the second version of the Data Management Plan (DMP) of the BridgeGap project. It updates the information supplied in the first edition, which was gathered at project's start through an internal partner survey. This version takes benefit from the guidelines and recommendations received after the first review meeting of the project held on 6 June 2025 and takes stock of the contents of deliverables uploaded until 31 December 2025. The document comprises the common sections for data management planning according to the EU Horizon Europe standards and provides an overview of all the BridgeGap datasets based on their current development levels. The DMP is a living document that is continuously updated throughout the project runtime. Its second version includes an initial overview of the sharing of responsibilities for data management after the end of the BridgeGap project's lifetime, which will be the main subject of the third and last DMP update.

Disclaimer

BridgeGap is funded by the European Commission's Horizon Europe Programme for Research and Innovation, under the Grant Agreement number 101132483. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, neither can the European Union be held responsible for them.

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Version History

date	version	who	description
16.01.2026	v0.1	Marieke Willems, Francesco Molinari, LC	ToC
24.04.2026	v0.2	Francesco Molinari, LC	Contents ready for the internal reviews
29.04.2026	V0.3	Marieke Willems, Evgenia Malikova, Luca Remotti, LC	Review
30.04.2026	v1.0	Marieke Willems, LC	Final version

Table of Contents

1. INTRODUCTION 9

1.1. LEGAL BASIS 9
1.2. TRANSITIONING BETWEEN EDITIONS OF THIS DMP 9
1.3. REVISED TABLE OF CONTENTS AND ORGANISATIONAL STRUCTURE FOR D11.3 11

2. DATA SUMMARY 13

2.1 OVERVIEW 13
2.2 CLASSIFICATION OF DATA TYPES AND FORMATS 13
2.3 VOLUME, PROVENANCE AND LIFECYCLE MANAGEMENT 15
2.4 REUSE POTENTIAL 15

3. FAIR MANAGEMENT OF DATA 17

3.1 MAKING DATA FINDABLE 17
3.2 MAKING DATA ACCESSIBLE 17
3.3 MAKING DATA INTEROPERABLE 17
3.4 MAKING DATA REUSABLE 18
3.5 FINDABILITY AND METADATA ENRICHMENT 18
3.6 ACCESSIBILITY AND THE 'OPEN BY DEFAULT' MANDATE 18
3.7 INTEROPERABILITY AND SYNCHRONISATION REQUIREMENTS 19
3.8 REUSABILITY AND LICENSING STRATEGY 19
3.9 THE ROLE OF THE DATA MANAGER AND COMMUNITY CURATION 19

4. DATA AND METADATA STORAGE AND BACKUP 21

4.1 GENERAL MEASURES 21
4.2 PARTNER-SPECIFIC PRACTICES (SELECTED EXAMPLES) 21
4.3 THE RELATIONAL ARCHITECTURE OF THE DATA HUB 21
4.4 SYNCHRONISATION WITH EUROPAM AND EUCAL 22
4.5 REQUIREMENTS FOR ZENODO EXTRACTION 22
4.6 PROPOSED DATASET INVENTORY FOR M36 23
4.7 VERSIONING STRATEGY FOR REGULATORY DATABASES 23

5. DATA SECURITY AND PROTECTION OF SENSITIVE DATA 25

5.1 GENERAL SECURITY AND GDPR COMPLIANCE RULES 25
5.2 PARTNER-SPECIFIC SAFEGUARDS FOR HIGH-RISK DATA 25
5.3 MANAGING DATA FLOW FROM AFFILIATED ENTITIES 25
5.4 GDPR COMPLIANCE AND SENSITIVE DATA PROTOCOLS 26

5.5 De-identification Protocols for Sensitive Research 26

6. DATA SHARING 28

6.1 Internal sharing 28
6.2 External sharing and embargoes 28

7. REPOSITORIES FOR REUSE AND PERSISTENT IDENTIFIERS 29

7.1 Preferred repositories and communities 29
7.2 DOI and licence policies 29

8. METHODS AND TOOLS FOR ACCESSING, ANALYSING AND UPLOADING DATA 30

8.1 Data access and analysis tools 30
8.2 Uploading and dissemination tools 30

9. DATA PRESERVATION 31

9.1 Selection and retention periods 31
9.2 Sustainability of online tools 31
9.3 Knowledge Transfer and Policy Impact 31

10. ETHICAL ISSUES AND CODE OF CONDUCT 33

11. ALLOCATION OF RESOURCES AND RESPONSIBILITIES 34

12. CONCLUSIONS AND NEXT STEPS 35

13. ANNEX. PROPOSED LIST OF DATASETS TO BE FEATURED ON ZENODO 36

Tables

Table 1 Action Plan and input Request Framework 11
Table 2 Data Type Classification (M1-M24 Deliverables) 14
Table 3 Overview of the Data Hub 22
Table 4 Overview of the M36 Dataset Inventory 23
Table 5 Updated split of tasks and responsibilities among project partners 26

Executive Summary

This second version of the BridgeGap Data Management Plan (D11.3) updates D11.2 to reflect the project's maturation at month 26 (April 2026), incorporating lessons from the first review meeting (June 2025), the amended Grant Agreement (April 2025) adding 15 TI-S affiliated entities, and deliverables up to M24 (December 2025).

It provides a comprehensive overview of the emerging data ecosystem - spanning the Data Hub (D1.2), European Transparency Index (D5.1), EUCAL (D8.1), EUROPAM extensions (D9.2), and sensitive WP3/WP6/WP7 datasets - while ensuring full FAIR compliance, GDPR adherence, and ethical standards. The DMP details handling, storage, sharing, preservation, and post-project sustainability for tabular, geospatial, network, and qualitative data, with a proposed Zenodo inventory to operationalise open access.

As a living document, it aligns with the "Roadmap for the new DMP" action plan, paving the way for a final version of the DMP, and maximising reuse by researchers, policymakers, and civil society. For the benefit of this deliverable, we will refer to this final version of the BridgeGap DMP, as the D11.4 Data Management Plan and Research Ethics - Final version.

1. Introduction

1.1. Legal basis

BridgeGap is a Research and Innovation Action funded under the framework of the Horizon Europe programme with the Grant Agreement number: 101132483. The project officially started on 01 January 2024 and will end on 31 December 2027. The consortium agreement was signed by all partners in October 2023.

BridgeGap *Task 11.4: Data Management, Privacy, Ethical, Societal and Legal Compliance & Analysis* engages the full consortium on delivering the Data Management Plan (DMP), which describes the overall data management lifecycle process for the data anticipated to be collected, processed, generated and further exchanged throughout the project duration.

In alignment with the principle of making a project's data Findable, Accessible, Interoperable and Reusable (FAIR), the DMP shall include details on:

1. how the data will be handled during and after the project's completion;
2. what data will be collected and/or processed, generated and further exchanged;
3. any standards that will be applied;
4. identifying whether the data will be made open access; and
5. how the data will be preserved, also considering the post project period, thus determining how partners might exploit data post project-lifecycle.

Task 11.4 also investigates the ethical and legal aspects that are relevant for BridgeGap by considering both the technical work within the project, as well as the regulatory and policy developments at EU level. To this end, the task aims at:

1. identifying the key ethical, legal, social and privacy related themes of the BridgeGap platform;
2. mapping the data flows amongst the BridgeGap technologies, users and the services with which they interface;
3. identifying key risks associated with these data flows;
4. engaging with all partners to suggest possible technical or operational solutions, mitigation measures with special attention given to privacy, social impact, security, lawful basis, and aspects of surveillance;
5. producing a final assessment of the proposed solutions from a legal and ethical viewpoint, including recommendations that will further support the delivery of BridgeGap tools to practitioners.

1.2. Transitioning between editions of this DMP

The evolution of the BridgeGap project from its conceptual inception into a multi-jurisdictional research engine necessitates a sophisticated and dynamic approach to data management. As a successor to the ANTICORRP consortium, BridgeGap (Grant Agreement 101132483) is tasked with producing scientific evidence that empowers reformers and civil society to prevent power abuse. The transition from the first

iteration of the Data Management Plan (D11.2) to this second version (D11.3) marks a critical shift from policy formulation to the active orchestration of diverse data flows across eleven work packages. This document also establishes the technical and ethical protocols required to manage the massive influx of data generated by 15 new affiliated entities and the integration of the central Data Hub with legacy repositories such as EUROPAM and the newly developed EU Compass on Anticorruption Law (EUCAL).

The primary objective of this second-version DMP is to ensure that all research outputs remain Findable, Accessible, Interoperable, and Reusable (FAIR), while strictly adhering to the European Commission's Horizon Europe standards and the General Data Protection Regulation (GDPR). The strategy moves beyond simple data storage to address the nuances of machine-readability for artificial intelligence, the complexities of anonymising sensitive cross-border criminal cases, and the versioning requirements of evolving regulatory databases.

Another objective is to pave the way for the delivery of the third and final version of this DMP (we refer to this as the D11.4 Data Management Plan and Research Ethics - Final version), including an initial overview of the sharing of responsibilities for data management after the end of the BridgeGap project's lifetime.

The roadmap for the development of D11.4 provides a strategic timeline for aligning research outputs with the technical infrastructure of the project. Spanning from month 25 to month 48, this roadmap prioritises the transition of raw data into curated datasets suitable for public dissemination and high-level policy analysis. The progression is characterised by a shift from the consolidation of metadata standards in Zenodo to the operational synchronisation of the Data Hub with external partner databases. This process is governed by a rigorous action plan that assigns specific data management focuses to each major deliverable issued up to the second-year milestone. A central component of this roadmap is the integration of the UNODC Statistical Framework model, which provides a harmonised structure for measuring corruption across diverse political and legal systems. The roadmap details the steps for populating the Data Hub with 42 distinct indicators, ranging from fiscal transparency to internet empowerment, ensuring that each data point is supported by a clear methodology and provenance trail. This structured approach allows the project to “bridge the gap” between academic theory and actionable policy interventions by providing a real-time map of corruption risks and regulatory deficits across the EU and its neighbourhood.

The roadmap is also accompanied by an input request for the authors of deliverables issued until M24, which will be complemented by a similar action for those pertaining to the second half of the project’s lifetime. This approach will be ensuring that data management is not a retrospective task but a fundamental part of the research lifecycle. By identifying specific focuses such as interoperability, anonymisation, and versioning, the project mitigates the risk of technical silos and ensures that datasets can be synthesised into the central Data Hub.

TABLE 1 ACTION PLAN AND INPUT REQUEST FRAMEWORK

Deliverable ID	Project Month	Primary Data Management Focus	Technical Requirement and Specification
D1.2	M10	Hub Interoperability	Relational database integration with OAI-PMH harvesting
D1.3	M18	Historical Provenance	Archival metadata tagging for cross-border corruption reports
D3.1	M12	Ethnographic Anonymisation	De-identification of actor-network maps in the energy sector
D3.3	M24	Sanction Data Anonymisation	Compliance with GDPR for cross-border undue influence cases
D5.1	M18	Sub-indicator Granularity	XLS/CSV mapping for the European Transparency Index
D6.1	M15	Machine-Readability	JSON/CSV suitability for AI training on procurement data
D8.2	M18	Legal Database Standards	UNODC-compliant coding for the EUCAL repository
D8.3	M24	Criminal Law Anonymisation	Protocols for sensitive enforcement and sanction records
D9.2	M24	DOI-Linked Versioning	Tracking regulatory updates through persistent identifiers

The technical implementation of this action plan requires a decentralised yet coordinated effort. Authors are tasked with ensuring that their research materials are processed for machine-readability from the outset, particularly in Work Package 6, where the adequacy of public procurement data for AI purposes is a primary concern. This involves moving beyond standard PDF reporting to provide structured data supplements that can be ingested by the Data Hub and the Follow the Money search engine.

1.3. Revised Table of Contents and Organisational Structure for D11.3

Compared with D11.2, the revised organisational structure of the second-version DMP reflects the increased complexity of the project's data ecosystem. Unlike the initial version, which focused on overarching principles, D11.3 provides granular protocols for the interaction between the Data Hub and the expanded consortium. The new structure is organised to facilitate the rapid identification of data management protocols for different research modalities, from quantitative indices to qualitative case studies. This organisational framework prioritises the Data Hub as the central repository while maintaining a strict adherence to the Zenodo community standards for all archival outputs. By mapping content onto the D11.2 template while

incorporating the new organisational structure, the DMP ensures continuity while addressing the specific technical challenges identified during the first 24 months of the project.

2. Data Summary

2.1 Overview

The BridgeGap project generates a vast array of data that must be categorised to ensure proper handling and preservation. This data ranges from administrative procurement notices and legal statutes to original survey data and ethnographic field notes. The technical quality of these data is verified through a rigorous process of validation, cleaning, and standardisation, ensuring that they are suitable for high-level statistical analysis and AI-driven corruption detection.

- **Regulatory and legal data**
  - EUCAL Online Comparative Regulation Database and Map (D8.1): structured, country-level legal and institutional data on criminal law frameworks for corruption and related offences, with associated map visualisation.
  - Updated and extended EUROPAM.eu (D9.2): longitudinal regulatory and implementation indicators on public accountability tools across EU MS and candidate countries.
- **Perceptual / analytical indices and transparency datasets**
  - European Transparency Index (ETI) map, visualisation and database (D5.1): composite transparency indicators and underlying microdata for EU and candidate countries, including contributions from 15 TI-S affiliated entities.
  - Data Hub (D1.2): pooled, interlinked datasets from Integrity Watch, T-Index, CRF, EUROPAM, YouControl and other sources, exposed through a central API and search interface.
- **Sensitive / high-risk research datasets**
  - Cross-border undue influence and security-related material (D3.3 and associated WP3 work): qualitative and mixed-method datasets on weaponised corruption and foreign undue influence, including elite interviews, case studies and, where applicable, derived variables from open-source intelligence and social media.
  - AI, digital-transformation and experimental data in WP4 and WP6 (e.g. D6.5, internal experimental datasets): survey responses, behavioural-experimental outputs, and training/test sets for AI tools on public procurement and related data.

Additional qualitative and quantitative datasets, largely at planning or early collection stage by month 24, are associated with WP1 background studies, WP2 ethnographic work, WP4 organisational case studies, WP7 media capture risk analysis, and WP8 enforcement capability studies (D8.3 first version). These are managed under the same principles but often remain embargoed and/or partner-internal at this stage.

2.2 Classification of Data Types and Formats

The project classifies its outputs into several distinct data streams, each with its own preferred format and management requirements. For quantitative datasets, such as the Index of Public Integrity and the European Transparency Index, the focus is on interoperability and machine-readability. These are typically provided in

CSV or Excel formats to allow for easy ingestion into the Data Hub and external analytical software. In contrast, qualitative reports and historical analyses are preserved in PDF/A format to ensure long-term accessibility and adherence to archival standards.

Across the project, the main **data types and formats** are:

- Tabular structured data: CSV, Excel, SQL exports (EUCAL, ETI, EUROPAM, procurement datasets, survey data).
- Semi-structured data: JSON, XML (APIs from YouControl, social media, web-scraping outputs, analytics logs).
- Textual documents: PDF, DOCX, RTF (interview transcripts, legal summaries, coding manuals, methodological annexes).
- Network and graph data: edge-lists and node tables in CSV / GraphML (actor and ownership networks, media capture networks).
- Multimedia: audio/video interview files (MP4) where strictly necessary for transcription and only under secure storage.

TABLE 2 DATA TYPE CLASSIFICATION (M1-M24 DELIVERABLES)

Deliverable	Data Type	Preferred Format	Technical Context and Purpose
D3.1	Ethnographic Maps	CSV, PDF/A	Mapping actor-networks and ties in the energy sector
D5.1	Transparency Indices	XLS, CSV	Monitoring political finance and lobby transparency
D6.1	Procurement Data	JSON, CSV	Comparative suitability for AI-driven anti-corruption
D8.2	Legal Repositories	SQL, CSV	Comparative database of criminal law and sanctions
D1.3	Historical Reports	PDF/A	Background research on cross-border corruption

The technical quality of the data is further enhanced by the inclusion of comprehensive documentation, such as README files and variable lists, which accompany every Zenodo deposit. This ensures that external researchers can understand the provenance of the data, including the sources of administrative records and the methodology used for the calculation of composite indices. For instance, the data utilised in the Data Hub is sourced from a variety of prestigious institutions and platforms, including OpenTender.eu, the Corruption Perceptions Index (CPI), and the UNODC, with each source being clearly documented in the database's relational structure.

2.3 Volume, Provenance and Lifecycle Management

Volume estimates at M24 remain modest in storage terms but high in complexity: individual datasets are typically in the tens to hundreds of megabytes, with total pooled data in the Data Hub expected to reach several gigabytes as interconnection with at least 20 external databases is completed.

Provenance is mixed, combining:

- Primary data collected by BridgeGap partners (surveys, experiments, interviews, expert assessments, event analytics).
- Secondary open data (official statistics, public procurement portals, public registers, existing indices).
- Licensed or purchased datasets (media content databases, commercial transparency datasets) acquired under “Other Direct Costs”, especially by TI-S affiliates for D5.1 and by partners in WP7.

All datasets are accompanied by dataset-level documentation describing source(s), sampling frames, processing steps, and links to related BRIDGEGAP deliverables.

The provenance of BridgeGap data is meticulously tracked to maintain research integrity and reproducibility. For datasets sourced from administrative portals, such as Tenders Electronic Daily (TED), the project documents the date of extraction, the filters applied, and the cleaning algorithms used to resolve issues with company names and beneficial ownership. This is particularly critical for Work Package 6, where the quality of the underlying data directly influences the effectiveness and fairness of the AI tools being tested to detect corruption.

The data lifecycle management protocol specifies that data should move from a raw collection phase to a cleaning and validation phase, followed by a final curation phase before being published in the Data Hub or Zenodo. During the cleaning phase, rigorous standardisation processes are undertaken to ensure that indicators such as "Single-Bid Procurement Rates" are comparable across different national contexts. This multi-stage process ensures that the final data products are of the highest technical quality and suitable for use by policymakers and practitioners alike.

2.4 Reuse Potential

Outside the project, the main potential users are:

- Researchers and students in political science, law, economics, data science and media studies.
- Practitioners in EU and national institutions, anti-corruption agencies, prosecution services and audit bodies.
- Civil-society organisations, investigative journalists and civic technologists interested in corruption, transparency and media capture.

Regulatory and index-type datasets (ETI, EUCAL, EUROPAM extensions) are designed for long-term reuse; sensitive micro-data (elite interviews, social-media content, individual-level survey responses) will normally be shared only in anonymised, aggregated or synthetic form, if at all.

3. FAIR management of data

The BridgeGap project is deeply committed to the FAIR principles, viewing them not just as a compliance requirement but as a means of maximising the societal impact of its research. The project utilizes the Zenodo repository, hosted by CERN, as its primary platform for long-term data preservation and public access. Zenodo provides a robust infrastructure for assigning persistent identifiers and ensuring that data remains findable and citeable for decades.

3.1 Making data findable

Findability is ensured through:

- Systematic use of DOIs for public datasets via Zenodo or equivalent repositories, including a BridgeGap community where possible.
- Descriptive, versioned dataset names (e.g. BRIDGEGAP_ETI_country_indicators_v1_2025, BRIDGEGAP_EUCAL_legal_database_v1_2025).
- Rich metadata in line with Dublin Core and, where appropriate, domain standards (e.g. for legal and social-science data).
- Cross-linking from the BRIDGEGAP website and Data Hub (D1.2) to repository landing pages.

For internal or restricted datasets, persistent internal identifiers and a central registry (maintained under T11.4) record location, access conditions and responsible partner.

3.2 Making data accessible

Accessibility follows the principle “as open as possible, as closed as necessary”:

- Open datasets (e.g. ETI, EUCAL, selected Data Hub exports, EUROPAM extensions) are made available under Creative Commons licences, preferably CC BY 4.0, via Zenodo and the project website.
- Data involving personal data, high-risk security content, or commercial licences are restricted; access may be provided on justified request, under data use agreements and, where necessary, non-disclosure agreements.
- Internal access to sensitive data is controlled through role-based access control, VPNs, password protection and encryption, as described in Section 5.

All repository entries clearly state access level (open / embargoed / restricted) and conditions.

3.3 Making data interoperable

Interoperability is driven by the Data Hub design and the roadmap’s emphasis on harmonisation:

- Common coding schemes for countries (ISO), years, sectors and institutional categories are applied across ETI, EUCAL, EUROPAM extensions and Data Hub datasets to enable cross-dataset queries.
- Metadata schemes (Dublin Core plus dataset-specific fields) are implemented consistently.
- Where feasible, open, non-proprietary formats (CSV, JSON) are preferred for public releases, even if partners work internally with Stata, R or SPSS.

Technical documentation for the Data Hub API and metadata is collected from the D1.2 team and referenced in this DMP.

3.4 Making data reusable

Reusability is supported through:

- Clear licensing: CC BY 4.0 as the default for BridgeGap-generated open datasets, with explicit statements where upstream licence constraints apply.
- Full methodological documentation (variable definitions, construction of indices, handling of missing data, sampling procedures).
- Versioning and changelogs for longitudinal datasets (ETI, EUROPAM, EUCAL), ensuring researchers can replicate specific deliverable versions.
- Where sensitive data cannot be shared, provision of detailed codebooks, analytical scripts and, if appropriate, synthetic or highly aggregated datasets.

3.5 Findability and Metadata Enrichment

Findability is ensured through the automatic registration of Digital Object Identifiers (DOIs) for every record published in the BridgeGap Zenodo community. The project adheres to the DataCite Metadata Schema v4, which includes essential fields such as title, creator, publication date, and funder. To improve the discoverability of its data, BridgeGap enriches its metadata with additional article-level fields, including journal titles, conference acronyms, and alternate persistent identifiers.

The use of controlled vocabularies, such as EuroSciVoc for field-of-science classification and the UNODC framework for corruption-related topics, further enhances the metadata's quality. Researchers are mandated to include their ORCID iDs and institutional ROR identifiers in all metadata submissions, ensuring that credit is accurately attributed and that the project’s outputs can be easily tracked by the European Commission and other funding bodies.

3.6 Accessibility and the ‘Open by Default’ Mandate

In line with the Horizon Europe requirements, all BridgeGap research outputs are open access by default. Metadata for all records is publicly accessible and licensed under the CC0 waiver, allowing for frictionless harvesting by search engines and data aggregators. While the project produces no sensitive personal data in its primary quantitative indices, certain qualitative or criminal law datasets may be subject to restricted access or embargo periods to protect investigative integrity or comply with legal restrictions.

Accessibility is facilitated through standard, open protocols such as REST API and OAI-PMH, which allow for the automated retrieval of metadata and datasets. The project also ensures that metadata remains available even if the underlying data is no longer accessible, a critical requirement for maintaining a permanent record of scientific inquiry. For restricted files, the project provides clear instructions on how to request access, ensuring that the process is transparent and fair.

3.7 Interoperability and Synchronisation Requirements

Interoperability is achieved through the use of formal, broadly applicable languages for knowledge representation, primarily JSON Schema within Zenodo and relational database structures within the Data Hub. The project's commitment to interoperability extends to its synchronisation with external repositories like EUROPAM.eu and EUCAL. These repositories provide critical data on preventive and criminal regulation, which are integrated into the Data Hub to provide a comprehensive view of the public integrity framework.

The synchronisation process requires that all incoming data from partner repositories is mapped onto the central Hub's coding system, which distinguishes between direct indicators of corruption and indirect indicators such as transparency, regulation, and enforcement. This semantic synchronisation allows the Hub to present a unified interactive dashboard where users can explore the drivers and deterrents of corruption across 41 countries.

3.8 Reusability and Licensing Strategy

To foster the reuse of its data, BridgeGap employs clear and accessible data usage licenses. Most datasets and reports are released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which allows others to share and adapt the material as long as appropriate credit is given. This license is compatible with the "Open Definition" and is the recommended standard for research data funded by the European Commission.

Reusability is further supported by the provision of detailed provenance and methodological documentation. Every record on Zenodo is traceable to a registered user, and the metadata describes the original authors and their affiliations. This high level of transparency allows third-party researchers to validate the project's findings and integrate BridgeGap data into new research initiatives, thereby extending the value of the original investment.

3.9 The Role of the Data Manager and Community Curation

The data management team, led by experts from CSD, LUISS and the TI Secretariat, plays a central role in ensuring the technical quality and FAIRness of the project's outputs. This team is responsible for managing the Zenodo community, reviewing deposits for metadata accuracy, and providing guidance to deliverable authors on data standards. The curation policy for the BridgeGap Zenodo community ensures that only data whose research principal is the organiation and that meets the required level of documentation quality is published.

Funded by
the European Union

4. Data and Metadata Storage and Backup

4.1 General measures

All partners store research data on secure institutional infrastructures with regular backups, access controls and adherence to local data-protection policies, as described in D11.2 and in partner-specific contributions. Central elements include:

- Storage on institutional servers, encrypted cloud services (e.g. OneDrive for Business, secure university clouds) or secure third-party providers that are GDPR-compliant.
- Daily or frequent backups, often to separate physical or cloud locations.
- Separation of working copies from archival copies, with clearly designated responsible persons per dataset.

The Data Hub itself is hosted on secure infrastructure specified in D1.2; its internal databases are regularly backed up and mirrored, with failover mechanisms where feasible.

4.2 Partner-specific practices (selected examples)

- University of Glasgow: uses password-protected cloud storage (OneDrive for Business) with long-term retention (minimum 10 years for data of long-term value).
- Lisbon Council: Matomo Analytics data are stored on a dedicated Hetzner server in Frankfurt with encrypted backups to AWS S3; only aggregated statistics are preserved to the end of the project.
- Università di Perugia: combines secure institutional servers and GDPR-compliant cloud services (e.g. AWS) with strong access controls and regular backups for media-related datasets.

4.3 The Relational Architecture of the Data Hub

The BridgeGap Data Hub represents a state-of-the-art platform for the visualisation and sharing of anti-corruption metrics. Developed through the collaborative efforts of Transparency International, LUISS University, the Romanian Academic Society (SAR), and YouControl, the Hub centralises high-quality data to enable cross-country evaluations of anti-corruption effectiveness.

The Data Hub's relational database structure is designed for scalability and flexibility, allowing for advanced queries and frequent updates. The final database consists of 42 indicators organised into six theoretical dimensions: Corruption Measures, Transparency, Regulation, Enforcement, Freedom of Expression, and Internet Empowerment. This theoretical structure is grounded in the belief that the control of corruption results from a balance between opportunities for corruption and the constraints posed by states and societies.

TABLE 3 OVERVIEW OF THE DATA HUB

Theoretical Dimension	Key Indicators and Metrics	Source and Provenance
Corruption Measures	CPI Score, Single-Bid Procurement Rate	TI-S, OpenTender.eu, TED
Transparency	T-Index (Fiscal, Judicial, Political), Lobby Index	LUISS, ERCAS/SAR
Regulation	Political Finance Score, Conflict of Interest Score	EUROPAM, EUCAL, World Bank
Enforcement	Cross-Border Cases, Sanctioned Companies	National Registries, UNODC
Freedom of Expression	Media Freedom, Civil Society Participation	V-Dem, Reporters Without Borders
Internet Empowerment	Online Services, E-Citizenship Score	UN E-Government Survey

The technical quality of the Hub is maintained through rigorous cleaning and validation processes that ensure the accuracy and reliability of the data. This is particularly important for composite indicators like the Index of Public Integrity (IPI), which combines multiple data points to provide a single risk score. The Hub's interactive dashboard allows users to visualise these complex data points through dynamic maps, bar charts, and tables, making the evidence accessible to a non-expert audience.

4.4 Synchronisation with EUROPAM and EUCAL

A primary technical challenge for the second half of the project is the full synchronisation between the Data Hub and repositories like EUROPAM.eu and EUCAL.eu. EUROPAM provides de jure transparency data based on internationally accepted legal standards, while EUCAL serves as an online comparative regulation database focusing on criminal law. The synchronisation requirement dictates that these external databases must adopt the Hub's standardised coding and temporal structures to allow for real-time data pooling.

The goal is to complete this integration by the end of 2026, thereby adding granular preventive and criminal regulation data to the Hub's existing dataset. This will allow researchers to correlate legal frameworks (de jure) with actual transparency outcomes (de facto), providing a powerful tool for identifying regulatory and impact gaps. The use of the UNODC Statistical Framework model as the common denominator for these databases ensures that the findings are robust and comparable across different jurisdictions.

To ensure the success of the Data Hub and the project's archival goals, deliverable authors are provided with a set of actionable requirements for data management. These requirements focus on the early identification of datasets for extraction and the adoption of the project's metadata and licensing standards.

4.5 Requirements for Zenodo Extraction

Authors are required to prepare a "data supplement" for every technical deliverable that generates or re-uses data. This supplement must include the raw or semi-processed data used for the report's findings, formatted

in a machine-readable way. Authors must also ensure that their datasets are assigned a DOI and published in the BridgeGap Zenodo community, using the CC BY 4.0 license unless there is a strong justification for a more restrictive approach.

4.6 Proposed Dataset Inventory for M36

The following datasets have been identified as high-priority candidates for Zenodo extraction and Hub integration up to the 36-month mark. Each dataset is proposed with a specific name and metadata standard to ensure consistency across the project.

TABLE 4 OVERVIEW OF THE M36 DATASET INVENTORY

Proposed Dataset Name	Description and Scope	Metadata Standard	License
BG_WP1_DH_v1.0	Data Hub consolidated indicators (42 variables)	DataCite v4 / UNODC	CC BY 4.0
BG_WP5_ETI_v2025	European Transparency Index sub-indicator data	DataCite v4	CC BY 4.0
BG_WP6_PPSA_v1	Procurement suitability for AI matrix (D6.1)	DataCite v4	CC BY 4.0
BG_WP8_EUCAL_v1	Online comparative regulation database records	DataCite v4 / UNODC	CC BY 4.0
BG_WP9_RE_v1	Data supplement for regulatory and enforcement gaps	DataCite v4	CC BY 4.0
BG_WP3_CBC_v1	Cross-border corruption case study metadata	DataCite v4	CC BY 4.0

The adoption of a uniform naming convention (e.g., BG_WP#_NAME_v#) ensures that the project's outputs are easily identifiable in global repositories. Furthermore, the use of semantic versioning allows for the tracking of updates to longitudinal datasets, such as the Transparency Index, which is updated annually.

4.7 Versioning Strategy for Regulatory Databases

As the project enters its second phase, the implementation of versioning and anonymisation protocols has become a technical priority. These protocols ensure that updates to regulatory databases are traceable and that sensitive information is protected in accordance with the project's ethical guidelines.

For deliverables such as D9.2, which maps regulatory gaps over time, a clear versioning strategy is essential. The project utilises Zenodo's versioning feature, which allows researchers to publish updates to a dataset while maintaining separate metadata and persistent identifiers for each version. This ensures that

any researcher citing a specific version of the EUCAL database or the Transparency Index can be confident that the files they used will not be modified, preserving the reproducibility of their findings.

The versioning protocol requires that:

1. Every major update is released as a new record with a unique DOI.
2. The "parent" DOI for the dataset is used to link all versions together, making it easy to find the most recent data.
3. Metadata for older versions remains available, even if the files are no longer current.

5. Data security and protection of sensitive data

5.1 General security and GDPR compliance rules

The project handles several categories of **potentially sensitive data** (personal data from surveys and interviews, public but potentially sensitive social-media content, security-relevant case studies). Security and GDPR compliance are ensured through:

- Encryption of stored data (e.g. AES-256) and encrypted transmission (TLS), especially for interview recordings, consent forms and confidential case-materials.
- Role-based access control, strong authentication (including 2FA in relevant systems), and strict limitation of access to authorised team members.
- Data minimisation (collecting only data strictly necessary for each task) and pseudonymisation/anonymisation wherever possible before analysis and sharing.
- Institutional ethics approvals and data-protection assessments (e.g. DPIAs where required), as outlined in D11.2.

Sensitive D3.3, D6.5, D8.3 and WP4/WP7 datasets are stored only on secure institutional infrastructure, never on personal devices without encryption, and are not copied into the Data Hub except in heavily aggregated or anonymised form.

5.2 Partner-specific safeguards for high-risk data

- Social-media data (WP3, WP6, WP7): partners use platform APIs and, where possible, conduct analysis inside secure platform environments; local storage is minimised, anonymised and time-limited, in line with AoIR and other social-media ethics guidelines.
- Elite interviews and high-profile experts (WP3, WP8): transcripts are pseudonymised, stored separately from identity keys, and protected with strong encryption; dissemination uses only aggregated or de-identified findings.
- TI-S affiliates (see below): data flows from national chapters into ETI and Integrity Watch follow harmonised GDPR-compliant protocols and are governed by data-processing arrangements with TI-S.

5.3 Managing Data Flow from Affiliated Entities

After a dedicated amendment to the Grant Agreement, in April 2025, the BridgeGap consortium has expanded to include 15 new affiliated entities, primarily from the Transparency International (TI) network, such as TI-S CZ (Czech Republic) and TI-S MT (Malta). This expansion necessitates a rigorous data flow and GDPR compliance framework to ensure that personal data is handled securely and in accordance with EU law.

The data flow within the project is structured to move from national-level collection by affiliated entities to central processing by lead beneficiaries like LUISS and the TI Secretariat. Affiliated entities are responsible for collecting data on local procurement notices, regulatory updates, and criminal sanctions, which are then transmitted to the central Hub via encrypted channels. This decentralised collection model ensures that the project benefits from local expertise while maintaining a unified data standard.

TABLE 5 UPDATED SPLIT OF TASKS AND RESPONSIBILITIES AMONG PROJECT PARTNERS

Entity Type	Roles and Responsibilities	Data Contribution Focus
Beneficiaries (e.g., LUISS, TI EV)	Project management, methodology, Hub hosting	Global indices, quality control, FAIR compliance
Affiliated Entities (e.g., TI-S CZ, TI-S MT)	National data collection, local dissemination	Procurement data, regulatory updates, case studies
Partners (e.g., YouControl)	Technical tool development, platform hosting	Follow the Money engine, business registry data

This structure is formalised through the Grant Agreement amendment, which specifies the terms of participation for the new entities. Each entity must adhere to the project's data management protocols, ensuring that the data they contribute is properly formatted and accompanied by the necessary metadata for Zenodo publication.

5.4 GDPR Compliance and Sensitive Data Protocols

GDPR compliance is a paramount concern, particularly for work packages that handle data on criminal law and political influence. The project adopts a "Privacy by Design" approach, ensuring that data protection is integrated into every stage of the research process. While most of the data collected is from public records—such as company registries and court rulings—the project still applies strict anonymisation and pseudonymisation protocols where necessary.

For deliverables like D3.3 and D8.3, which focus on cross-border corruption and criminal sanctions, the data management focus is on de-identifying personal data while preserving the analytical value of the information. This involves replacing names with alphanumeric codes and aggregating data at the organisational or national level where possible. The Data Hub platform itself does not store sensitive personal data, instead providing links to original public sources or displaying aggregated risk scores.

5.5 De-identification Protocols for Sensitive Research

The anonymisation focus for deliverables like D3.3 and D8.3 involves a multi-step de-identification process. This is particularly critical for datasets involving sanctioned companies and individuals, where the goal is to analyse trends in cross-border corruption without inadvertently disclosing protected personal data. The

project utilises a combination of data masking, aggregation, and pseudonymisation to ensure that individuals cannot be re-identified through the combination of multiple data points.

Authors are required to submit their anonymisation plans to the project's data manager for review before any data is shared or published. This review process ensures that the techniques used are appropriate for the sensitivity of the data and comply with both GDPR and the specific requirements of the Grant Agreement.

6. Data sharing

6.1 Internal sharing

Internal sharing is facilitated through:

- The Data Hub (D1.2), which gives partners a controlled access to pooled open and restricted datasets under agreed governance.
- Project-internal collaboration platforms (e.g. SharePoint) for sharing non-public data and associated documentation.
- Clear rules, set out in the Consortium Agreement and internal policies, on when data can be shared within the consortium and under what conditions.

6.2 External sharing and embargoes

External sharing is organised in three tiers:

1. **Open data**: ETI, EUCAL, EUROPAM updates and selected Data Hub exports, shared via Zenodo or equivalent, under CC BY, with no access restrictions.
2. **Controlled-access data**: certain anonymised micro-data (e.g. survey datasets, AI training data) may be shared upon request under data-use agreements, provided this does not conflict with consent forms or licences.
3. **Non-shareable data**: raw interview recordings, full elite interview transcripts, non-anonymisable social-media data, and commercially licensed media content remain confidential and are not shared outside authorised teams.

Embargo periods may apply where data underpin forthcoming publications; such embargoes, if used, are documented in the dataset metadata.

7. Repositories for reuse and persistent identifiers

7.1 Preferred repositories and communities

The project uses the following repositories:

- **Zenodo** (with a dedicated BRIDGEGAP community where feasible) as the default for open datasets and code.
- **Subject-specific or institutional repositories** (e.g. dataverses operated by partner universities) where required by institutional policy.
- For specific tools and code, GitHub connected to Zenodo for automatic DOI assignment (as already envisaged by the Lisbon Council for certain datasets).

All deposited datasets receive DOIs and clear version numbers.

7.2 DOI and licence policies

For each dataset intended for long-term reuse, the following are mandatory:

- DOI and recommended citation format.
- Licence (CC BY 4.0 default; more restrictive only where necessary).
- Statement on provenance, processing and limitations.
- Link to relevant BRIDGEGAP deliverable(s).

8. Methods and tools for accessing, analysing and uploading data

8.1 Data access and analysis tools

Partners use a range of established tools, including:

- Statistical and data-science software: R, Stata, SPSS, Python, Excel.
- Text and network analysis: QDA Miner, WordStat, CooRnet (R), topic-modelling libraries (e.g. LDA, NNMF).
- Survey and panel platforms: Qualtrics, Prolific, institutional survey tools.
- Web and social-media APIs and associated Python/R libraries for data collection under platform terms of use.

8.2 Uploading and dissemination tools

- Data Hub interfaces (APIs, ingestion scripts) for integrating datasets like ETI, EUCAL and EUROPAM updates.
- Repository web interfaces and CLIs (Zenodo, institutional dataverses).
- GitHub for code and some data, linked to Zenodo.

9. Data preservation

The long-term sustainability of the BridgeGap project is dependent on the quality of its data stewardship. By utilising Zenodo and the CERN infrastructure, the project guarantees that its data and metadata will remain available for at least 10 years, far outlasting the duration of the project itself.

9.1 Selection and retention periods

The project distinguishes:

- **Core reference datasets** (ETI, EUCAL, EUROPAM updates, key Hub exports), preserved for at least 10 years via Zenodo and/or institutional repositories.
- **Analytical and intermediate datasets** (e.g. specific modelling inputs, derived variables) preserved for at least 5 years after project end, where they are necessary for reproducibility.
- **Highly sensitive personal data** (raw interviews, identifiable social-media data), retained only as long as required for analysis and auditing, then anonymised or securely destroyed, in line with institutional and EU retention rules.

Each partner implements retention schedules consistent with its institutional policies; for example, University of Glasgow commits to at least 10 years for data of long-term value.

9.2 Sustainability of online tools

The DMP recognises that D1.2 (Data Hub), D5.1 (ETI map), D8.1 (EUCAL) and D9.2 (EUROPAM updates) entail ongoing hosting and maintenance costs. Sustainability measures include:

- Clear allocation of post-project responsibilities among partners (e.g. TI-S and SAR for Integrity Watch and EUROPAM, LUISS/UNIPI for EUCAL, YouControl for certain pooled databases).
- Use of open-source components and modular architectures to ease handover and maintenance.
- Designation of repository copies (e.g. full database snapshots on Zenodo) as a fall-back in case interactive services cannot be indefinitely maintained.

9.3 Knowledge Transfer and Policy Impact

The ultimate goal of BridgeGap's data management strategy is to facilitate the transfer of knowledge from academia to policy. By providing a single, user-friendly Data Hub that centralises anti-corruption metrics, the project empowers civil society, the press, and policymakers to hold power to account. The integration of the "Follow the Money" engine and the EUCAL legal database provides these users with the same investigation and analytical tools as the project researchers, truly bridging the gap between evidence and action.

As the project moves toward its conclusion in 2027, the focus will shift toward ensuring that these tools are institutionalised and that the Data Hub continues to be updated by partner organisations. The technical and ethical foundations laid in this second version of the Data Management Plan provide the necessary

framework for this long-term impact, ensuring that the scientific evidence generated by BridgeGap remains a valuable resource for the global anti-corruption community.

10. Ethical issues and code of conduct

Ethical governance continues to follow the principles and procedures described in D11.2 and the Ethics Self-Assessment, reinforced by the appointment of an Ethics Mentor under T11.4. Key elements include:

- Informed consent for all human-subjects research (interviews, surveys, experiments), with clear explanation of data usage, storage, sharing and rights.
- Respect for privacy, confidentiality and non-discrimination, including careful treatment of potentially re-identifiable qualitative material.
- Compliance with GDPR and relevant national legislation, overseen by institutional ethics bodies and data-protection officers.
- Special care in work on security-sensitive topics (weaponised corruption, foreign undue influence), including dual-use assessments and risk mitigation.

11. Allocation of resources and responsibilities

CSD, as the new coordinator and lead of WP11, retains overall responsibility for DMP maintenance and for collecting updated information from partners via structured input requests (as outlined in the “Roadmap for the new DMP”). Each dataset has a designated data owner and contact person, responsible for documentation, repository deposition (where applicable) and compliance with this DMP.

Resource needs for data management (person-months, hosting costs, repository fees where applicable) are covered within existing WP11 and relevant WP budgets, including the redistributed TI-S and affiliate resources identified in the amended Grant Agreement.

12. Conclusions and next steps

The second version of the Data Management Plan (D11.3) photographs a maturing project that has successfully navigated the complexities of multi-country, interdisciplinary data generation. The BridgeGap project has established a robust ecosystem for anti-corruption evidence, grounded in the FAIR principles and a commitment to transparency. The consolidation of 42 indicators into the Data Hub and the integration of the UNODC Statistical Framework model provide a powerful and standardised tool for corruption monitoring.

The successful expansion of the consortium to include 15 new affiliated entities demonstrates the project's capacity to manage large-scale data flows while maintaining strict GDPR compliance. Furthermore, the pioneering work in WP6 on the AI-suitability of public procurement data positions BridgeGap as a leader in the next generation of digital anti-corruption research.

Moving forward, the project must prioritise the full technical synchronisation between the Data Hub and external legal repositories like EUCAL, while continuing to refine its anonymisation and versioning protocols. By adhering to the actionable requirements and technical standards outlined in this DMP, the BridgeGap consortium ensures that its research outputs will remain citeable, reproducible, and impactful for years to come. The journey from evidence to impact is paved with high-quality data, and BridgeGap has built the technical infrastructure necessary to reach that destination.

The plan will continue to be updated and adjusted regularly to align with project progress and evolving data management best practices. The next steps involve finalising the upload of identified datasets to Zenodo and ensuring all data management practices are consistently applied across the consortium.

Between M29 and M48 the DMP will be updated mainly to:

- Incorporate fully empirical versions of D2.1, D3.4, D4.2, D6.6, D7.1–D7.4 and the M36 version of D8.3, with their associated datasets.
- Finalise and document all Zenodo (or equivalent) depositions, ensuring DOIs, licences and metadata are complete.
- Refine long-term sustainability plans for the Data Hub, ETI, EUCAL and EUROPAM once their final project-phase versions are known.

13. Annex. Proposed list of datasets to be featured on Zenodo

Dataset name (proposed)	Source deliverable / WP	Lead partner(s)	Content and scope	Suggested Zenodo status
BRIDGEGAP_ETI_country_indicators_v1_2025	D5.1 – European Transparency Index map and database	SAR, TI-S + 15 TI-S affiliates	Country-level ETI scores, dimensions and underlying indicators for EU and candidate countries, with time coverage as in D5.1 and contributions from national TI chapters.	Already one dataset exists for D5.1; ensure this is the canonical ETI dataset with full documentation and CC BY; update/extend as needed.
BRIDGEGAP_ETI_microdata_national_contributions	D5.1 / T5.4	TI-S + 15 TI-S affiliates	Anonymised, harmonised national-chapter datasets used to build ETI (legal assessments, expert scorings, sub-indices), where licence and confidentiality permit.	Zenodo, CC BY or CC BY-NC, possibly aggregated; consider separate files per country with common schema.
BRIDGEGAP_EUCAL_legal_database_v1_2025	D8.1 – Online Comparative Regulation Database and Map EUCAL	LUISS, UNIPI	Structured database underlying EUCAL (offence definitions, sanctions, limitation periods, liability regimes, enforcement institutions), covering EU27 and candidate countries.	Zenodo, CC BY; provide full codebook and mapping from legal sources to coded variables.

Funded by the European Union

BRIDGEGAP_EUCAL_map_geodata_v1_2025	D8.1	LUISS, UNIPI	Geospatial and ready-to-visualise data (e.g. shapefiles/GeoJSON with EUCAL indicators attached) to reproduce EUCAL maps outside the project site.	Zenodo, CC BY.
BRIDGEGAP_EUROPAM_extended_panel_v1_2025	D9.2 – Updated and extended EUROPAM.eu	SAR	Longitudinal public-accountability indicators added/updated under BRIDGEGAP, in a flat panel suitable for econometric use, with variable-level documentation.	Zenodo, CC BY; coordinate with existing EUROPAM licence and branding.
BRIDGEGAP_DataHub_reference_snapshot_M24	D1.2 – Data Hub	HUJI, TI-S, SAR, YC	A frozen, documented snapshot (or a documented export subset) of the pooled Data Hub at M24, including links and schemas for connected external databases (Integrity Watch, T-Index, CRF, YouControl, etc.).	Zenodo, CC BY for BRIDGEGAP-owned tables; document external licences and, if needed, provide metadata-only records for non-redistributable sources.
BRIDGEGAP_public_procurement_AI_dataset_v1	WP6 – especially D6.1, D6.2, D6.5	UU, KEDGE	An anonymised subset of public procurement data prepared in AI-ready form (features, labels, documentation) as used in the analyses of AI tools'	Zenodo, preferably CC BY; if licence constraints apply, provide feature-engin

Funded by the European Union

			effectiveness/fairness/explainability.	eered, anonymised subset and full description. Zenodo, CC BY (subject to upstream source conditions); include separate file describing how proprietary inputs were combined without redistributing them.
BRIDGEGAP_media_capture_risk_index_dataset_v1	WP7 – D7.1 All-EU media capture risk map	UNIPG	Country-level media-capture risk scores and underlying indicators built from Media Pluralism Monitor, Euromedia Ownership Monitor, financial indicators and other sources.	
BRIDGEGAP_social_media_corruption_coverage_aggregates	WP7 – D7.2– D7.4	UNIPG	Aggregated statistics and topic- or network-level outputs from social-media analyses (e.g. frequency of corruption-related content, network measures), fully anonymised and platform-compliant.	Zenodo, CC BY or CC BY-NC; avoid including raw posts or handles.
BRIDGEGAP_crossborder_corruption_gravity_model_data	WP3 – D3.4 (preparatory work from D3.3)	UU, LC	Panel dataset used for gravity-model estimation of cross-border corruption/illicit financial flows, combining T-Index, Integrity	Zenodo, CC BY; where specific sources are restricted, distribute only derived variables and

Funded by the European Union

			Watch, sanctions data and other sources.	document inputs.
BRIDGEGAP_highlevel_prosecution_events_dataset	WP8 – D8.3 (M36 final version, but already architected)	UNIPI	Event-level data on criminal proceedings involving high-ranking public officials for corruption-related offences (2000–2025), in anonymised, coded form.	Zenodo, CC BY; ensure no personal data beyond what is already public record; provide legal-source references rather than full documents.
BRIDGEGAP_experimental_datasets_integrity_management	WP4 – integrity management experiments	UA, UG	Anonymised experimental datasets on organisational integrity tools, whistleblowing, etc., including treatment conditions and outcome variables.	Zenodo, CC BY or CC BY-NC, depending on consent forms; include full experimental protocols.
BRIDGEGAP_managerial_perceptions_survey_dataset	WP4 / WP6 – surveys with managers (e.g. UA, UG)	UA, UG	Anonymised survey micro-data on managers' perceptions of corruption risks and integrity systems, suitable for secondary analysis.	Zenodo, CC BY-NC or similar; ensure robust anonymisation and removal of indirect identifiers.
BRIDGEGAP_website_analytics_aggregated_statistics	WP10 – dissemination analytics	LC	Aggregated Matomo statistics on website visits, geographical distribution, and access to data/tools over the project lifetime.	Zenodo, CC BY; no personal data; useful for open-science

				and impact reporting.
BRIDGEGAP_training_and_MOOC_materials_dataset	WP10 – MOOC and training outputs (D10.9)	KEDGE, CSD, LC	Structured package of MOOC datasets (if any embedded exercises use open data), plus scripts and example data used in training materials.	Zenodo, CC BY; ensure that any third-party material remains excluded or properly licensed.

Items 1 and 6–8 correspond most directly to GA-defined “DATA” and “OTHER” deliverables (D5.1, D8.1, D9.2, D1.2). Items 7–13 are not deliverables per se but are natural candidates for open or controlled-access datasets in line with the Roadmap and D11.2’s open-science commitments.