Funded by the European Union

EuropeanCity2	
Project number:	**101178170**
Project name:	European City Squared: Computational Social Science Simulation for Democracy
Topic:	HORIZON-CL2-2024-DEMOCRACY-01-06
Type of action:	HORIZON-RIA
Starting date of action:	01/01/2025
Project duration:	36 months
Project end date:	31/12/2027
Deliverable number:	D1.2
Deliverable title:	Data Management Plan
EC document version:	v1
WP number:	WP1
Lead beneficiary:	1-AU
Main author(s):	Kristoffer Nielbo (AU)
Internal reviewers:	Per Møldrup-Dalum (AU)
Nature of deliverable:	R
Dissemination level:	PU
Delivery date from Annex 1:	30 June 2025 (M6)
Actual delivery date:	30/06/2025

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor European Research Executive Agency (REA) can be held responsible for them.

Abbreviations

D	Deliverable
EC	European Commission
WP	Work Package
WT	Work Task

1 Data Summary

- **Purpose:** The datasets described are directly aligned with the EC2 project objectives: simulation of democratic processes, evaluation of voting mechanisms (including quantum and quadratic voting), and integration of classical and synthetic data to enhance collective decision-making.
- **Types of Data:**
  - Real-world data from pilot studies.
  - Synthetic data generated via LLMs.
  - Agent-based simulation outputs.
  - Quantum simulation data.
  - Event logs and derived preference graphs.
- **Formats:**
  - Structured text: JSON, CSV.
  - Documentation and metadata: Markdown, PDF.
- **Expected Data Size:** Estimated to range across datasets from 200 MB to 10 GB, totaling less than 1TB across the project lifecycle.
- **Generation and Reuse:**
  - Real-world data collected in Aarhus and Basel.
  - Synthetic data generated using LLMs (Llama 3.3 instruct-tuned model in 70B size).
  - Simulation outputs from internal computational models (agent-based and quantum).
- **Usefulness:** The data supports computational research in political science, quantum social choice, and civic technology. Data may be reused by researchers in CSS, governance, and AI ethics.
- **Metadata Specification:** Detailed metadata for each dataset is available in the project's structured data specification (data-collections.csv and associated files).
  - **Example Documentation:** Format and provenance documentation are available for each dataset. See, e.g., `LLM_Synthetic_Preferences_data_format.md` and `LLM_Synthetic_Preferences_data__Outputs_provenance.md`.
- **Repository:** The living DMP and data specification files are maintained at: https://github.com/centre-for-humanities-computing/dmp-EC2

2 FAIR Data

2.1 Making Data Findable

- Persistent identifiers (DOIs) will be assigned via Zenodo.
- Metadata records linked from project outputs and public repositories.

2.2 Making Data Accessible

- Data will be stored on AU-SIF (sensitive data) and AU-ERDA (open data).
- Metadata standards consistent with the DCAT v3 schema will be used.
- Public datasets will be deposited in Zenodo with open licensing.
- Sensitive data will be accessible upon request under joint controller agreements.

2.3 Making Data Interoperable

- Use of standard file types (CSV, JSON).
- Metadata vocabularies consistent with RDF standards (DCAT v3).
- Encoding practices follow UTF-8 and binary storage for high-performance data if necessary.

2.4 Increasing Data Reuse

- Open access datasets licensed under CC-BY 4.0.
- Data and metadata will be maintained for at least five years after project conclusion.
- Documentation of processing and provenance is provided per dataset (see, e.g., `LLM_Synthetic_Preferences_provenance.md`).

3 Other Research Outputs

- Public simulation software and model code.
- Policy briefs and guidelines.
- Academic publications and open peer-reviewed outputs.
- Metadata specifications and markdown-based data format profiles.

4 Allocation of Resources

- Data storage infrastructure provided by AU-SIF and AU-ERDA.
- Cloud and HPC compute resources allocated via DeiC and LUMI.
- IBM Quantum used under academic access for quantum simulations.
- Data management responsibilities led by AU with support from consortium.

5 Data Security

- Sensitive data stored on ISO-27001-certified AU-SIF and DeiC Interactive HPC infrastructures.
- Secure access controls via institutional authentication (WAYF) and audit logs.

- Anonymization and pseudonymization applied as needed.

6 Ethics

- Legal basis for sensitive data collection established via joint controller agreements.
- Open simulation data does not require a legal basis for processing personal data for research purposes.
- Ethical review managed through WP11 and monitored by Ethics Advisory Board.
- All processing activities documented in data-provenance.csv.

7 Other Issues

- Continued alignment with AI Act and GDPR.
- Monitoring of emerging standards for LLMs and synthetic data.
- Long-term sustainability secured through ERDA/Zenodo archiving strategies.