



Baltic Engagement
Centre for Combating
Information Disorders

BECID DATA MANAGEMENT PLAN

DECEMBER 2025

THIS PROJECT WAS FUNDED BY THE BALTIC ENGAGEMENT CENTRE FOR COMBATING INFORMATION DISORDERS (BECID) 2025–2028; PART OF D1.4 OF THE EU-FUNDED PROJECT WITH THE ID: 101226242, COORDINATED BY THE UNIVERSITY OF TARTU (UTARTU).

MARJU HIMMA | PRIMARY INVESTIGATOR

GRETEL JUHANSOO | PROJECT MANAGER

... in co-creation with the whole hub and its stakeholders.



Co-funded by the
European Union

1. Data Summary

The BECID project will generate and collect the following types of data:

Primary data:

- Platform-specific data accessed via APIs (Facebook, Instagram, LinkedIn, Snapchat, TikTok, YouTube, X/Twitter, Telegram).
- Interviews with key stakeholders (e.g. policymakers, journalists, researchers, practitioners).
-

Secondary data: Anonymised statistical data exports from cleanrooms (e.g. Meta Researcher Platform).

Generated data: Analytical outputs, reports, statistical tables, visualisations, and Virtual Interns' automated outputs.

Training and capacity-building data: Materials for repository training, online workshops, and media literacy education.

Formats: JSON, XML, CSV (raw data); MP3/WAV (audio recordings); DOCX/PDF (transcripts); PDF, XLSX, HTML (analytical outputs).

Estimated volume: API-based data (variable); interview recordings and transcripts (50–200 GB depending on scope).

2. FAIR Principles (with interviews added)

Findable. Interview transcripts will be pseudonymised and catalogued in the University of Tartu secure repository (OwnCloud) with restricted access metadata. Metadata (non-sensitive) will follow Dublin Core/Datacite for discoverability.

Accessible. Interview recordings and full transcripts: Stored confidentially; access restricted to authorised researchers in BECID. Anonymised excerpts or coded transcripts: May be shared with external collaborators upon ethics approval. Participants are informed that their data will be stored securely, used only for research, and never disclosed in identifiable form.

Interoperable. Transcripts stored in DOCX/PDF with coding in NVivo/Atlas.ti or equivalent, exportable to interoperable formats (e.g. CSV for coded data).

Reusable. Only pseudonymised/anonymised data may be reused, and only with

THIS PROJECT WAS FUNDED BY THE BALTIC ENGAGEMENT CENTRE FOR COMBATING INFORMATION DISORDERS (BECID) 2025–2028; PART OF MS4 OF THE EU-FUNDED PROJECT WITH THE ID: 101226242, COORDINATED BY THE UNIVERSITY OF TARTU (UTARTU).

participants' consent and under controlled conditions.

Following the guidance of the BECID2 Ethics Report (D1.3) and the project's glass house approach, only non-sensitive outputs (incl. analyses, reports, summaries) will be shared openly via the project website.

Notion is used as an internal collaboration and documentation platform for project coordination. It is **not** used for long-term data preservation or public data sharing. All shareable outputs are deposited in appropriate institutional or open repositories.

3. Data Collection and Management

Interviews:

1. Conducted with informed consent, explicitly stating that data will be stored confidentially and used solely for research.
2. Audio recordings (MP3/WAV) will be uploaded immediately after collection to the University of Tartu secure cloud (OwnCloud), encrypted at rest.
3. Transcripts prepared and pseudonymised (removal of names, locations, identifiers).
4. Original recordings retained for verification but accessible only to the core research team.

Access policy:

1. Raw audio: Access restricted to PI and interview team.
2. Transcripts: Pseudonymised versions accessible to the wider BECID research team.
3. Shared datasets: Only anonymised excerpts and coded datasets may be shared outside the team.
4. For structured datasets (e.g. fact-checking databases), machine-readable formats and standardised metadata schemas will be prioritised to support interoperability and future reuse by researchers and practitioners.

4. Data Sharing Policies

Restricted interviews:

1. No direct sharing of audio recordings or full transcripts.
2. Only pseudonymised, aggregated insights will be made publicly available.
3. If transcripts are to be archived for reuse, this will only happen under explicit additional consent and with restricted repository access.

5. Data Security

Interviews:

1. Stored on University of Tartu secure cloud (OwnCloud) with **role-based access**



control.

2. Audio files encrypted; transcripts pseudonymised before storage.
3. Transfer of files via secure university channels only.
4. Retention policy: Interview data retained for project duration + 5 years, then deleted unless archiving is consented.

6. Ethical and Legal Aspects

- All interviews conducted under **informed consent** (written or recorded).
- Participants informed about data use, confidentiality, and right to withdraw.
- GDPR compliance ensured by pseudonymisation and secure storage.
- Ethics approval obtained through University of Tartu procedures.
- Platform- and API-based data will be handled in aggregated or pseudonymised form. While full anonymisation cannot always be guaranteed, the project will not seek to identify or profile individual users, and all data processing will comply with GDPR and platform terms of service.

7. Allocation of Resources

- Interview management coordinated by WP leaders.
- Secure storage provided by University of Tartu secure service (OwnCloud).
- Training for interviewers in data protection and ethical handling of sensitive material.

8. Data Governance, Responsibilities and Updates

- The University of Tartu acts as the data controller for data collected within BECID2. WP leaders are responsible for data quality and compliance within their tasks. The PI oversees DMP updates and compliance with requirements.
- This Data Management Plan is a living document and will be updated as the project progresses, particularly when new data types, repositories, or sharing conditions emerge.

As of December 2025, EDMO Central has not yet launched its new funding round, and the future structure of its communication channels is therefore not fully defined. As communication channels within the EDMO network evolve, the plan will be reviewed and updated accordingly. Any updates will be communicated via the project website and documented in the annual reports.