

Project BoLI – Data Governance Policy

Bolstering Linguistic Resources in the Indigenous Languages of India

Attribution

This document is authored and published by:

UnReaL-TecE LLP
and
Council for Diversity and Innovation (CoDIN)

In connection with the BoLI initiative.

License

This document is made available under the
Creative Commons Attribution 4.0 International License (CC BY 4.0).

You are free to:

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially

Under the following terms:

- Attribution — You must give appropriate credit to *UnReaL-TecE LLP* and *CoDIN*, provide a link to the license, and indicate if changes were made.
License link: <https://creativecommons.org/licenses/by/4.0/>

Machine-readable attribution:

CC BY 4.0 — © 2025 UnReaL-TecE LLP & Council for Diversity and Innovation (CoDIN)

Document Version

Version: 1.0 (2025)
Part of the BoLI Guidelines



Project BoLI – Data Governance Policy

(Issued by UnReal-TecE LLP and Council for Diversity and Innovation on behalf of the BoLI Consortium)

1. Purpose and Scope

This Policy governs the ethical, legal, and technical management of linguistic data collected, processed, stored, and shared under **Project BoLI**.

It applies to all partners—including linguists, community members, and affiliated researchers—engaged in data creation, annotation, storage, and distribution within the BoLI network.

2. Guiding Principles

BoLI's data governance is grounded in the following principles:

- **Fairness and Transparency** – equitable recognition, compensation, and attribution to linguists and speakers.
- **Community Rights and Consent** – respect for linguistic rights, cultural sensitivities, and informed consent of all contributors.
- **Accountability and Oversight** – continuous audit, quality assurance, and redressal mechanisms for ethical and regulatory compliance.
- **Open and Responsible Access** – commitment to openness through Creative Commons licensing balanced with controlled access for sensitive data.
- **Non-Exclusivity** – datasets remain co-owned by linguists and speech communities; BoLI acts as a non-exclusive distribution and visibility platform.

3. Data Lifecycle Governance

3.1 Collection and Consent

- Data must be collected directly by approved linguists or their associates from speakers with full, informed, and documented consent.
- All contributors must be compensated fairly under mutually agreed terms.
- No subcontracting or secondary data collection is permitted without prior written approval.



- Ethical protocols and cultural norms prescribed by BoLI must be observed at every stage of fieldwork.

3.2 Processing and Annotation

- All audio recordings shall be transcribed in IPA and in the most widely used or official script, inter-linearly glossed following Leipzig Glossing Rules, and translated into English or another relevant language.
- Linguists are responsible for ensuring data quality, completeness, and adherence to BoLI annotation standards.
- BoLI may review, revise, or reject datasets based on compliance and quality assessment.

3.3 Storage and Access

- Accepted datasets are hosted securely on UnReal-TecE servers and mirrored on selected public repositories for visibility.
- Controlled access is maintained through the LiFE App or equivalent platforms.
- Access logs and audit trails will be preserved for accountability and traceability.

3.4 Licensing and Intellectual Property

- All datasets are released under **Creative Commons CC BY-NC-SA 4.0**.
- Ownership of intellectual property remains with the original speakers, speech communities, and linguists.
- BoLI functions as a non-exclusive distributor, enabling broader discovery and reuse under agreed license conditions.
- Commercial and non-commercial reuse requests are handled jointly by the primary stakeholders and BoLI through transparent agreements.

3.5 Retention, Removal, and Exit

- Rejected datasets are securely deleted within six months.



- Contributors may withdraw or remove datasets at any time by written notice.
- Upon exit or termination, datasets remain accessible to existing authorized users under the license in force but are withdrawn from future sales or redistribution.
- Removed datasets are retained only for prior licensees and archival integrity.

4. Roles and Responsibilities

- **UnReal-TecE LLP** – Custodian of the BoLI repository; ensures system security, training, ethical compliance, and public communication.
- **Linguists** – Data stewards responsible for ethical collection, consent, transcription, and metadata accuracy.
- **Community Members / Speakers** – Rights holders and co-owners entitled to fair remuneration and royalties.
- **BoLI Network Council** – Oversight and advisory body representing linguists, communities, and UnReal-TecE; approves amendments and audits.

5. Audits, Compliance, and Review

- Periodic audits will verify data privacy, consent documentation, and ethical compliance.
- Corrective action will be initiated upon detection or notification of non-compliance.
- Violations may result in suspension or termination of network membership.
- The Policy shall be reviewed annually, incorporating member feedback and regulatory updates.

6. Amendments and Dispute Resolution

- Amendments to this Policy may be proposed through a written request submitted to the BoLI Network Council.



- Such a proposal must be endorsed by a simple majority of the BoLI network members, evidenced through signed or digitally verified consent.
- The proposed amendment shall be reviewed and ratified by the BoLI Network Council, and will come into effect upon such ratification,
- Disputes shall be resolved through internal arbitration as specified in the BoLI Terms and Conditions.

7. Effective Date and Applicability

This Policy takes effect upon approval of BoLI membership and remains binding throughout participation in the network. All data handling under BoLI must conform to this Policy, in addition to applicable national and international data protection laws.

Issued by:

UnReaL-TecE LLP and Council for Diversity and Innovation

For: Project BoLI – Bolstering Linguistic Resources for Indian Languages