



Language Data Sovereignty Agreement

Between [the Nation] and [the Implementing Partner] concerning the Nation's language data, cultural protocols, and AI systems.

This is a template, not legal advice. It is a starting point for a language authority and tribal counsel to adapt. Bracketed text is for the Nation to complete. Where the Nation has its own data governance code, research review process, or tribal IRB, that code governs and this agreement should be conformed to it. Codetalkers.io publishes this template so that any nation can use it, whether or not Codetalkers is involved in the implementation.

Recitals

The Nation is a sovereign government with inherent authority over its language, its cultural knowledge, and all information about its citizens. The Nation's language is the collective property and heritage of the Nation and its people. The Implementing Partner wishes to assist the Nation in deploying digital tools to preserve, teach, and create in the Nation's language, and both parties intend that every such tool be governed by the Nation. This agreement records that intent as enforceable terms, guided by the CARE Principles for Indigenous Data Governance and the First Nations principles of OCAP®.

1. Definitions

- 1.1 **Language Data** means every recording, transcript, text, image, video, lexicon entry, lesson, story, song, annotation, and metadata concerning the Nation's language that is placed in, created by, or derived from the Toolkit, including all corrections and approvals.
- 1.2 **Derived Data** means anything computed from Language Data, including transcripts, embeddings, indexes, search results, and Models.
- 1.3 **Model** means any machine learning model, fine-tune, adapter, or weights trained, tuned, or adapted using Language Data in whole or in part.
- 1.4 **Language Authority** means the body the Nation designates in Schedule A to govern Language Data under this agreement.
- 1.5 **Vault** means the Nation's isolated data store for Language Data: its own database, file storage, and encryption keys, as described in Schedule B.

- 1.6 **Cultural Protocol** means a rule set by the Language Authority that governs who may access an item of Language Data, when, and for what purposes, including whether it may be processed by an AI Service.
- 1.7 **Sacred Material** means Language Data the Language Authority designates under the Sacred tier of Schedule C, or any tier the Authority marks as excluded from AI processing.
- 1.8 **AI Service** means any software that performs transcription, translation, speech synthesis, tutoring, generation, summarization, search, or similar functions using machine learning, whether hosted by the Implementing Partner, a third party, or the Nation.
- 1.9 **Toolkit** means the software, configuration, and documentation described in Schedule D as deployed for the Nation.

2. Ownership

- 2.1 The Nation owns all Language Data, all Derived Data, and all Models, from the moment of creation. Ownership does not depend on who recorded, transcribed, uploaded, or processed the material.
- 2.2 The Implementing Partner acquires no ownership, copyright, license, or other interest in Language Data, Derived Data, or Models except the limited, revocable permission in Section 2.3. The Implementing Partner will not register, assert, or claim copyright or any other intellectual property right in Language Data, Derived Data, or Models, and waives any such claim that might otherwise arise by operation of law.
- 2.3 The Nation grants the Implementing Partner permission to access and process Language Data solely to provide the services in Schedule D, subject to Cultural Protocols, and only for so long as this agreement is in force and the permission has not been revoked under Section 9.
- 2.4 Individual speakers and contributors retain whatever personal and moral rights the Nation's law and custom recognize. Nothing in this agreement transfers a speaker's rights to the Implementing Partner.
- 2.5 Where the Implementing Partner develops general software improvements while serving the Nation, those improvements may be reused for other nations only if they contain no Language Data, Derived Data, or Nation-specific configuration.

3. The Language Authority

- 3.1 The Nation designates the Language Authority in Schedule A. The Nation may change the designation at any time by written notice.
- 3.2 The Language Authority has sole power to: set and change Cultural Protocols; assign tiers and roles to items and people; approve or reject any AI-generated entry before it enters the record; designate Sacred Material; direct export or deletion; and revoke the Implementing Partner's access.

- 3.3 The Implementing Partner will act on Language Authority instructions within [5] business days, or immediately for revocation and for removal of Sacred Material from any system where it should not be.

4. Cultural Protocols

- 4.1 Every item of Language Data carries a Cultural Protocol. Items without an assigned protocol default to the most restrictive tier until the Language Authority assigns one.
- 4.2 Cultural Protocols are enforced in the Vault's data and storage layer, so that no application, agent, or AI Service can read an item its protocol excludes. Interface-only restrictions do not satisfy this section.
- 4.3 The Toolkit supports the tiers in Schedule C and must allow the Language Authority to add, rename, and combine tiers, including role-based, seasonal, and named-individual restrictions.
- 4.4 The Toolkit allows the Language Authority to attach Local Contexts Traditional Knowledge Labels to any item, and preserves those labels on export.

5. Hosting and Possession

- 5.1 The Nation chooses where the Vault is hosted: on infrastructure the Nation operates, or in a cloud region the Nation selects. The choice is recorded in Schedule B and may be changed by the Nation on [60] days' notice, with the Implementing Partner's assistance in migration.
- 5.2 The Vault is dedicated to the Nation. Language Data is never stored in a database, bucket, index, or model shared with any other nation or customer.
- 5.3 The Nation holds the encryption keys to the Vault. Sacred Material is encrypted under a separate key held only by the Nation and the individuals the Language Authority names. The Implementing Partner and all AI Services are never given that key.
- 5.4 The Implementing Partner will not move Language Data outside the hosting location in Schedule B, including for backup, support, or debugging, without written approval of the Language Authority.

6. AI Services and Models

- 6.1 Only Language Data the Language Authority marks as AI-eligible may be processed by an AI Service. Sacred Material is never used for training, fine-tuning, indexing, embedding, retrieval, prompting, evaluation, or testing of any AI Service, anywhere.
- 6.2 No third-party AI vendor may retain, log, or train on Language Data. The Implementing Partner will use hosted AI Services only under written zero-data-retention terms listed in Schedule D, and will offer the Nation a self-hosted option that sends nothing off site.

- 6.3 Every Model is the Nation's property under Section 2. The Implementing Partner will deliver Model weights and training configuration to the Nation on request and on termination.
- 6.4 Output from any AI Service is a draft. It enters the language record only after a fluent speaker designated by the Language Authority approves it, and the record retains the approver's name and date. AI Services must ground generated language in approved lexicon entries and flag, rather than present, any output that cannot be traced to one.
- 6.5 Where an item is deleted or its protocol is changed to exclude AI processing, the Implementing Partner will remove it from every index and embedding within [10] business days and will retrain or retire any Model that was trained on it within [90] days, or sooner if the Language Authority directs.

7. Access, Audit, and Transparency

- 7.1 The Vault keeps a log of every access to Language Data, by person or system, with time and purpose. The Language Authority can read the full log at any time through the Toolkit and export it.
- 7.2 The Implementing Partner maintains a transparency page for the deployment listing what is stored, where, which AI Services and Models are in use, and who has administrative access, and updates it within [10] business days of any change.
- 7.3 Before Sacred Material or any role-restricted tier is enabled in the Vault, the protocol enforcement will be reviewed by an independent reviewer acceptable to the Language Authority, at the Implementing Partner's expense, and the review report delivered to the Nation.
- 7.4 The source code of the protocol enforcement layer is made available to the Nation and to any reviewer the Nation names.

8. Speakers, Compensation, and Attribution

- 8.1 Speakers, elders, and other contributors are compensated for recordings, reviews, and approvals at the rates in Schedule E, paid by [the Nation / the Implementing Partner].
- 8.2 Each entry in the record credits the contributors who created and approved it in the manner the Language Authority considers appropriate, which may include withholding a name where custom requires.
- 8.3 The Implementing Partner will not publish, present, or describe Language Data, or identify the Nation as a customer, without the Language Authority's written approval.

9. Export, Deletion, and Revocation

- 9.1 The Nation may export all Language Data, Derived Data, Models, protocols, logs, and configuration at any time, without stating a reason, in open documented formats, through a

control on the Language Authority dashboard and on written request. The Implementing Partner will complete any export within [15] business days.

- 9.2 The Nation may direct deletion of any or all Language Data at any time. Deletion includes Derived Data and Models under Section 6.5, and all backups within [30] days, and the Implementing Partner will certify completion in writing.
- 9.3 The Language Authority may revoke the Implementing Partner's access to the Vault at any time, immediately, through the dashboard or by written notice, with or without terminating this agreement. Revocation does not require cause.
- 9.4 On termination for any reason, the Implementing Partner will assist in transition for [90] days, deliver everything in Section 9.1, then delete every copy in its possession under Section 9.2.

10. Confidentiality and Security

- 10.1 The Implementing Partner treats all Language Data as confidential regardless of tier, applies the security measures in Schedule B, limits staff access to named individuals approved by the Language Authority, and requires each of them to sign a personal confidentiality undertaking.
- 10.2 The Implementing Partner will notify the Language Authority of any unauthorized access, loss, or disclosure within [24] hours of discovery, and will cooperate fully in response.
- 10.3 If a court, agency, or other third party demands Language Data from the Implementing Partner, the Implementing Partner will notify the Nation immediately, will assert the Nation's ownership and sovereignty, and will disclose nothing beyond what it is finally compelled to disclose.

11. Term

- 11.1 This agreement begins on the date of the Nation's authorizing resolution referenced below and continues until terminated by either party on [90] days' written notice, or by the Nation immediately under Section 9.
- 11.2 Sections 2, 6.1, 6.3, 8.3, 9, 10, and 12 survive termination.

12. Sovereignty, Governing Law, and Disputes

- 12.1 Nothing in this agreement waives, limits, or is a submission of the Nation's sovereign immunity, in whole or in part.
- 12.2 This agreement is governed by the laws of the Nation. Disputes will be resolved in the Nation's courts or by the dispute process the Nation designates in Schedule A, after good-faith discussion between the Language Authority and the Implementing Partner.
- 12.3 Where a term of this agreement conflicts with the Nation's data governance code or research policy, the Nation's code or policy controls.

Authorization and Signatures

This agreement is authorized by [Council / Governing Body] Resolution No. [] dated [] , a copy of which is attached.

For the Nation · Name, title, date

For the Language Authority · Name, title, date

For the Implementing Partner · Name, title, date

Schedule A - Language Authority and Dispute Process

ITEM	NATION TO COMPLETE
Language Authority (body)	[e.g., the Nation's Language Department; a Language and Culture Committee; a council of designated elders]
Chair or primary contact	[Name, title, contact]
Designated fluent speakers for approval (Section 6.4)	[Names, or "as designated by the Authority in the Toolkit"]
Dispute process (Section 12.2)	[Nation's court; mediation process; other]
Applicable Nation data governance code or research policy (Section 12.3)	[Title and date, or "none at this time"]

Schedule B - Hosting, Keys, and Security

ITEM	NATION TO COMPLETE
Hosting location	[Nation-operated infrastructure at ____ / cloud provider ____, region ____]
Vault components	Dedicated database; dedicated file storage; dedicated encryption keys; dedicated AI index. None shared with any other nation or customer.
Application key holder	[Nation IT; or Implementing Partner under Authority approval]
Sacred Material key holder(s)	[Named individuals designated by the Authority. Never the Implementing Partner.]
Backup location and retention	[Location within the hosting choice above; retention ____ days]
Security measures	Encryption at rest and in transit; role-based access enforced in the data layer; multi-factor authentication for all administrative access; access logging under Section 7.1; annual security review.

Schedule C - Cultural Protocol Tiers

A starting set. The Language Authority may add, rename, or combine tiers at any time.

TIER	WHO MAY ACCESS	EXAMPLES	AI PROCESSING
PUBLIC	Anyone	Greetings, numbers, place names the Nation chooses to share, outreach material	Permitted if marked AI-eligible
COMMUNITY	Enrolled members and approved learners	Everyday vocabulary, family stories, classroom recordings	Permitted within the Nation's own Models only
ROLE	A defined group: women, men, a clan, a society, initiated people, elders	Gender-specific teachings, society songs, clan knowledge	Only within that role's scope; default not permitted
SEASONAL	Community, only in the designated season	Winter-only stories, ceremonial-calendar material	As its base tier, gated by date
SACRED	Named individuals set by the Authority	Ceremonial language, prayers, names of the deceased where custom requires, anything the Authority designates	Never. Separate key under Section 5.3.

Schedule D · Services, AI Services, and Vendor Terms

ITEM	TO COMPLETE
Services the Implementing Partner provides	[e.g., deployment of the Vault and protocol engine; transcription pipeline; Authority dashboard; training for Nation staff; support hours]
Hosted AI Services in use, with zero-retention terms	[Vendor, service, date of written zero-data-retention terms, copy attached]
Self-hosted AI Services in use	[Model name and version, where it runs]
Models owned by the Nation under this agreement	[Name, purpose, training data tiers, date]
Protocol enforcement source code location	[Repository or delivery method]

Schedule E · Compensation and Attribution

ITEM	NATION TO COMPLETE
Speaker recording rate	[per hour / session]
Review and approval rate	[per hour / per entry]
Attribution practice	[Name shown on each entry; name withheld where custom requires; other]
Payer	[Nation / Implementing Partner / grant source]

Language Data Sovereignty Agreement · Template v1.0 · Published by Codetalkers.io as part of the Language Toolkit (codetalkers.io/language-toolkit). Free for any nation to adapt. Informed by the CARE Principles for Indigenous Data Governance (Global Indigenous Data Alliance) and the First Nations principles of OCAP® (First Nations Information Governance Centre). Not legal advice; review with tribal counsel.