21 March 2025

English

United Nations Group of Experts On Geographical Names
2025 session
New York, 28 April – 2 May 2025
Item 4 (b) **of the provisional agenda** *
Reports: Governments on the situation in their countries and on the progress made in the standardization of geographical names

Report of the Republic of India

Submitted by the Republic of India**

Summary

India is a multilingual nation with 22 scheduled languages (Included in the Eighth Schedule of the Constitution of India), 99 non-scheduled languages (with more than 10,000 speakers in India) and several other languages (with less than 10,000 speakers in India) and dialects, representing the complex socio-cultural structure of the country.

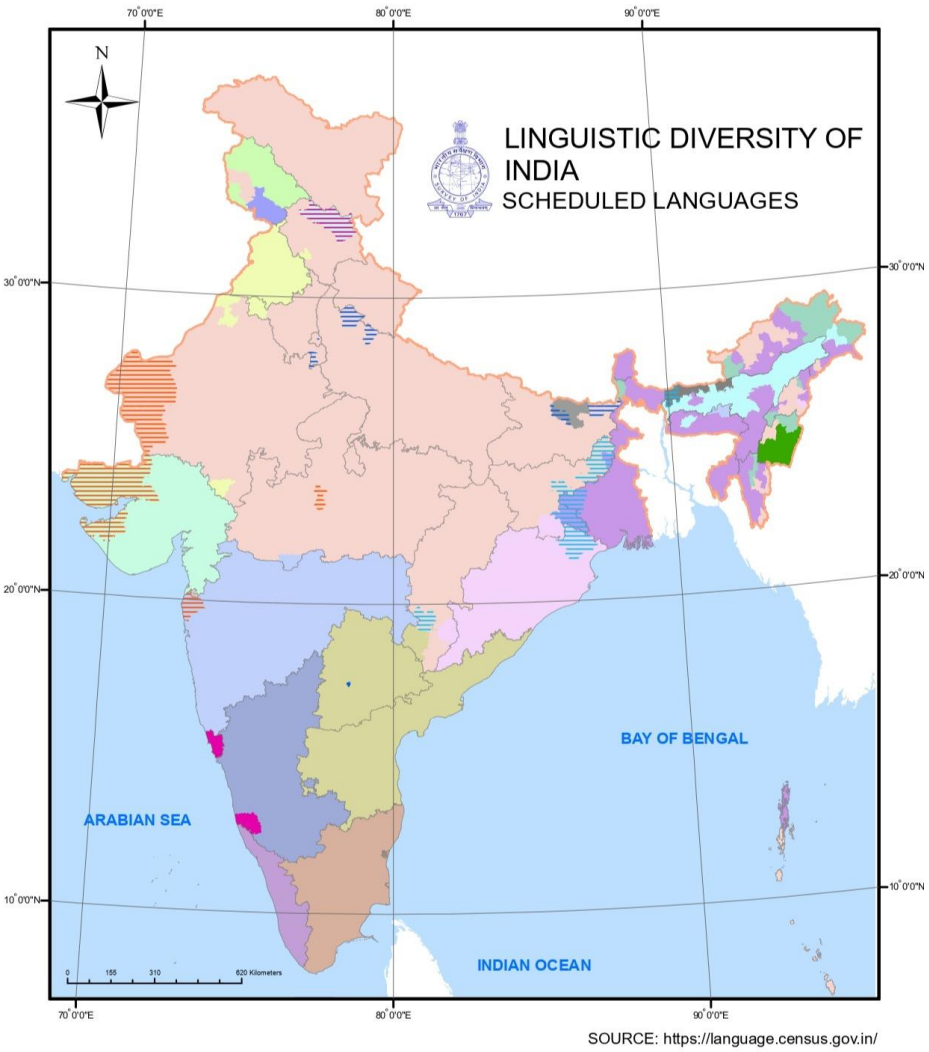
The Ministry of Home Affairs, Government of India is the National Names Authority and Survey of India, the National Geospatial and Mapping Organisation, is responsible for the transliteration of the geographical names from Vernacular Scripts to Roman and Devanagari Scripts, thereby standardizing the place names in English and Hindi languages. Further, the National Geospatial Policy, 2022 of India has entrusted the Survey of India for the implementation of the global fundamental geospatial data theme of Geographical Names.

This report will outline the complex linguistic population distribution in India (indicative of the rich history and cultural legacy of India), the roles and responsibilities of various stakeholder government agencies in the process of standardization of Geographical Names, Survey of India's initiatives in the creation and maintenance of the database of Geographical Names in all the 22 scheduled languages of India – the traditional methodology, current practice in digital environment for collection of audio bytes and the future roadmap.

* GEGN.2/2025/1
** Prepared by Survey of India, Department of Science & Technology, Government of India

Introduction

India's linguistic landscape is among the most intricate and diverse in the world, with a rich tapestry of languages that reflects its cultural, historical and geographical complexity. Languages in India belong to several language families, the major ones being the Indo-European languages, spoken by around 77% of Indians and the Dravidian languages, spoken by 21% of Indians. Languages spoken by the remaining 2% of the population belong to the Austro-Asiatic, Tibeto-Burmese, Semito-Hamitic and a few other minor language families and isolates each with deep historical and cultural underpinnings. With 121 languages and 19,500 dialects, including 22 constitutionally recognised languages under the Eighth Schedule, India represents a microcosm of global linguistic plurality.



*

[ ]: *Most spoken scheduled language district-wise*

[ ]: *Maximum speakers of Santhali, Sanskrit, Sindhi and Urdu scheduled languages (which are not most spoken scheduled language district-wise)*

Fig-1: Linguistic diversity of India: Scheduled Languages

Although the official language of the Union of India is Hindi in Devanagari script (as per Article 343), the Article 345 of the Constitution allows individual States to adopt, by law any one or more of the languages in use within the State or Hindi as the official language(s) for official purposes of that State. In line with the Articles 344(1) and 351 of the Eighth Schedule, the Indian Constitution includes twenty-two (22) scheduled languages (Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu, Urdu, Bodo, Santhali, Maithili and Dogri) and a total of ninety-nine (99) non-scheduled languages. In addition to this, several other languages and dialects exist, representing cultures and complex systems of knowledge developed and accumulated over thousands of years, reflecting distinctive worldviews and perspectives of the indigenous population. Hence it is no surprise that correspondingly, the geographical names in India are also rich and varied. Having a map of a region in its local language enhances the understanding of its topography, heritage, and cultural significance.

Hindi, the most widely spoken language functions as one of the official languages of India alongside English, which serves as a vital link facilitating administration, education and commercial exchange. The coexistence of multiple languages has necessitated a framework of multilingual governance and policy to balance linguistic inclusivity and national cohesion. This linguistic heterogeneity, while enriching India's cultural fabric, also poses challenges in governance, interregional communication and representation, requiring policies and methods that promote both preservation and adaptability.

Roles and Responsibilities of Stakeholders

The Government of India has laid down the principles and procedures with respect to the spelling of geographical names in India. The spelling of any geographical name in the script used as the official language of the centre i.e. Hindi and English is determined by the Government of India through the Survey of India (SoI), the National Geospatial and Mapping Agency. State Governments have been vested with the authority to name geographical places of their area in their respective vernacular script and will consult SoI for determining their spelling in Roman and Devanagari scripts. SoI is responsible for the transliteration of geographical names from vernacular scripts to Roman and Devanagari Scripts. This ensures the standardization of geographical names in English and Hindi. Other government departments are advised to consult SoI for accurate spellings of new installations like railway stations, aerodromes, etc. The Ministry of Home Affairs, Government of India, serves as the National Names Authority under the Government of India (Allocation of Business) Rules, 1961, and is responsible for overseeing any changes to Geographical Names.

Fig-2: Railway Station name in three languages viz. Gujarati, Hindi and English

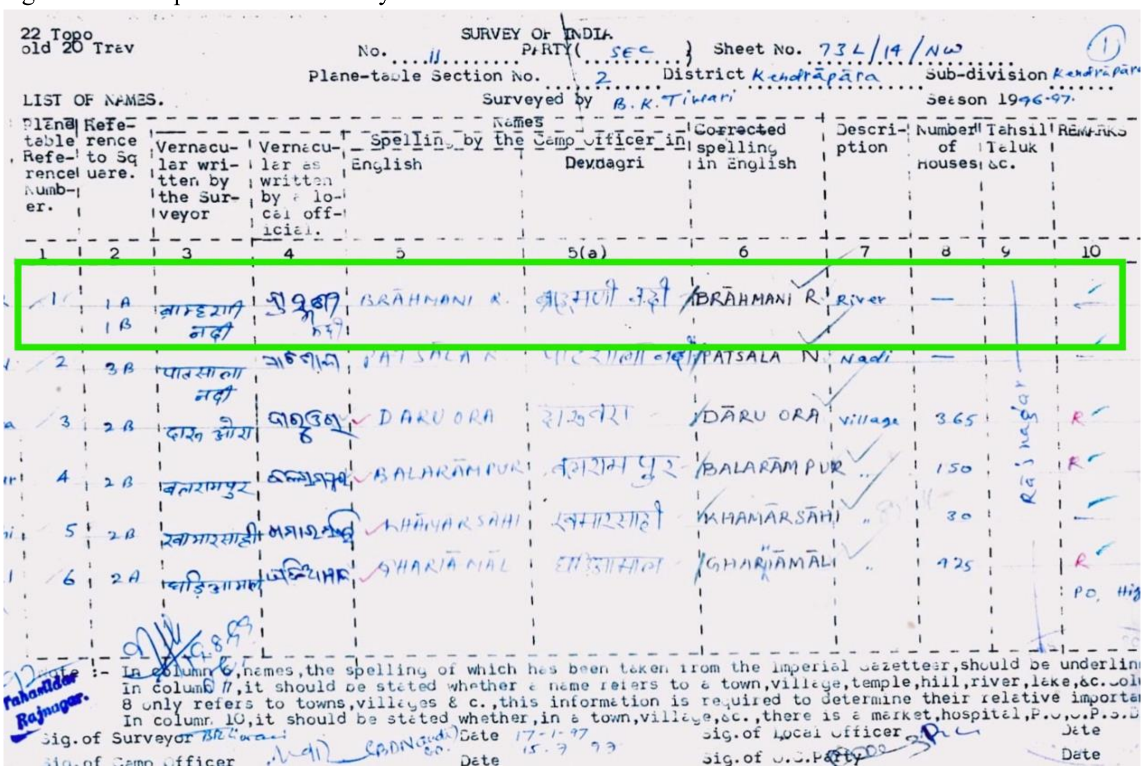
The Government of India published the National Geospatial Policy, in December 2022 and encompassed the fourteen fundamental data themes recognized by UN-GGIM, as themes fundamental to the development of a country's Integrated

Geospatial Information Infrastructure and supporting the Sustainable Development Goals. The policy has assigned each of these fourteen fundamental data themes to a nodal agency under the appropriate ministry. Along with the data themes of Geodetic Reference Frame, Ortho-imagery, Elevation and Functional Areas, the creation and maintenance of Geographical Names along with the development of national standards for the same has been entrusted with the SoI.

Collection of Geographical Names

The initial exercise for the creation of a geographical name database for the country was started by SoI during 1960s to 1990s. The geographical names were acquired during the field survey for topographical mapping in a standard format known as 22 Topo. Geographical names in the vernacular script of the concerned area were collected by the field surveyor by way of listening to the pronunciation of the corresponding place name by the local community. The same was further vetted and authenticated by a local government official. The Camp officer of the field teams then converted the names from vernacular to Roman and Devanagari scripts. The final spellings of the geographical names in all the scripts was checked and confirmed by the Officer in charge of the field unit. The final names were then published as part of the topographical maps of the SoI.

The database, over time, got enriched with additional names from subsequent field activities, as additions or updates of geographical names. Currently, SoI holds an extensive repository of over 1.4 million ground-verified geographical names encompassing the entire expanse of the country.



22 Topo
old 20 Trav

SURVEY OF INDIA
No. ..11.. PARTY (SEC) Sheet No. 73 L/14/NW
Plane-table Section No. 2 District Kendrāpāra Sub-division Kendrāpāra
LIST OF NAMES. Surveyed by B. K. Tiwari Season 1996-97.

Plane table Refe-rence Number	Refe-rence to Square	Vernacular written by the Surveyor	Vernacular as written by a local official	Names Spelling by the Camp officer in English	Names Spelling by the Camp officer in Devnagri	Corrected spelling in English	Description	Number of Houses	Tahsil Taluk &c.	Remarks
1	2	3	4	5	5(a)	6	7	8	9	10
1	1A 1B	ब्राह्मणी नदी	ବ୍ରାହ୍ମଣୀ ନଦୀ	BRAHMANI R.	ब्राह्मणी नदी	BRĀHMANI R.	River	—		
2	3B	पातसाला नदी	ପାଟଶାଳା	PATSALA R	पातसाला नदी	PATSALA N.	Nadi	—		
3	2B	दारु ओरा	ଦାରୁଓରା	DARU ORA	दारुओरा	DĀRU ORĀ	Village	365	Rajnagar	R
4	2B	बलरामपुर	ବଳରାମପୁର	BALARĀMPUR	बलराम पुर	BALARĀMPUR	"	150		R
5	2B	खामारसाही	ଖମାରସାହି	KHĀMARSAHI	खमारसाही	KHAMARSĀHI	"	30		
6	2A	घड़िआमाल	ଘଡ଼ିଆମାଳ	GHARIĀMĀL	घड़िआमाल	GHARIĀMĀLI	"	425		R, PO, Hig

Note :- In column 6, names, the spelling of which has been taken from the imperial Gazetteer, should be underlined. In column 7, it should be stated whether a name refers to a town, village, temple, hill, river, lake, &c. Column 8 only refers to towns, villages & c., this information is required to determine their relative importance. In column 10, it should be stated whether, in a town, village, &c., there is a market, hospital, P.O., ...

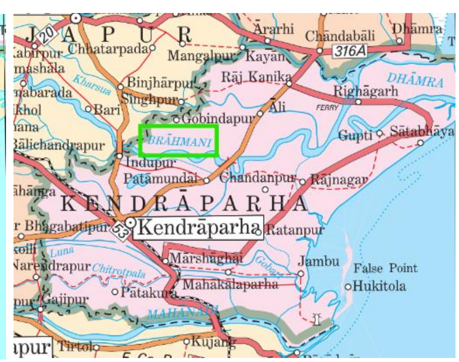
Tahasildar Rajnagar.

Sig. of Surveyor ... Date 17-1-97 Sig. of Local Officer ... Date
Sig. of Camp Officer ... Date 15-3-97 Sig. of O.C. Party ... Date

22-Topo filled proforma

Odia (Odia Script)

Hindi (Devanagiri Script)



English (Roman Script)

Fig-3: 22-Topo filled proforma and corresponding geographical name (in green box) published State Map of Odisha in three languages (Odia, Hindi and English)

SoI has further enriched this database by software-enabled conversion of all the 1.4 million geographical names into 22 scheduled languages of India and the same constitutes a national geographical name database (GNDB). The transliteration tool uses ASCII & Unicode storage standards & is custom built for all 22 scheduled languages. This geographical name database has been published by Survey of India on the Online Maps Portal (https://onlinemaps.surveyofindia.gov.in/) for the benefit of all users. The geographical names of States that use the Devanagari script and Hindi as the primary spoken language have been converted into all other scheduled languages using Devanagari as the base script. Likewise, for the remaining States, the database has been compiled based on field-collected records (22 Topo) and converted using the respective regional language script as the base script through the tool.

Future Roadmap

The challenges faced in the traditional processes include the non-comprehension of the local language/script by the field surveyor which may lead to human error in the collection and transliteration process, no earlier record or reference to the pronunciation of the particular geographical name, no consultations with linguistic experts involved for the transliteration and verification of the spelling. Additionally, automation errors have also contributed to challenges.

To address the above challenges, a holistic system for collection and transliteration of geographical names based on the sound bites / audio bytes is being developed. This will also aid capturing the actual pronunciation of the geographical name. This process is facilitated through mobile applications equipped with linguistic expert interface and integrated into a centralized server-based system.

The system involves leveraging the capabilities of the pan-India Common Services Centers (CSCs) scheme implemented in the country under the Digital India initiative of the Ministry of Electronics and Information Technology, Government of India. CSCs are the access points for the delivery of essential Government services, social welfare schemes, healthcare, financial, education, agriculture, and a host of business to citizen services in rural and remote areas of the country. The CSCs are managed by Village Level Entrepreneurs (VLE) who are well integrated into the community. A collaborative effort with the VLEs for the collection of the sound bites / audio bytes, will aid in identifying local speakers with native pronunciations thus ensuring capturing authentic pronunciation of geographical names.

In the proposed system geospatial coordinates of the location also would be captured along with the sound bites/ audio bytes thus integrating toponymy data into the national Geographic Information System datasets being maintained by SoI. Subsequently, a National Geographical Name Information System (NGNIS) will be developed by SoI for access by through a Geo-Portal by all stakeholders. The NGNIS shall contain the Geographical Name Data Base (GNDB) associated with spatial data. This GNDB shall contain;

i. Existing geographical Name in Vernacular script (local language of the State), Devanagari script (Hindi), Roman script (English).
ii. Variant/Alias Name
iii. Notified Name
iv. Sound bite/ Audio-byte file(s) linked with the Geographical Name
v. Description of Geographical name e.g. Village, Town, River, Mountain, etc.
vi. Latitude, longitude
vii. Reference System - Datum
viii. Reference to the national mapping organizations topographic Sheet Numbers.

Points for discussion

The Group of Experts is invited to:

1. Take note of the procedures of the collection and standardization of Geographical Names in India.
2. Provide feedback and insights on similar efforts to enhance collaboration among multiple stakeholders