PRESS INFORMATION BUREAU
(Research Unit)
Ministry of Information and Broadcasting
Government of India

BHASHINI - National Language Translation Mission

BHASHINI- BHASa INterface for India

Digital India BHASHINI - India's AI-led language translation platform

(Ministry of Electronics & IT)

August 26, 2022

"In this digital age, we have an opportunity to transform lives of people"

- **Prime Minister Narendra Modi**

Introduction:

'Digital India BHASHINI' seeks to enable easy access to the internet and digital services in Indian languages, including voice-based access, and help the creation of content in Indian languages. It was launched by Prime Minister Narendra Modi at the inauguration of the Digital India Week [1] 2022 at Gandhinagar, Gujarat.

Digital India BHASHINI, as India's Artificial Intelligence (AI)-led language translation platform, will enable massive citizen engagement to build multilingual datasets through a crowd-sourcing initiative called Bhasha Daan.

Aim of Digital India BHASHINI:

- This mission aims to **empower Indian citizens** by connecting them to the Digital Initiatives [2] of the country in their own language, thereby leading to **digital inclusion**.

- The Bhashini platform [3] will catalyze the entire Digital Ecosystem and is a giant step towards realizing the goal of **Digital Government**.

---

[1] https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1839175
[2] https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1827997
[3] https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1827997

- BHASHINI[4] aims to build a **National Public Digital Platform for languages** to develop services and products for citizens by leveraging the power of artificial intelligence and other emerging technologies.
- BHASHINI also aims to increase the content in Indian languages on the Internet substantially[5] in the domains of public interest, particularly, governance-and-policy, science & technology, etc., thus **encouraging citizens to use the Internet in their own language**.

Salient Features:

- BHASHINI[6] shall act as an orchestrator to unify and align a large diverse network across government, industry, academia, research groups and start-ups to bring all their contributions into an open repository.
- BHASHINI will involve the development of a national digital public platform for language[7] to **provide universal access to content** i.e. boost the delivery of digital content in all Indian languages. This would result in the creation of a knowledge-based society where information is freely and readily available and would make the ecosystem and citizens *"Atmanirbhar."*
- Citizens shall be provided easy tools [8]and motivated to contribute towards the languages of their liking through BHASHINI's crowd-sourcing platform. They will be the prime beneficiaries of this whole initiative.
- Startups[9] will also be encouraged to create innovative applications by using Bhashini's resources.

---

[4] https://bhashini.gov.in/en/about
[5] https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1827997
[6] https://bhashini.gov.in/en/about
[7] https://bhashini.gov.in/images/Bhashini_-_Whitepaper.pdf
[8] https://bhashini.gov.in/images/Bhashini_-_Whitepaper.pdf
[9] https://bhashini.gov.in/images/Bhashini_-_Whitepaper.pdf

Road Map for BHASHINI

Bhashini's journey has been designed to consist of key milestones segregated across four tracks, namely Foundation Track, Contribution Track, Innovation Track and Grand Challenge Track.

Foundation Track

- Publish Universal Language Contribution API
- Data repository
- Model repository
- Benchmarking system
- Data collection tools

Contribution Track

- Training and benchmark datasets
- Data contributions from government entities, language chapters, communities etc
- Crowd sourcing initiatives
- Open source language models

Innovation Track

- Hackathons and challenge rounds for developing applications
- Inter-ministerial projects that leverage Bhashini to provide citizen centric services

Grand Challenge Track

- One grand challenge related to Bhashini's goals every year Participation from academia and industry

BHASHA DAAN

Bhasha Daan is an initiative to crowdsource language inputs for multiple Indian languages as part of Project BHASHINI. It calls upon citizens to help build an open repository of data to digitally enrich his/her own language. The aim here is to create large datasets for Indian languages, which can be used to train AI models for use by different stakeholders to create products or services for the

betterment of society. One can contribute to the project anonymously. The bigger aim is to create an open-source repository that anyone can access.
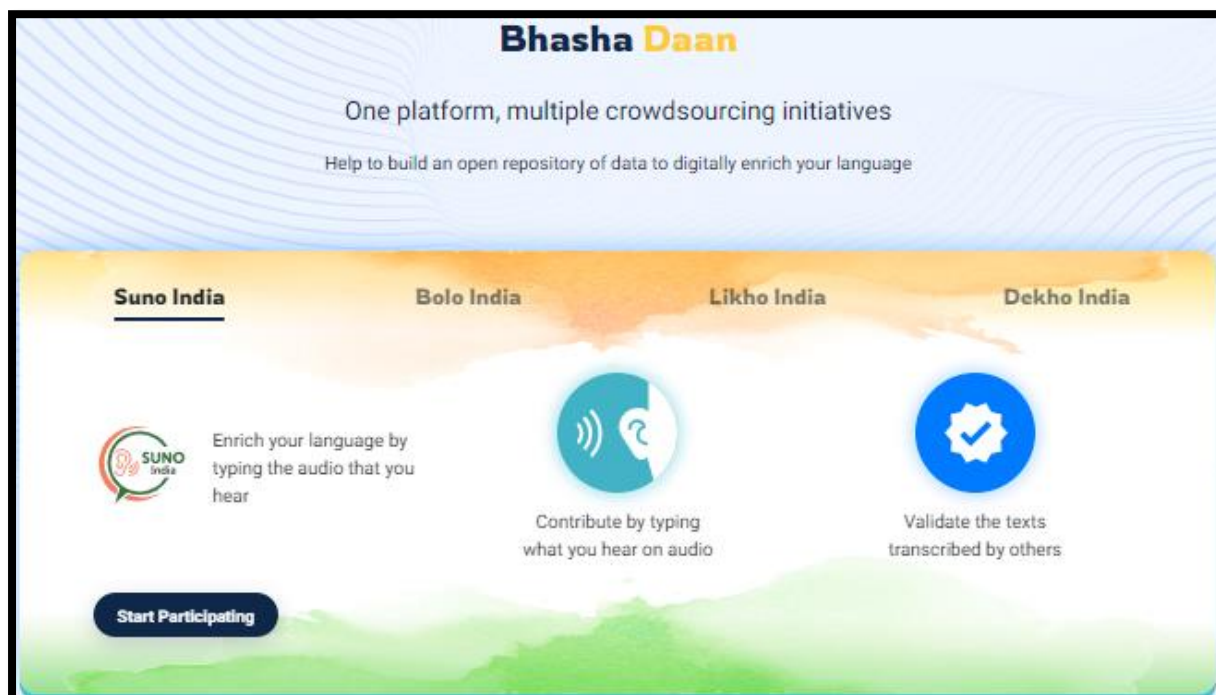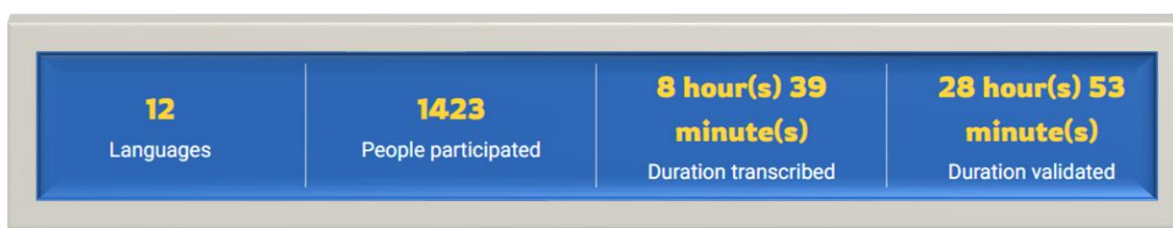

Categories of Bhasha Daan

- ➤ **Suno India**
- One can enrich one's language by typing the audio he/she hears, or by validating the text transcribed by others.
- [31.3 % of the Suno India Target has already been achieved](#)[10]

Contributions made to Suno India

12 Languages	1423 People participated	8 hour(s) 39 minute(s) Duration transcribed	28 hour(s) 53 minute(s) Duration validated

---

[10] https://bhashini.gov.in/bhashadaan/en/suno-india

- ➤ **Bolo India**
- One can enrich one's language by donating the voice, through recording of sentences
- One can also validate the audios recorded by others
- [17.1% of Bolo India Target Achieved](11)[11]

Contributions made to Bolo India

12	750	14 hour(s) 39 minute(s)	5 hour(s) 50 minute(s)
Languages	People participated	Duration recorded	Duration validated
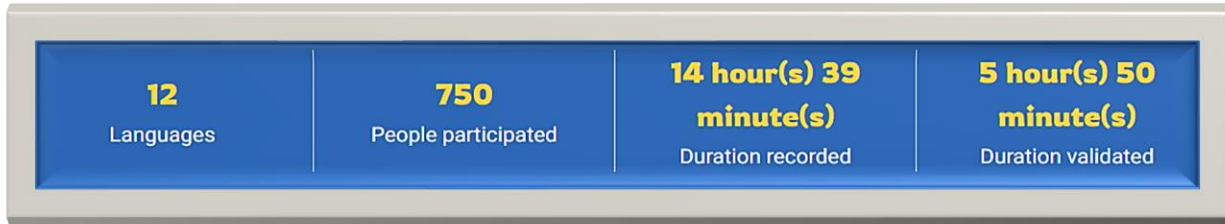
- ➤ **Likho India**

- One can contribute by translating the prompted text.
- One can also validate the translations contributed by others.

Contributions made to Likho India

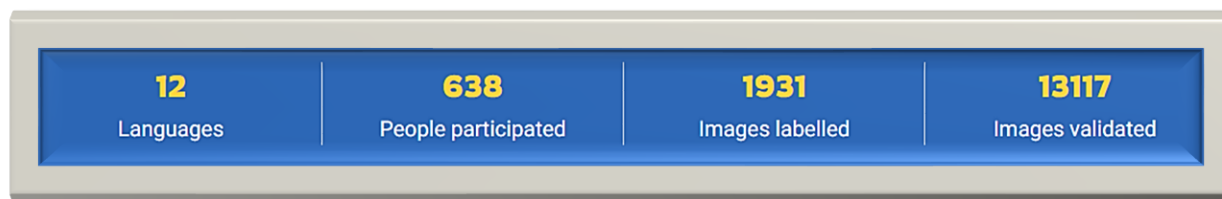| 53	583	1990	4169
Language pairs	People participated	Translations done	Translations validated
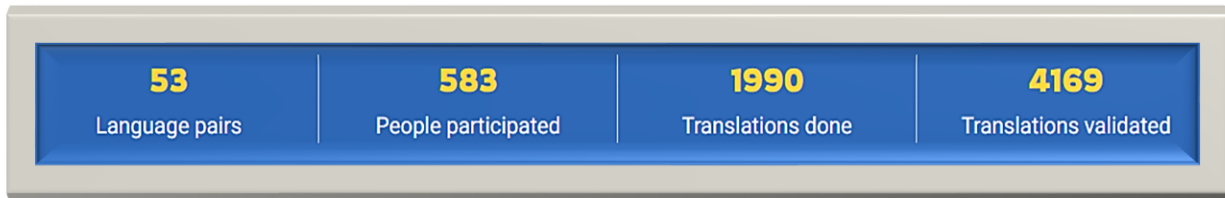
- ➤ **Dekho India**

- One can enrich one's language by typing the text seen, or by labelling the image.
- One can also validate the images contributed by others

---

[11] https://bhashini.gov.in/bhashadaan/en/bolo-india

Contributions made to Dekho India

12	638	1931	13117
Languages	People participated	Images labelled	Images validated

Conclusion[12]

Despite having centuries of literary tradition and lakhs of speakers, India's official languages have all been grappling with one common problem – adapting to the demands and the potential of the internet. BHASHINI has developed a roadmap for enabling an ecosystem to make digital content more accessible to Indian citizens by offering it to them in their native language. Bhashini has introduced certain novel ways in which the ecosystem players can contribute towards the success of this objective. At the same time, Bhashini shall also introduce standards for data collection, curation, benchmarking and adoption. This shall ensure widespread accessibility to data and in turn, should result in the development of innovative applications and services for the citizens.

References:

- https://bhashini.gov.in/en/about
- https://bhashini.gov.in/images/Bhashini_-_Whitepaper.pdf
- https://pib.gov.in/PressReleaseDetail.aspx?PRID=17860
- https://pib.gov.in/PressReleaseIframePage.aspx18279975
- https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1839175
- https://pib.gov.in/PressReleaseDetail.aspx?PRID=1786560
- https://pib.gov.in/PressReleaseDetailm.aspx?PRID=1693887

Twitter Links

- https://twitter.com/NWRailways/status/1543912513070563330?s=20&t=QI812cMPbA4s1pvXGvDdcw
- https://twitter.com/GoI_MeitY/status/1268436782766452736?s=20&t=QI812cMPbA4s1pvXGvDdcw
- https://twitter.com/AkashvaniAIR/status/1356131239955099649?s=20&t=QI812cMPbA4s1pvXGvDdcw
- https://twitter.com/investindia/status/1366749702138527746?s=20&t=QI812cMPbA4s1pvXGvDdcw
- https://twitter.com/_DigitalIndia/status/1531938608101269505?s=20&t=QI812cMPbA4s1pvXGvDdcw
- https://twitter.com/PBNS_India/status/1544649008731529216?s=20&t=QI812cMPbA4s1pvXGvDdcw

Video Links

- https://www.youtube.com/watch?v=ilSJR_vTHdc

AG/HP/RC/SA/MK

---

[12] https://bhashini.gov.in/images/Bhashini_-_Whitepaper.pdf