



SUCCESS STORY

Developing India's First National AI Data Repository: AlKosh

Customer

National e-Governance Division

Country

India

Industry

Government Technology



ABOUT THE CLIENT



IndiaAI, a department under the Ministry of IT, has created a seven-pillar strategy to strengthen the country's artificial intelligence initiatives. To implement this strategy, IndiaAI works with the National e-Governance Division (NeGD), which serves as the government's execution arm for major digital projects. One of these pillars is AlKosh - India's first national AI datasets platform. AlKosh is designed to bring together datasets, models, and resources from government ministries, private organizations, and research institutions, giving users a trusted source of data to develop and train AI solutions.

BUSINESS SITUATION

Before AlKosh, datasets in India were siloed across ministries, universities, hospitals, and private organizations. Data engineers had to spend significant time sourcing, cleaning, and validating information before they could begin training AI models. Accessing private datasets was also expensive, and without consistent validation or licensing frameworks, the quality and reliability of data varied greatly. In addition, training required GPUs, which were costly to outsource and not always easily available.

To address these challenges, IndiaAI envisioned a single national repository that would consolidate datasets from government and private contributors into one trusted platform. The goal was to make data affordable, authentic, and compliant with privacy

SERVICES DELIVERED

AI Development



Software Engineering



TECH STACK



regulations, while also providing built-in tools for experimentation and model training. The platform also needed to be designed for long-term growth, ensuring that more ministries, organizations, and contributors could be onboarded over time.

To implement this vision, NeGD partnered with Daffodil Software to develop AIKosh. Daffodil Software provided the technical expertise required to build a platform capable of AI development and data management at a national level. To achieve this vision, NeGD required a solution that would:

KEY REQUIREMENTS

01

Unify fragmented sources

Aggregate datasets, models, and resources from ministries, universities, hospitals, and private contributors into one searchable repository.

02

Enforce validation and approvals

Introduce a multi-level review process so only verified and trustworthy datasets were published.

03

Define usage rights clearly

Attach licenses to each dataset, making rules for access, sharing, and application transparent.

04

Control data access

Allow contributors to decide how their datasets are shared - whether openly downloadable, available only on request, or kept private for use within their own workspace.

05

Maintain data quality

Introduce a scoring system based on uniqueness, completeness, and regular updates, encouraging contributors to keep their datasets accurate and valuable.

06

Provide a sandbox for experimentation

Offer an in-platform GPU-enabled environment where data engineers could test datasets and train AI models directly.

07

Scale with adoption

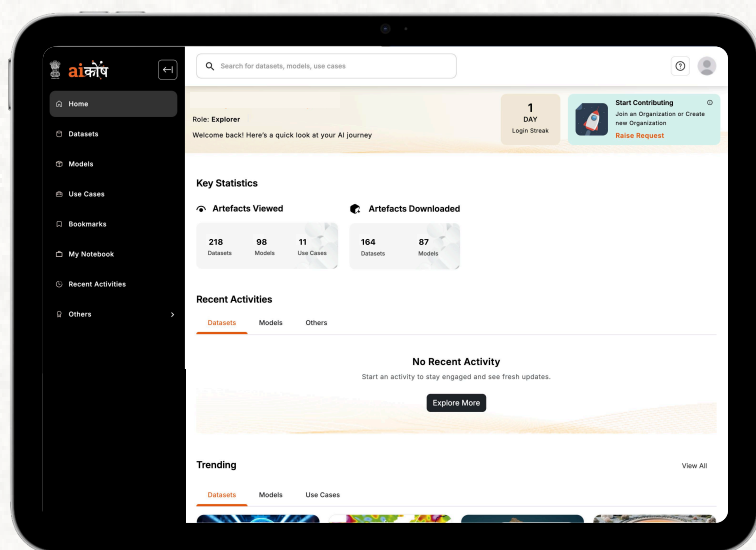
Ensure the architecture could seamlessly onboard more ministries, organizations, and private contributors over time.

THE SOLUTION

The AIKosh platform was built with several key components working together to create a comprehensive AI data collection platform. Each feature was designed to address specific challenges in data accessibility and AI development.

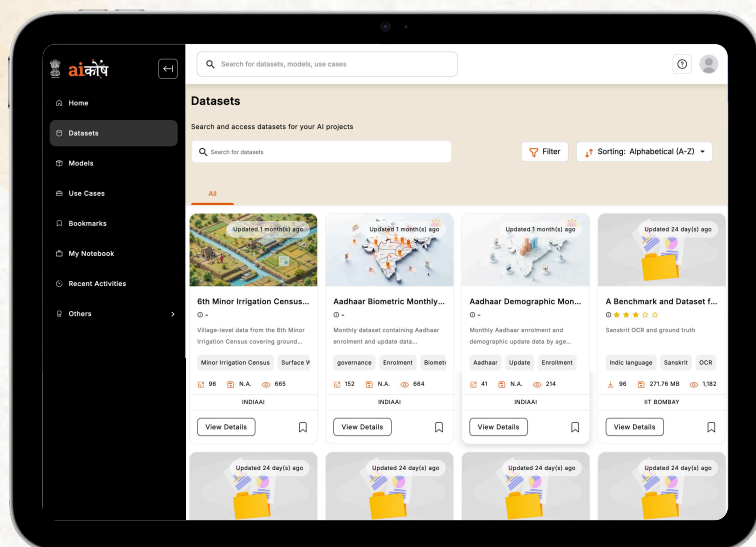
Unified Repository Of Artefacts

AIKosh serves as a single destination for datasets, AI models, toolkits, and use cases, eliminating the silos that previously existed across ministries and organizations. By offering a central hub, it allows researchers, startups, and enterprises to discover and leverage resources from multiple domains without duplication of effort. The repository is searchable by sector, organization, and usage tags, ensuring ease of discovery and efficient knowledge sharing.



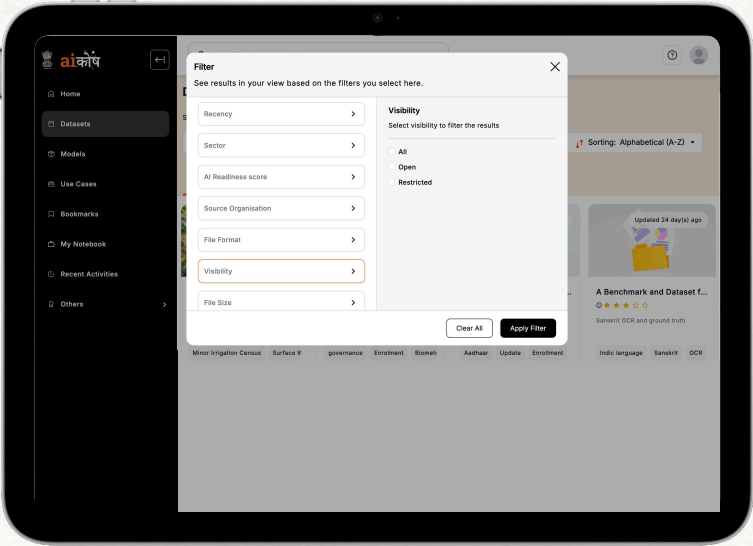
Role-Based Governance And Access Control

The platform introduced structured roles - Explorer, Contributor, Organization Admin, and Platform Admin—each with distinct privileges. Contributors could upload datasets, models, and use cases, while organization admins oversaw approvals and access requests. This model ensured that data sharing remained secure and accountable, while still giving contributors flexibility to decide whether their artefacts were open, request-based, or private.



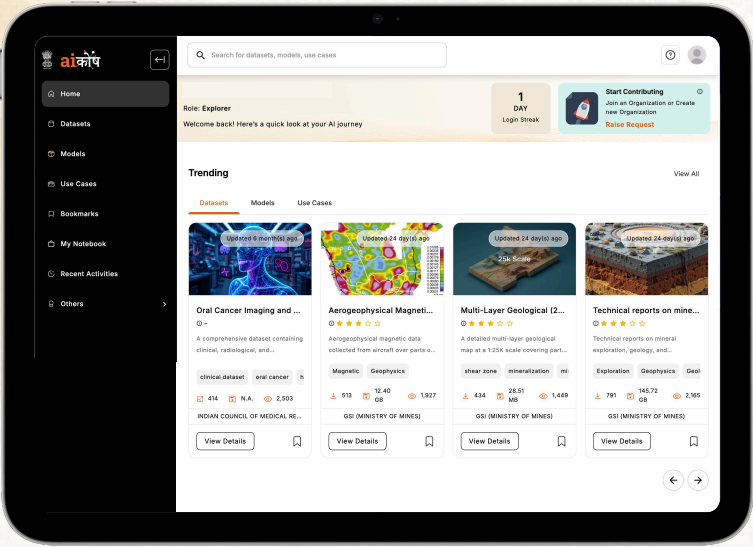
Multi-Level Approval Workflow

Every dataset or model published on AIKosh passed through a rigorous multi-step review process. Uploads were first validated by organization admins, followed by platform moderators, and finally the Ministry, before being visible on the platform. This ensured authenticity, compliance, and consistency across all artefacts, building trust for users who rely on verified resources.



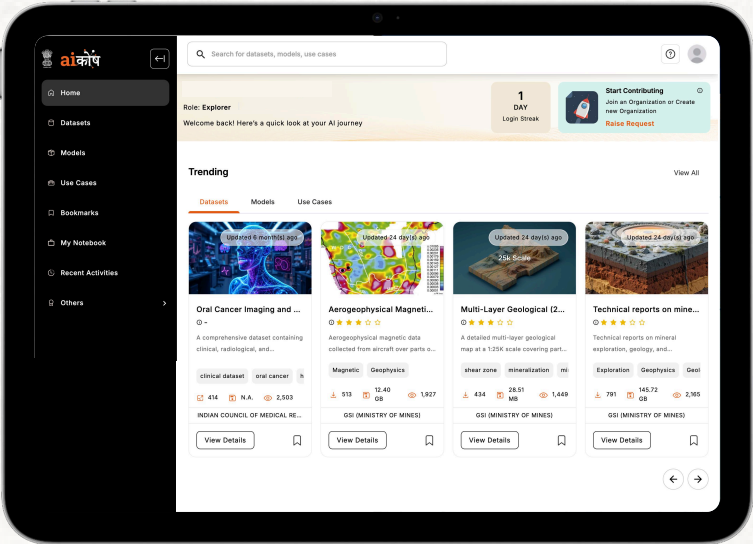
Flexible Data Ingestion

AIKosh supported multiple methods of onboarding datasets, manual uploads, push and pull APIs, secure SFTP transfers, and direct integrations with popular platforms like GitHub, Kaggle, and Hugging Face. This flexibility reduced barriers for contributors, enabling both government ministries and private organizations to seamlessly bring their data into the ecosystem. API keys further allowed programmatic access, enabling developers to integrate datasets directly into their workflows.



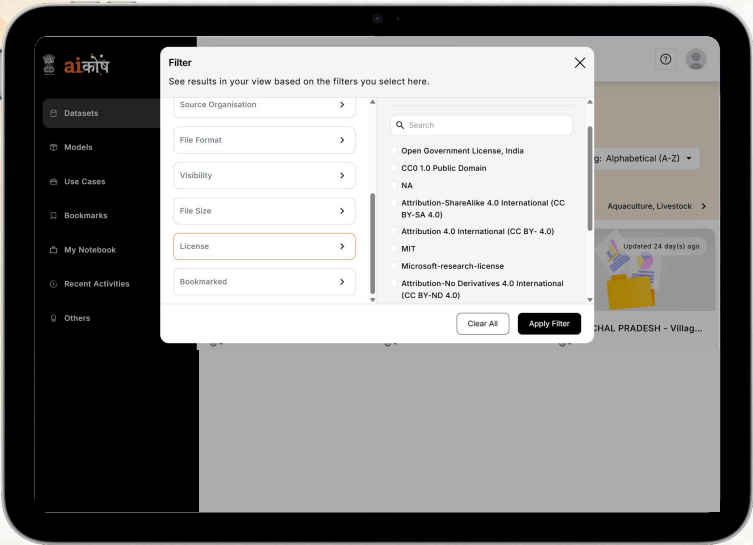
Data Quality And Validation

To encourage reliability, Daffodil implemented automated pipelines that scored datasets on factors such as uniqueness, completeness, frequency of updates, and absence of duplication. Contributors received a star-rating, motivating them to improve data quality over time.



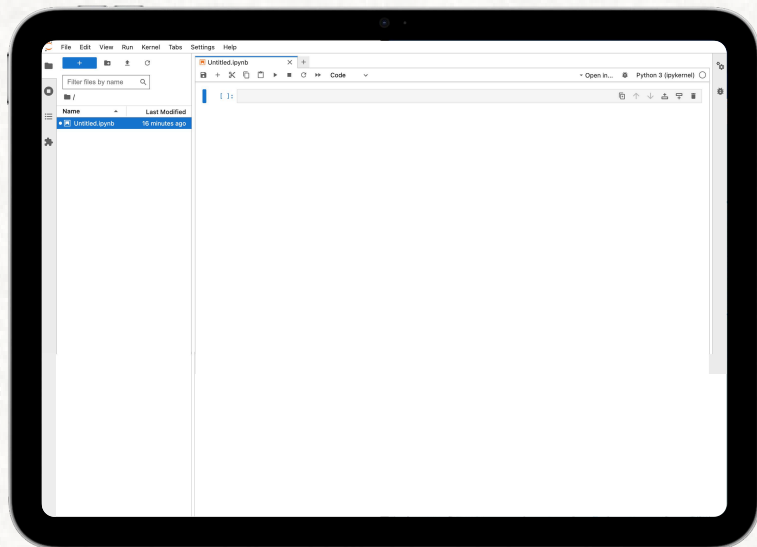
Licensing And Compliance

Each dataset was published under a clearly defined license, ranging from open-source to restricted or government-specific terms. This provided transparency on how data could be accessed, shared, or reused. Anonymization tools were also included to safeguard sensitive information, ensuring compliance with national data protection policies and ethical AI practices.



Integrated Notebook Environment

One of AIKosh's key impactful features was its GPU-enabled notebook, which gave users a ready-to-use sandbox for experimentation. Instead of outsourcing expensive GPUs, researchers and engineers could directly train and test AI models on the platform. This lowered costs and accelerated the development cycle, making AI research more accessible to a wider community.



IMPACT

AIKosh has consolidated 1,447 datasets and 217 AI models from 34 organizations across 20+ sectors into a single platform, replacing the fragmented system where data engineers previously searched multiple disconnected government and private sources. The integrated notebook environment eliminated the need for teams to outsource GPU computing for AI model training and testing.

The platform established standardized access controls and automated quality scoring across government ministries and private contributors, enabling cross-sector data sharing that was previously limited by organizational silos. This created the foundation for India's national AI development infrastructure, supporting broader AI advancement initiatives across the country.

1,447

Datasets
Available

217

AI
Models

20

Sectors
Covered

34

Contributing
Organizations

HAVE A SOFTWARE PRODUCT VISION IN MIND?

Set up a personalized consultation with our technology expert

Let's Talk 



info@daffodilsw.com



www.daffodilsw.com