Open DataStack to power Government of Uttar Pradesh's AgriStack

An approach note on Agri DataStack

from

International Finance Corporation (IFC) / The World Bank Group

To

The Government of Uttar Pradesh.

Open DataStack to power Government of Uttar Pradesh's AgriStack

Background:

Digitalization in Agriculture (including Analytics, AI, Robotics) is transforming existing business processes and enabling new means to deploy innovative services and products at an unimaginable scale and pace.

Digital agriculture has a much broader scope with the potential use in the areas of supply chains, access to finance and markets, contextual advisory, building farmer networks etc. The term fourth agricultural revolution is also used interchangeably with digital agriculture - the use of digital and precision tools across agriculture – where hyper spectral satellites, IOT, robotics, AI etc are being used in agriculture and allied sectors and are giving rise to innovations across all stages of the food, inputs and water value chain.

Project Goal

The goals for an Agri stack for Uttar Pradesh should include:

1. **Increase Agricultural Productivity:** By providing farmers with actionable insight through access to real-time data on weather conditions, soil health, and crop yields, and help make better decisions about planting, irrigation, and fertilization.

2. **Improve livelihoods:** Improve farmers' livelihoods by connecting them with markets and buyers, and help them sell their produce at better prices.

3. **Enhance Resilience:** Improve financial resilience through insurance products as well as to climate change by identifying areas that are most vulnerable and using Data and information to protect their crops.

4. **Digitization of Agriculture and farmers:** The project should aim to digitize agriculture and farmers by creating a digital platform that allows farmers to access important information and connect with buyers and the government.

5. **Improve the Agri economy and hence overall economic condition of Farmers**: Increase agricultural productivity, profitability, and resilience for farmers and have a significant impact on the overall economy.

6. **Enhancing Farmer Trust and Digital literacy**: Improve the acceptability of digital solutions and technologies by the farmers through improved accreditation, validation of Ag Tech claims and improved digital literacy.

7. **Balancing Farmers interests for Data confidentiality and economic benefit:** Ensure that the Farmer data is not misused and the Farmers / State / National interests are protected. At the same time, while protecting confidentiality, ensure that Agri Data is leveraged for its economic potential and a regulated but commercially viable Agri Data market place is created across the Govt, Private sector and the Farmer.

Alignment of AgriStack and DataStack:

There has been a lot of work done by the Central and State Governments, Academic, Research Institutions, and the Private Sector in conceptualising different approaches and aspects of an "AgriStack" to transform Indian Agriculture. An AgriStack is envisioned as an "enabling framework of Services (applications) and Agri Data around a data exchange", built to support several Agri use cases, while encouraging application and data interoperability using well defined interfaces.

A well designed AgriStack will help reduce duplication by integrating data sources and a vast backend of new and existing applications: Govt's eNam, ITC's eChoupal, NCDEX's NeML, APEDA's TraceNet, etc related to logistics, weather, supply-chain, warehousing, assaying, recommendation engines, etc.

There is a need to integrate into all this a focussed discussion and engagement on an **Open Agri Data Stack** which forms the backbone of the AgriStack aligned to the several critical use cases chosen by the State Government of Uttar Pradesh.

At the core of Digital Agriculture is an increasing amount of live and real-time "Data" which is being rapidly generated by both - the Government through its farmer friendly programs (including DBT and other schemes) and also the Private Sector as a by-product of business applications or through remote sensing (satellite imagery). Physical maps being digitized is an example, while another is through the multiplicity of payment and Agri trading platforms. In addition, there is a rapid increase in precision farming applications generating "machine-data" from:

- **Drones / UAV** *(e.g. real-time monitoring of farms enabling mapping, surveying, 3D modelling)*
- **Satellites and remote sensing** *(providing direct farmer advisory, for yield estimation, early pest warning etc.)*

- **Robotics** *(e.g. nutrient injection, harvesting, targeting high precision weed removal)*
- **Automated Material Handling** *(e.g. harvesting equipment, grain/food processors)*
- **Farm sensors and management** *(e.g. for soil moisture, water and nutrient injection)*
- **Mobile cameras** *(e.g. for identifying pest, nutrient deficiencies)*

This mountain of data, once anonymised, aggregated and processed can be re-purposed using AI for different use cases - raising yields, optimising national resources and doubling farmer income.

Approach to a Data Stack for Uttar Pradesh:

Given this context, this approach paper, seeks to lay down some practical suggestions to start with – using which we can to contribute to the Farmer's benefit as well as the broader debate on an AgriStack architecture and the specific data building blocks needed.

Through this process we will need to understand and uncover both the explicit and the implicit expectations of the Government of Uttar Pradesh and several interested entities (Public, Private, Research, and Academia) from the AgriStack and translate these into defining and designing a data stack reference design that becomes the core building block of the proposed AgriStack. Ultimately, this will need work across multi-stakeholder interest groups to enable deeper clarity in understanding and defining "Open Agri Data" for the AgriStack and help converge ideas and actions.

Use Case led approach:

Very often attempts to work on complex, multi-year Agri problems, however well-intentioned (and regardless of how these attempts begin) end up adopting a "technology first", feature focussed approach, sometimes driven by a specific company's solution.

Instead we suggest, adopting a farmer-centric use case approach, with the use cases prioritised carefully by the Government for maximum societal impact. This allows the Government to coalesce several related Agri technologies and overlapping Ag Tech Start-ups together for working on the use case.

This shift in focus, though subtle, requires tough implementation focus to hold course, but produces long term sustainable results.

Elements to consider for aligning an Agri Data Stack to an Use Case led approach:

We recommend, that to inject more specificity, engagement be held as suggested above, in line with key use cases identified by the State Government, which they wish to be enabled by AgriStack and the consequent Agri Data Stack. We need to identify several objectives for each use case, focussing on specific data elements needed for each:

a) Build a 360° view of the "problem statement" for the Use Case with a "To-Be" state for the short / medium / long term.
b) Identify the strategic levers needed to solve the use case.
c) Identify Data elements and then the consequent Data sets needed for each strategic lever.
d) Identify the sources of Data sets / elements with a view of authenticating the data.

Discussion around each use case will focus on answering the following types of questions (illustrative, not comprehensive):

1) What data (spatial and temporal resolution) is needed for the use case to be deployed for a farmer?
2) What data is currently available with central Govt. or the state Govt. or private enterprises?
3) What are the data gaps for this use case to be enabled at the desired level of scale and granularity?
4) How can these data gaps be filled? Does the Govt. need to revamp or adapt existing data collection methods or incorporate secondary or private data sets to fill these gaps?
5) What are the building blocks for each use case which can developed in open source as a Digital Public Good (DPG)?
6) How to build the Agri Data Stack while being cognizant of data governance standards – access, privacy, interoperability, quality and incentive for the data owners?
7) Who are the diverse stakeholders who could play a role in each of the thematic areas?
8) What are the policy measures required to make the data based digital tools deployment beneficial for the Indian farmers?

Integrating State Govt's Use cases with Public & Private data sources through an Agri Data Stack:

Some key areas of interest to the Government of Uttar Pradesh:

1) Credit and Rural finance for the smallholder farmers.
2) Actionable and context specific advisory for farmers, especially to prevent pest and disease.
3) Market Linkages and Price forecasting to reduce food wastage.

Process steps for UP State Govt to AgTechs and Data Stack for different types of use cases:

Adopting a use-case-led approach requires identifying and mapping data sets needed for Agri technologies provided by start-ups. Identification and validation of these use cases and the shortlisting of startups to implement solutions in turn needs thorough research and review of the market dynamics, the existing startup landscape, and their tech capabilities/solutions. Scoping of the Use case, identifying strategic levers associated to the use case, technology/capabilities maturity and the adoption of emerging technologies can be a few parameters in evaluating startups to participate in a Data Stack Project.

Based on these parameters, and feedback from other Stat es, we suggest three categories of use cases that require slightly different approaches in evaluating and selecting the right technology partner for the pilot and eventually the implementation. These categories are briefly described below:

1. **"Broad Scope" Use Cases**: These are problem Identification and selection of start-ups to work on Broad Scope Use Cases can be done by floating a request for proposal. The scope can be narrowed down to suit the requirements of the use case. Advanced validation is needed to identify the top 2 or 3 start-ups post the evaluation and shortlisting of proposals. An example of a broad-scope Use Case can be Farmer Credit, Market access, Actionable Advisory, Irrigation, etc.

2. **Narrow Scope Use Case:** Use Cases that are narrow in scope are those that typically could serve to be building blocks as a part of several other use cases. For instance, Electronic Farm Records, Crop acreage etc are relatively narrow in scope and could be used as building blocks for several "Broader" use cases. These would require a request for a limited "discover pilot" to assess the specific usability and technical capabilities of the start-ups. Working with a limited number of start-ups who could carry out a pilot to meet the requirements is a way forward.

3. **Emerging Tech-Based Use Cases**: Solutions on emerging technologies require considering various use cases to apply the technology. This can help review and narrow down the capabilities and also allows the start-ups in showcasing and demonstrating their technologies. An example can be the use of Robotics for various use cases like the plucking of cotton, precision seeding, removing weeds, etc

The process steps for all three categories of use cases in the implementation of a Data Stack are shown below (refer to Diagram 1).
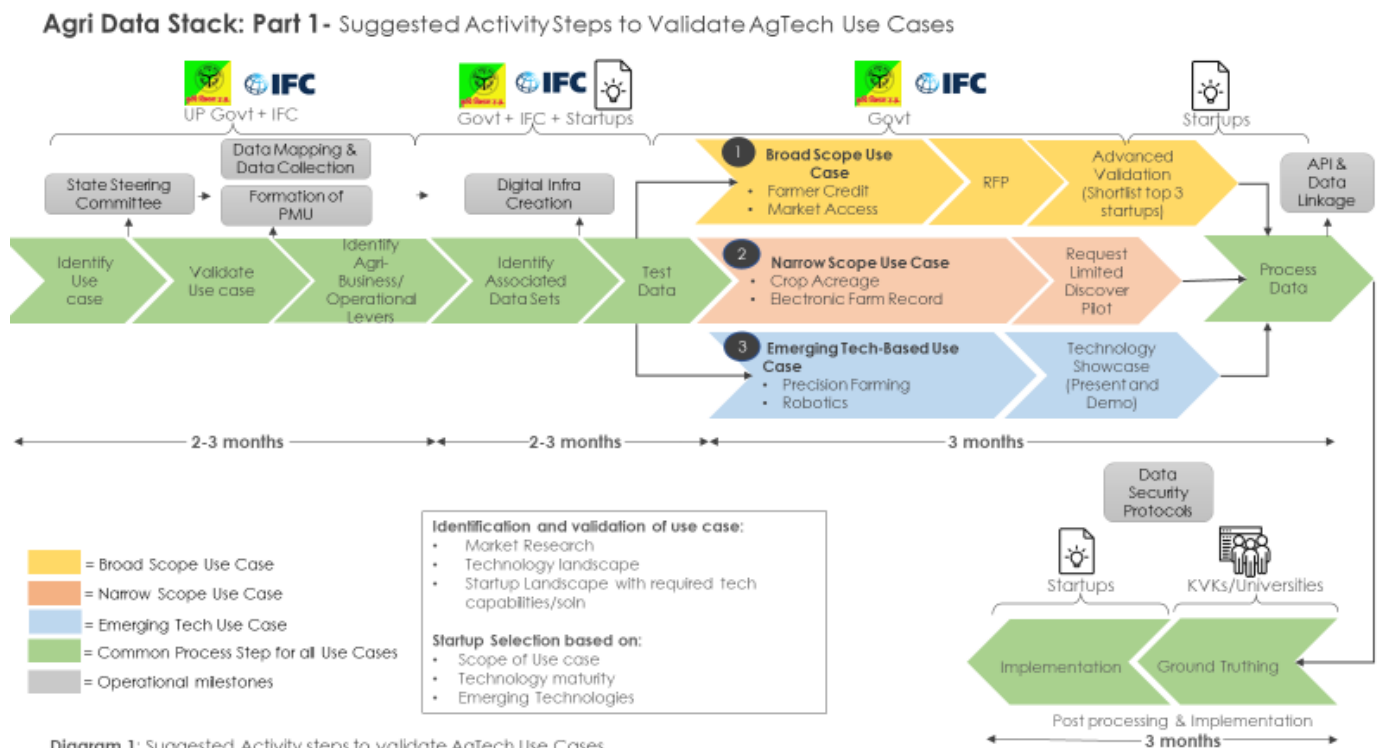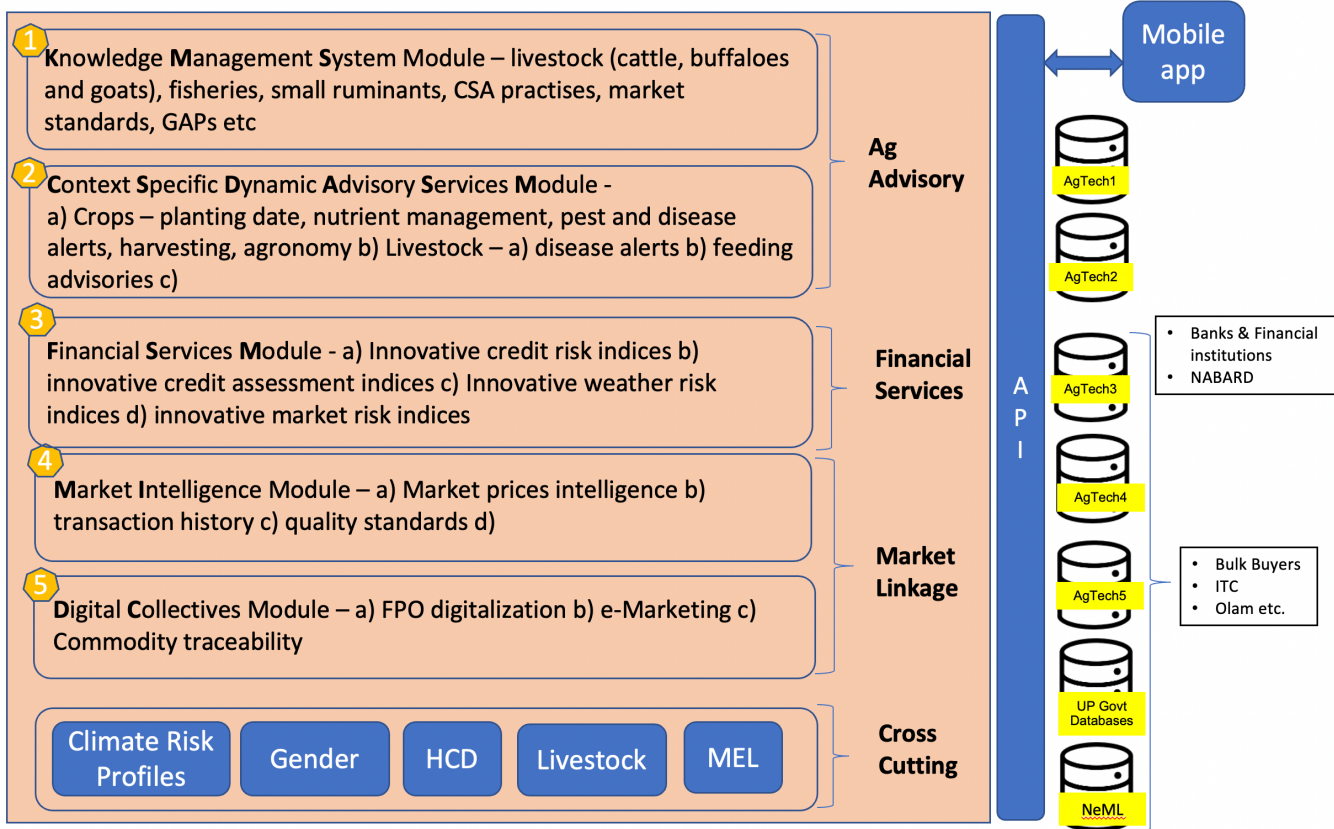

Agri Data Stack: Part 1- Suggested Activity Steps to Validate AgTech Use Cases

Diagram 1: Suggested Activity steps to validate AgTech Use Cases

1 Knowledge Management System Module – livestock (cattle, buffaloes and goats), fisheries, small ruminants, CSA practises, market standards, GAPs etc

2 Context Specific Dynamic Advisory Services Module -
a) Crops – planting date, nutrient management, pest and disease alerts, harvesting, agronomy b) Livestock – a) disease alerts b) feeding advisories c)

3 Financial Services Module - a) Innovative credit risk indices b) innovative credit assessment indices c) Innovative weather risk indices d) innovative market risk indices

4 Market Intelligence Module – a) Market prices intelligence b) transaction history c) quality standards d)

5 Digital Collectives Module – a) FPO digitalization b) e-Marketing c) Commodity traceability

Climate Risk Profiles | Gender | HCD | Livestock | MEL

Ag Advisory

Financial Services

Market Linkage

Cross Cutting

Illustrative Approach Courtesy: ILRI / CGIAR

Mobile app

API

AgTech1
AgTech2
AgTech3
AgTech4
AgTech5
UP Govt Databases
NeML

- Banks & Financial institutions
- NABARD

- Bulk Buyers
- ITC
- Olam etc.

Annexure1: Strategic considerations for an Agri Data Stack:

Digitization of Agriculture and Agri Data often raise questions about Data Interoperability (standardisation, calibration, certification), policies, privacy, regulation, data sharing, scope, sources (including govt and private), farmer trust, data confidentiality, and data governance. Even though these are initial days for the State Govt of UP's work on Agri Data Stack, below are some important considerations that merit discussion in the context of a Data Stack:

- **Agricultural Data Interoperability:**

Given the overall potential for the Agri sector, it is important for us to ensure an enabling, transparent and well governed framework for data, information, and domain knowledge sharing to allow all players to succeed and grow. Agriculture's economic viability will improve via democratized and streamlined access to finance, advisory, supply chain, and markets.

The Agriculture sector is plagued by dozens of siloed software applications and islands of data, providing little benefit. Disaggregated and non-standardized data is deemed un-trustworthy and rendered partially ineffective for further processing. Standardization will help improve "data-trust" furthering automation using AI models, also avoiding real-world biases creeping into AI prediction. Agri data interoperability across the Government and the private sector will incubate new business models (Agri and non-Agri) providing transformative impact similar to how Aadhaar with UPI transformed digital payments..

- **Involve the AgTech Ecosystem into creating an Agri Data Stack:**

Like in any digital innovation, digital agriculture also flourishes within an ecosystem. The existence of a spectrum of data is the key to spurring innovations in digital agriculture. The concepts of data hubs, data stacks, and data exchanges have emerged to enable a frictionless way of aggregation (data from different sources), organization, management, distribution, sharing, and sub-letting of data while enforcing data governance.

- **Data governance will be crucial to engender Farmer and AgTech Trust**

A key challenge lies in finding a balance between leveraging the Data potential for the sector's growth and farmers' economic interests by fostering greater trust through carefully drafted State policies for monetisation, access, usage etc- while protecting the Farmer privacy and confidentiality of agricultural data.

State level Data usage policies should be framed for several tricky issues like personal vs public ownership, shared monetization, portability, oversight and protection, processing rights, inter-operability standards, etc., keeping these in context of newly introduced terms in the centrally drafted "Data Protection Bill" like Data "Controller", "Processor", "Principal", "Trustee or Custodian".

These issues have been grouped below for simpler understanding and discussion:

1) Data Usage (ownership, access, monetization etc.),
2) Data Trust through data Interoperability (standardization, calibration certification), and enhanced focus on digital literacy.
3) Data Portability (across digital Agri service providers to enable a competitive environment)

For long term sustenance of these efforts, while we inject digital technology and innovation into Agriculture to drive growth and enable Farmers and Farming to become profitable, resilient, and sustainable - we must adopt sound data governance. Much of this data pertains to small and marginal Farmers largely at the bottom of the financial and digital pyramid and hence we need to ensure they are not unknowingly exploited, financially bypassed by others using their farm data, for selfish business benefit and market manipulation.

- **Regulatory Agri Sandbox can streamline Regulation and Innovation needs of the Government of Uttar Pradesh.**

An Agri Sandbox suitably designed can be a catalyst to help synthesise the often conflicting regulatory and Innovation requirements between the State Govt along and the Ag Tech Start Ups. The **International Finance Corporation (IFC) / The World Bank Group,** aspires to partner with the Government of Uttar Pradesh**,** to build and run India's first Hybrid Regulatory Sandbox for Agriculture, created in a unique "Hub & Spoke" arrangement.

This will help scale-up AgTechs by removing inhibitors and advancing key policy, regulatory, innovation enablers that enhance AgTech competitiveness, scale and impact to the State's Agriculture within a four-year project period.

Within this project we will seek to:

a) increase the number of AgTechs in the State, across the entire agricultural value chain
b) increase revenue for AgTechs operating in the State,
c) increased number of AgTechs graduating from early stage to commercial scale.

- **Data Partnerships will accelerate innovation**

Efforts to harness this dataflow and facilitate "data partnerships" between Government of Uttar Pradesh and AgTech Start-ups, Corporates, Research, Academia based on either direct or indirect business benefit will accelerate the penetration of innovation.

This will create new business models and partnerships across the value-chain (insurance, market access, assaying etc.)that help monetise data further contributing to improved productivity and profitability of Agriculture and other sectors as well.

Annexure2: Illustrative Approach using Use Case: Credit and Rural finance for the smallholder farmers

Problem Statement

Simplify Agri loan assessment for Small/Medium farmers to facilitate loan products at minimal cost in the shortest period and help lenders mitigate their risk, monitor credit usage, and improve loan recovery.

Context
- A holistic approach towards Agri Finance and credit to extend purpose-neutral Loans based on digital assessment, backed by insurance.
- Consider cash flow lending acknowledging the collective economic need of rural households.
- Streamline process by leveraging tech to democratize farmer outreach and minimize touch points by embedding all regulatory & KCC requirements in the system.

Aspects to consider
- Creating a Digital Profile/Identity of the Farmer (*reliable, digital, private, anonymous)*
- An end-to-end approach - farmers avail loan/credit, secure crop insurance and exit by selling produce loan repayment.
- Use data from multiple sources – farmer, government, autonomous, digital, manual, to assess the farmer after consent.
- Credit assessment of the farmer to be linked to the farmer's household's credit history including loans from microfinance, NREGA, NBFCs, banks, savings, insurance, and cash flows, to increase visibility of savings and assessing financial discipline.
- Credit/Loan products to be created based on history of loan payments and savings.
- Cash flow lending; and direct payment to supplier (vs cash to farmer) and recover dues in bullet repayments (vs EMI)
- Motivate lenders to expand the loan book and reduce NPAs by improving usage visibility.
- Validate the Risk Scores with assessment on the ground.
- Break current barriers of physical proximity of a branch/office of the lender in managing the credit lifecycle.

Strategic Levers	Data Elements	Sources
Farmers		
Personal history	Aadhar, bank account, mobile no.	Loan form – verified by UIDAI, mobile company & bank database
Crop and other insurance	Insurance details	Loan form – verified by insurance company
Credit, payment discipline & income patterns	Bank details – I Tax details, electricity, mobile payments, labour income, microfinance, lending details, KCC, Fasal Bima, PM Kisan	Database of banks, HiMark, Cibil, Equifax, Itax, NRLM, KCC, Mobile company database, electricity department
Other sources of income	Dairy, fisheries, poultry, livestock, Kirana shop, tailoring, trading mechanic	Loan form – verified/populated from database of fisheries, animal husbandry department
Standard of living	White goods owned, house structure, vehicle ownership	Loan form, traffic department database
Farm		
Ownership details, along with area and boundary of land	Own or leased, digital land coordinates, land papers with farm land boundary	Revenue department, Agri department, copy of land document – with loan form. If formal lease – letter of owner.
Details on quality of soil, water availability	Type, quality, irrigated or rainfed, source of irrigation	Soil database, irrigation department database – canals, water table
Crop grown		
Crop grown and timing	Type of crop, sowing date, seed used	Loan form
Revenue history	Last 3 years – crop grown, yield and revenue	Loan form – validated by market prices, productivity of that area, weather conditions – derived outcome

Unit price for last season	Sale price – farmgate, mandi, contract production, contract purchase. Any value addition through conversion or affiliated o FPO before sales?	Loan form, market prices, derived value provides insight into value addition