

For the AgriSage project focused on the Karnal region (Haryana), the most effective workflow leverages the recommended research paper—"Digital Twins in Agriculture: Orchestration and Applications"—and well-established, open datasets covering soil, weather, yield, and market prices. Below is an end-to-end workflow with the paper and datasets, including direct links, fully tailored to your project needs.

1. Anchor Research Paper

Paper for Methodological Backbone

- **Title:** Digital Twins in Agriculture: Orchestration and Applications
- **URL:** <https://pmc.ncbi.nlm.nih.gov/articles/PMC11100011/>
- **Use:** Reference for system architecture, AI/IoT/Digital Twin orchestration, module integration, and workflow inspiration.^[1]

2. Datasets for Each Module

Below are the best, regionally relevant, open datasets for each AgriSage component:

A. Soil Health & Fertilizer Recommendation

- **Dataset:** Haryana Soil Health (Soil Health Card, government stats)
 - [indiastat.com Haryana Soil Data \(district-wise\)](#)^[2]
 - [Haryana Agriculture Statistics - Soil Health](#)^[3]
- **Type:** Use for soil parameter extraction, field-level recommendations.

B. Weather Data

- **Dataset:** Indian Meteorological Department (IMD)
 - [IMD Weather Data - Open Government Data \(OGD\)](#)^[4]
- **Type:** For real-time and historical weather inputs to models.

C. Remote Sensing for Environmental Monitoring

- **Dataset:** Sentinel-2 and Landsat Satellite Imagery

- [Google Earth Engine Dataset Portal](#)
- [Kaggle Indian Crop Remote Sensing](#)^[5]
- **Type:** For NDVI, pollution, drought/thermal stress mapping.

D. Market Price & Economic Analytics

- **Dataset:** Agmarknet Haryana/Crop-specific Market Data
 - [Agmarknet Official Portal](#)
 - [OGD Crop Price, Area, Production Haryana](#)^[6]
- **Type:** For price forecasting, AgriMarket contracts, ROI analytics.

E. Crop Yield Prediction

- **Dataset:** District-wise Wheat Production/Yield
 - [Karnal district cereal yield by season, 1997–2023](#)^[6]
 - [ICRISAT District Data](#)^[7]
- **Type:** Historical yield, used for time-series and ML modeling.

F. Water Pollution Detection (if implemented)

- **Dataset:** Water pollution & quality stats for Haryana
 - [CSSRI ICAR Karnal Annual Report 2023](#)^[8]
- **Type:** Use for smart robotics/IOT module validation.

G. Blockchain & Smart Contracts

- **Dataset:** Use pilot data or mock datasets from FAO or Haryana seed certification websites.
 - [Seed data \(India Open Government platform\)](#)^[4]

3. Workflow Steps

Step 1: Literature Review & Technology Reference

- Study "Digital Twins in Agriculture: Orchestration and Applications" for system design, module integration, and use case inspiration.^[1]

Step 2: Data Ingestion and Centralization

- Download relevant district-level (Karnal) soil health, yield, market price, and weather datasets from the provided links.
- Set up cloud storage (Firebase, MongoDB, or PostgreSQL) for central access.

Step 3: Module Setup

Module	Dataset Link	Model/Algorithm
Soil/Fertilizer Engine	Soil Health Card ^[2]	Classification/ML (Random Forest, XGBoost, OpenCV for image)
Weather Prediction/Inputs	IMD Weather Data ^[4]	Time series models (ARIMA, LSTM)
Price Prediction	Agmarknet, OGD ^[6]	ARIMA, LSTM, XGBoost, ensemble
Yield Prediction	District-wise yield ^[6]	CNN, LSTM, hybrid ML
Remote Sensing/NDVI	GEE Sentinel-2	NDVI/thermal band analytics, clustering
Water Pollution Detection	CSSRI Karnal ^[8]	IoT with sensors, OpenCV, thresholds
Blockchain/Seed	OGD Seeds Data	Smart contract pilot, Web3/Ethereum

Step 4: AI/ML Model Development

- Use the twin architecture: ingest real-world and satellite/sensor data, run AI/ML models for insight generation and prediction, then serve output to the app/dashboard.
- Reference model stacks and pipeline from the "Digital Twins in Agriculture" paper.^[1]

Step 5: Interface and Deployment

- Build dashboard/mobile app using React, Flutter, Django, or similar, referencing workflows and UI/UX patterns from the paper.^[1]

Step 6: Regional Adaptation and Validation

- Validate and demo platform output (recommendations, visualizations) using Karnal-specific data for highest practical impact.

4. Validation & Future Scope

- Compare results, workflows, and integration approaches with those detailed in "Digital Twins in Agriculture: Orchestration and Applications".^[1]
- Propose further work (e.g. extension to XR-based farmer/game training modules) with reference to the frameworks found in your attached XR-based education paper and the review's future directions.

5. Summary Table

Project Task	Research Paper	Dataset URL
Architecture/Workflows	Digital Twins in Agriculture: Orchestration and Applications ^[1]	https://pmc.ncbi.nlm.nih.gov/articles/PMC11100011/
Soil/Fertilizer Module	Same	https://www.indiastat.com/Haryana-state/data/agriculture/soil
Weather Module	Same	https://data.gov.in/dataset-group-name/weather
Price Prediction	Same	https://dataful.in/datasets/5750/
Yield Prediction	Same	https://dataful.in/datasets/5750/
Remote Sensing	Same	https://developers.google.com/earth-engine/datasets
Water/IoT Sensors	Same	https://cssri.res.in/PSSU/Publications/Annual Report/Annual Report 2023.pdf
Blockchain/Seed	Same	https://data.gov.in/dataset-group-name/agriculture-seeds

This workflow and set of resources will enable a robust, scalable implementation of AgriSage, fully grounded in the best current research and open data that directly matches your vision.^{[2][3][5][7][8][4][6][1]}

1. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11100011/>
2. <https://www.indiastat.com/Haryana-state/data/agriculture/soil>
3. https://agriharyana.gov.in/data/AYP_MSP_VitalOfAgriDoc/Vital_of_Agriculture.pdf
4. <https://www.data.gov.in/dataset-group-name/Field Crops>
5. <https://www.kaggle.com/datasets/thammuio/all-agriculture-related-datasets-for-india>
6. <https://dataful.in/datasets/5750/>
7. <http://data.icrisat.org/dld/src/biophysical.html>
8. <https://cssri.res.in/PSSU/Publications/Annual Report/Annual Report 2023.pdf>
9. AgriSage.pdf