A Presentation to DoLR GOI

Basemap Generation and Incorporation of Available Datasets to Create a Digital Land Records System

> Milind Sohoni, Aaryan Dangi, Asim, IITB 16th October 2024, New Delhi

Agenda/Themes

- Overview and Key Inputs (8)
 - Problem statement and solution framework
 - Legacy village survey/cadastral maps
 - Textual information: sub-survey maps
 - Ground control points: ground reality
 - Google farm plots: derived from satellite
- GIS Basemap Generation: Pipeline (9)
 - Global jitter
 - Global GCP-based georeferencing
 - Local georeferencing
 - Validation: statistics and heatmaps
- Metrics and Field Use (8)
 - Key metrics: shape, bund distance, etc
 - Field testing and validation
 - Land records rating and zone identification

- Incorporation of Other Datasets (5)
 - Fixing village boundaries through stitching
 - Reconciliation process: analysis and editing tools to resolve conflicts
 - Subdivisions: enabling subdivisions and merge operations over time
- Applications and Engagement (7)
 - Monitoring of land assets by ownership
 - Urban planning, agricultural schemes, etc
 - Methods of engagement

Introduction

Problem Overview and Key Input Datasets

Problem Statement

Objective: To reconcile legacy land record datasets, and generate **accurate village basemaps** which are (i) legally valid, and (ii) closely match ground reality (as far as possible), and digital land record systems to support DLR functions.

Existing Land Record Data

- Village Survey Map: Digitised set of cadastral/survey polygons
- Approximate Georeferenced Village Boundaries: Used to determine scale and position of village survey map
- Textual Sub-Survey Records: Parcel data at the finest level, mapping ID/area to ownership information

Additional Inputs

- Segmented Farm Plots (Google): Geography of individual farm plots, defined as the lowest granularity of agricultural land use in a given season
- Ground Control Points: Labeled
 locations of large rocks signifying land
 tri-junctions

Input 1: Village Survey Map



Paper maps: behind 66% of civil court cases in India (property disputes) They were scanned, converted to GIS geometries by the department. Is this enough? Digital drawing (Unscaled)

Overlaying the scaled and approximately georeferenced survey maps on modern satellite images finds a mismatch of over 25m across villages in multiple states!



A survey map in Maharashtra

Overlaid on satellite post geo-referencing

Input 2: Approximate Village Boundaries



GIS village cadastral map, approximately georeferenced and stitched into an area partition by MRSAC in the early 2000s

An approximate village boundary polygon is used to scale and position the survey map. These boundaries are derived from MRSAC cadastrals in Maharashtra.

Input 3: Google Farm Plots

POLYGON((75 29, 77 29, 77 29, 75 29))



Segmented farm plots are crucial for the discretization of land. They divide a region into polygons that are visually distinct in satellite images; ideally, the gaps between these polygons indicate changes in cropping pattern or physical gaps like bunds/roads.

Input 4: Sub-survey textual records

क्षेत्राचे । होजाचे । अ) स	त्र, एक		Idcidid	44-1			_		शेताचे र	-थानिक न	नाव :	परटाची	पारी	_
क्षेत्राचे । अ) स	आकारण	कव गी	वाते क्र.	ओग	वटादारार	ये नांव	क्षेत्र	आकार	पो.ख.	के.का.	कुळ.	खंड व इ	तर अधिका	τ
तामा सत एकुण व रेषे (त्या) (त्या (त्या) (त्या) (त्या (त्या) ((त्या) (((त्या)) ((((()))) ((((तमक है. तगवड़ यी त थी थे. ता यी थे. जीद उर्दा गवड अद थे. के ख थे. ते ख थे. ते ख थे. ते ख	9772 877 772 877 14.90 14.90 14.90 14.90 14.00 19.10 14.00	10		bes di		0.34.90	5.69	0.09,10	(1550) 9 9 1 1	कुळापे माव इतर अधिका त्वबित केरका विटपा केरफा 2/02/2017	व खंड र ारः नाही. ार क्रमांकः	2750 d fizicite	
रूने केंग	प्रदर क (1202)	(131 x 15 (1203 x 1-	3 x 182 x 447 x 1450 षद् जमील 3	189 x 282 x 2057 x	н 315 н 5 2446 н 20 7 Төтс эгійн	18 x 530 x 688 x 2756 गाव नम्न रेब आणि न	.548 x 685) ा बारा (1 दिवहवा (तग	× 730 × 7 पेकांची न तर करने व	31 x 1059 नोंदवही) मुस्थितील	× 1104) _{हा} ठेवणे) लिय	गमा आणि मु म, १९७१ यार	मापन चिन् तिस नियम व	r 1	
ते केप	शार क (1202) न क्रम	(131 x 15 (1203 x 1-) महाव गाव :- ख कि व उप	3 x 182 x 447 x 1454 पटु जमील म ालापूर (विभाग :	189 x 282 1 <u>x 2057 x</u> 1 x 2057 x 1 x 2057 x) (2446 x 20 (2446 x 20 (कार अमिर्ग)	मार्ड ४ ८३० ४ १८४४ ४ २७२७ गाव नमुन रेख आणि ने तालु	548 x 685) ा बारा (दिवहमा (तन का :- खाल	× 730 × 7 पेकांची न तर करणे व तापुर	31 x 1059 नोंदवही) मुस्थितील	× 1104) _ह ठेवणे) निय जि	गमा आणि मु म.१९७१ वाल नेल्हा :- रा	मापन चिन् ोस नियम व यगड	N]	
ते केप	रुर क (1202) न क्रम	(131 x 15 (1203 x 1-) गाव :- ख नंक व उप	3 x 182 x 447 x 1454 पटु जमील म ालापूर (विभाग :	189 x 282 x 2067 x 553748 101/2/4	x 315 x 5 2446 <u>x 2</u> कार अमिर्ग) रिकाखार	माइ ॥ 530 ॥ १८३४ ॥ 2750 गाव ममुन रेष आणि मे तालु- नील क्षेत्रा	.548 x 655) त बारा (1 दिवहमा (तप का :- खाल चा तपशीत	× 730 × 7 पेकांची न तर करने व गपुर	31 x 1059 नोंदवही) मुस्थितील	x 1104) ह ठेवर्ण) निय	गमा आणि भु म.१७४१ बाल नेल्हा :- रा लागवः	मापन पिन् ोत नियम र यगड डीसाठी	ह स। जल	शेर
ते केप	रुर क (1202) न क्रम	(131 x 15 (1203 x 1) (1203	3 x 182 x 447 <u>x 145</u> पद जमील म ालापूर (विभाग : मि	189 x 282 x 2067 x 553748 101/2/ब 14 पिका) (2446 <u>x</u> 20 (2446 <u>x</u> 20) (2446 <u>x</u> 20) (246 <u>x</u> 20) (2	418 x 530 x 638 <u>x 2750</u> गाव नम् न रेख आणि न तालु नील क्षेत्रा होटक पिके कारणान्दी	548 x 685) ा बारा (1 दिवरण (तप का :- खाल या तपशी य य क्षेप	x 7.30 x 7 पेकांची न तर करणे व ापुर न निर्मळ	31 x 1059 नोंदवही) मुस्थितीत पिकाखा	x 1104) ह हेवर्ण) निव जि नील क्षेत्र	ीमा आणि भु म.१५७१ बाले जेल्हा :- रा लागवः उपलब्ध जम्	मापन पिन् विम निवम न यगड डीसाठी नसलेली नीन	त्स । जल सिंचनाचे साधन	शेर
हो केवर 1170) मुमाप	श्रेर क (1202) न क्रम हंगाम	(131 x 15 (1203 x 1-) महारत गाव :- ख क व उप सिर्भणाच् संकेल कज्माक	3 x 182 x 447 x 1454 पर जमीन म ातापूर (विभाग : मि सिचित	189 x 282 3 <u>2057 y</u> 553748 101/2/ब (अ पिका सिचित	x 315 x 5 2446 <u>x</u> 20 7 कार अभिने) पिकाखात खालील (पिकाखा पिका पिकाखा पिकाखा पिका पिकाखा पिकाखा पिकाखा पिका पिकाखा पिका पिकाखा पिका पिका पिका पिकाखा पिका पिकाखा पिकाखा पिकाखा पिकाखा पिकाखा पिकाखा पिका पिका पिका पिका पिका पिका पिका पिक	मंत्र x 530 x आज x 2750 माव ममुन रेष आणि मे तालु लील क्षेत्र तालु लील क्षेत्र तालु सेत्र जल सिंचित	545 x 685) न बारा (1 दवरचा (तपशीत व त क्षेत्र अजल सिचित	x 730 x 7 पेकांची न त करणे व गपुर न निर्भेळ गिवकांचे नाव	31 x 1089 गोंदवही) बुभियतीत पिकाखा पिकाखा सिंचित	× 1104) _ह रुवणे) निव जि नील क्षेत्र सेचित	तमा आणि भु म. १७४१ बार्स मेल्ह्या :- रा लागवः उपलब्ध जम् स्वरूप	गापन चिन गेम नियम यगड डीसाठी नसलेली गेन धेत्र	त्र । जल सिंचनाचे साधन	शेर
त्र केप् (१)	सर क (1202) न कम हंगाम (२)	(131 x 15 (1203 x 1-) महारा गाव :- ख ांक व उप मिश्रणाच् संकेत क्रमांक (3)	3 x 182 x 447 x 1454 पद जगीत म रा नापूर (विभाग : मि जल सिचित (४)	189 x 282 3 x 2057 x 10 x 2057 x 10 1/2/ब 10 1/2/ब 14 पिका सिंधित (9)	× 315 × 5 2446 × 20 7 राज्यर अमिन) प्रिकाखात खालील । घ प्रत्ये मिकाचे नाव (६)	भाष भ 530 % अब ह 2750 गाव नमुन रेव आणि न तालु नील क्षेत्रा के जि का खाती जित सिंचित (७)	545 x 685) ा बारा (1 दवस्य (तपशीत व त क्षेत्र अजल सिंचित (८)	× 730 × 7 पेकांची न तर करणे व ामुर निर्झेळ पिकांचे नाव (९)	31 x 1089 मॉदवही) बुग्धितीत पिकाखा पिकाखा सिंचित (१०)	× 1104) _ह रुवर्ण) निव जि रील क्षेत्र सेवित (११)	तमा अगि भु म.१९७१ बाल मेल्हा :- रा लागवः उपलब्ध जम् स्वरूप (१२)	मापन चिन ति नियम द यगड डीसाठी नसलेली तीन क्षेत्र (१३)	*। जल सिचनाचे साधन (१४)	शेर (१९
हो केव 1170) नुमाप वर्ष	सर क (1202) न कम हंगाम (२)	(131 x 15 (1203 x 1- गाव :- ख कंक व उप मिश्रणाच् संकेल क्रमांक (3)	3 x 182 x 447 x 1454 पद जमीन म रा नापूर (विभाग : मियित सियित (४) इ.आर.	189 x 282 3 x 2057 x 10 5537 48 5537 48 10 1/2/ब 14/ पिका सिंपित (%) (%) (%)	× 315 × 5 2446 × 20 7 कार अमिन) पिकाखातील (प्रत्ये पिकाखे माव (६)	भार x 500 x क्व x 2750 माव नमुन तात्यु रील क्षेत्र तात्यु रील क्षेत्र तात्यु रील क्षेत्र तात्यु रील क्षेत्र तात्यु रिक पिके काखालीत कि काखालीत कि काखालीत कि काखाली कि काखालीत कि कि कि कि कि कि कि कि कि कि	543 x 685) ा बारा (1 दबरपा (तप का :- खाह या तपशी या तपशी व द से क्षेत्र अजल सिंचित (C) हआए ज.म	x 730 x 7 पकांची न तर करणे व तापुर निर्भळ पिकांचे नाव (९)	31 x 1059 गोंदवही) मुक्तिवतीत पिकाखा पिकाखा सिंघित (१०) हे.अर. च.म	x 1104) _ह ठेवणे) निव जि सील क्षेत्र सिंधित (११) है.आर.	ালা এগলি মু ম. ৭৬ গ বান সল্লা :- বা লাগবং उपलब्ध ডাফ स्वरूप (१२)	मापन चिन् गेत नियम न यगड डीसाठी नसलेली गिन क्षेत्र (१३) हेआ र	स् जल सिचनाये साधन (१४)	शेर (१५

Sr_No	Gat_No	Gat_No_Area_Ha_R_Sqm
1	1	0.20
2	2	0.15
3	3	0.12
4	4	0.09
5	5	0.04
6	6	0.03
7	7	0.02
8	8	0.03
9	9	0.02
10	10	0.06
11	11	0.03
12	12	0.03
13	13	0.04

Sub-survey 7/12 textual records for 101/2/B: total area is 0.34 Ha

Textual Gat area data for Anterweli, Taluka :- Gangakhed, Parbhani

Input 5: Ground Control Points (GCPs)



GCPs (black): Important junctions on the map

- Stones installed into the ground by the British are identified by local surveyors
- These stones are collected on ground via rovers
- Each GCP's corresponding vertex on the map is identified and labelled

Survey Basemap Generation

Review of Software Pipeline Developed

Input Data Sources for Basemap Generation



आकारबंद (गटाचे)	
गांवचा नमुना नं.1	
वाघलगांव ता.गंगाखेड	
गट नंबर	एकूण क्षेत्र हे.आर.
1	3.57
2	3.30
3	4.23

Vectorised survey map (Map 0) + Akarbandh



Approximately georeferenced village boundary



Google farm plots



Ground Control Points (GCPs)

Overview of Problem Statement



Main Idea: To match approximately georeferenced basemaps with farm plots so that areas are broady preserved and distortion is minimal

Pipeline Summary



Original plan: adjust the survey polygon shapes using vertex editing to better align with agricultural possession. Current: Upon request of the DoLR GoM, limited changes to local translation only (LG).

Pipeline Summary



Original plan: adjust the survey polygon shapes using vertex editing to better align with agricultural possession. Current: Upon request of the DoLR GoM, limited changes to local translation only (LG).

Step 1: Global Jitter



Input	Reference	Output	Processing
Original survey map, or Map 0 (green)	Village boundary (purple), Google farm plots (not shown)	"Jitter map", or Map 1 (red)	Map 1 is produced by performing global scale , translate and rotate operations to best fit survey map with respect to cadastrals, and then farm plots.

Step 2: Global GCP-based Georeferencing



Input	Reference	Output	Processing
Jitter Map (Map 1)	GCPs (orange), Farm plots (pink)	Global georeferenced map (Map 2)	Survey nodes are mapped onto GCPs, and global transformations are used to optimize GCP-survey distance; the method maximizing farm plot match with low distortion is chosen.

17

Step 3: Local Georeferencing: Allowing local transformations

Input	Global georeferenced map, or Map 2 (blue)					
Reference	Farm plots (pink)					
Output	LG Map, or Map 3					
Processing	 (a) Iso-jitter: Each survey plot is translated to minimize excess area with respect to farm plots. (b) Anchor polygons: Plots that meet quality standards are selected; namely excess area (<5%), farm intersection (>50%), area (>1.5 hectares). (c) Global spline: Nodes of anchors 					
	are used as a reference to spline transform the map.					



Anchor polygons (red) are translated to fit farm plots, and Map 2 is transformed to match these polygons. 18

Validation



		Number of Survey Plots G				Georeferenced Refined Output				Possession Boundaries		
Sr. No.	village	Total	Non-Integer	Vaild (unique, alphanumeric survey numbers)	Akarbandh Validity	% of survey numbers with over 95% farm rating	% of survey numbers within 5% of akarbandh area	% of survey numbers within 5% of geo-referenced area and perimeter	% of survey numbers within 5% of geo-referenced area and deviation	% of survey numbers within 5% of akarbandh area	% of survey numbers within 5% of geo-referenced deviation	% of survey numbers within 5% of geo-referenced area and deviation
0	deolanabk	184	12	138	103	39.86	32.61	45.65	23.19	34.78	20.29	16.67
1	waghalgaon	67	2	65	60	50.77	61.54	58.46	30.77	63.08	33.85	32.31
2	khatnapur	23	4	19	18	10.53	42.11	42.11	10.53	26.32	26.32	21.05
3	shekhapur	66	4	57	11	28.07	10.53	49.12	14.04	8.77	7.02	5.26
4	matargaon	41	4	36	33	58.33	63.89	77.78	61.11	61.11	63.89	61.11
5	deolanakh	41	2	36	35	50.00	63.89	66.67	47.22	69.44	41.67	41.67
6	dagdagad	52	3	49	48	83.67	71.43	75.51	63.27	75.51	69.39	65.31
7	kharburdi	59	3	40	0	50.00	0.00	75.00	47.50	0.00	20.00	15.00
8	gopa	78	3	75	73	76.00	74.67	62.67	46.67	82.67	64.00	53.33
9	virshi	84	3	79	76	69.62	65.82	68.35	43.04	78.48	46.84	43.04
10	nirgudibk	60	5	54	52	9.26	51.85	53.70	16.67	51.85	7.41	5.56

19

Input	Reference	Output	Processing
LG Map (3)	GG Map (2), Farm plots, GCPs	Heatmaps and Statistics	Various quality metrics such as shape deviation, textual area difference, and average distance to bund are added to the survey map tables. Village statistics tables are also generated.

Similar issues in Telangana: mismatch with ground reality



DMV_Code_235	6005
▼ V_Name	Kurmaidu
(Derived)	
(Actions)	
OBJECTID	394
Parcel_num	293
Remarks	NULL
V_Name	Kurmaidu
M_Name	Chintha Palle
D_Name	Nalgonda
DMV_Code	2356005
Shape_Leng	0.00906448206003
LandUse_Fr	Dry
LUCodeTiff	08
STCode	0
Shape_Le_1	0.00888692423572
Shape_Area	4.13068249945e-06

A village cadastral map with attributes in Telangana



Zoom-in: vertex distance between survey map (blue) and farm plot (red) gap is 26 meters 20

Statistics and Validation

Formulation of metrics and application via heatmaps on generated basemaps

Key Metrics

Textual Metric: Area difference



GIS Metric: Shape deviation with original map



Deviation (shaded area) = 100 * ((A∪B)-(A∩B)) / (2*A) = 11 %

Possession Metric: Average Distance to Bund





The average distance to farm bund is computed for each survey edge individually, representing the distance to bund on the field.

Heatmaps for a Good Village: Taluka Bhatkuli

0 - 3

3 - 5

5 - 100





Shape deviation heatmap for "Indapur" village in Bhatkuli



Average DTB: Appropriate fit found with farm plots through local transformations

Heatmaps for a Bad Village: Taluka Gadhinglaj

0 - 3

3 - 5

5 - 100



Average distance to bund heatmap for "Mankadevi" village in Gadhinglaj

Shape deviation heatmap for "Mankadevi" village in Gadhinglaj



For bad talukas, field work is needed to reconcile internal boundaries

Visualizing DTB Metric Change across the Pipeline



Statistics for villages in Pilot MoU

A summary of LG OUTPUTS for chosen villages across 5 talukas for Phase 4 delivery.

village	No. of survey plots	No. of anchor polygons	% of survey numbers within 3% of georeferenced area, perimeter and deviation	% of survey plots in LG with average bund distance < 2.5m	% of survey plots with average bund distance < 5m
akoli	92	17	86.67	34.44	70.00
shipgaon	58	12	100.00	55.56	72.22
<u>kanfodi</u>	66	14	96.23	30.19	66.04
jasapur	54	6	100.00	17.65	50.98
indapur	41	9	97.14	51.43	80.00
harangul	64	14	62.30	16.39	40.98
bhaddarpur	63	13	100.00	41.67	66.67
waghoda	39	4	94.74	15.79	36.85
dahatonda	13	4	92.31	23.07	69.23
banpimpla	43	7	90.00	12.50	42.50
kharbi	31	7	100.00	64.29	78.57

Overall Observations

- Total number of villages with successful map generation: **107**
- Total number of invalid villages:
 20 (largely torn survey map or missing village boundary data)
- Area-perimeter-deviation metric: almost all polygons consistently pass 3% constraint
- Average DTB metric: variable from taluka to taluka; generally good in areas like Bhatkuli, but poor in Gadhinglaj

Field Validation



A village in MH: coloured by quality



Black: points measured on the field

Red: georeferenced map

Avg. error: 2.2m

There was a large correlation between rating indicators developed and the actual field error in pilot villages.

MoU-I achievements

Execution

- Data validation and basemap generation of 120 villages in 5 districts
- Heatmaps and statistics created for all villages, to help with GIS verification and field use
- Pilot software developed and deployed on cloud

F1 F2 F4



- Metrics developed: DTB, deviation, etc
- Bad farm plot recognition
- Farm bund generation
- Void handling: roads, rivers, gaothans
- Distortion-limiting LG transformations





- Data Validation Protocols
- GCP Labeling and Spatial Distribution SOPs

Processes

• Software Training and Execution

MoU Addendum: Scale, Upstream, Downstream Integration

Module 1: Production of **Accurate Georeferenced Maps** for 6 Talukas (773 Villages)





Module 2: Development of Land Record Rating for Individual Survey Plots based on available land record data for villages in Pilot talukas



Module 3: Fixing of Village Boundaries for selected talukas

Digital Land Records System

Creating a complete land record management system that supports DLR functions

Module 2: Zone Identification for Internal Reconciliation



Identification: Plots that require internal reconciliation



Average DTB heatmap: original map



Merged boundaries: polygons merged across red edges



Average DTB heatmap: merged map

Module 3: Stitching Village Boundaries



Objective: To create a complete village partition, made through stitching village borders where appropriate and amending boundary plots without distortion.

32

Junction-Midline Mapping and Issues with Scale





The border tri-junctions are mapped to the shared boundary, ensuring that roads remain aligned.

Issues to be tackled while performing bottom-up village survey map stitching:

- Road alignment: Preserving continuity of roads spanning across villages
- Unmapped area: Detecting village roads and other government land that is unmapped at boundaries

Beyond Basemaps: Reconciliation through Real-time

Tools



A land reconciliation system: The way to systematically resolve boundary conflicts through real-time visual metric tracking and editor tools

Use Cases and Engagement

Direct and indirect applications of produced basemaps and developed systems

Applications



E-Mojani: Automatic Fixing for 50%+ Plots (High LRR)



Urban Planning: Better base maps are useful for road mapping, etc



Agricultural Monitoring: Insurance and Subsidy Transfer



Efficiency in Land Development and Transfer

Land Records Framework: Overview



Introduction to Terrastack Tech. Pvt. Ltd.



Milind Sohoni

- Co-founder at TTPL
- Professor CSE, IIT Bombay
- Head CTARA, IITB ('11-'16)
- Executed projects like PoCRA, UMA, etc with state and national governments



Incubated by SINE, IIT Bombay in February 2024

Prototyping grant given for land records work: October 2024



Aaryan Dangi

- Co-founder at TTPL
- B. Tech. IIT Bombay CSE
 ^{'25}



GISE Geospatial Information

HUB

Science & Engineering

Team invited to present our work on land records: November 2023

Introduction to Terrastack Tech. Pvt. Ltd.



Asim Rama Praveen

- Core Team, TTPL
 - Software
 Engineering at
 Greenplum, etc
 for 20+ years



Bharat Adsul

- Core Team, TTPL
- Professor CSE at IIT Bombay



Lisan Kadivar

- Core Team, TTPL
- IIT Bombay
 CSE '25

Our Mission

We want to solve problems for individual farmers and landowners, and the agencies that support them.

Our vision

Our aim is to build an end-to-end digital land records system, that allows for the modeling and analysis of land assets and events.



Phase-wise delivery: MoU between GoM and IIT Bombay Multi-partite engagements: building and scaling state-wise land record systems

What do we need to know to work with other states?

- Structure and quality of existing maps, and their formats
- Availability and accuracy of Ground Control Points (GCPs)
- Choice of metrics for optimization
- Current formalisation processes and legal status of already "modernised" land parcels
- Procedures for reconciliation of disputes and subdivision of land
- Decision on a reliable proxy for ground reality

A Peek into Our Team

