



Land Records Information System Data Extraction

Objective

The primary objective of this exercise is to **convert Record of Rights (RoR) PDF files, generated through the front-end of the existing LRIS system, into structured Excel files** that conform to the prescribed schema of the new LRIS.

This approach is being pursued because the backend of the legacy LRIS system is inaccessible currently, and manual data entry of ~8 crore pages would be time consuming, inefficient and error-prone. By automating PDF-to-Excel conversion, we aim to:

- Enable seamless ingestion of land records into the new LRIS database.
- Minimize human intervention and reduce the scope for data entry errors.
- Establish a repeatable and scalable process to digitize historical records and keep pace with ongoing updates.

This will enable the Revenue Department to upload structured Excel files directly into the LRIS system, ensuring smooth ingestion of land records **without** overburdening field staff with manual entry tasks. Which will expedite the implementation of Agri Stack in the UT.

Data Extraction Requirements

12	11		10	9	8			7		6	5	4	3	2	1
kaifiyat	havalat intakal		mutalbat tashari maal va sewai	lagan jo mujariya ada karta hai meh shara va tadad	raqba bakiyad kisam			number khasra v nam khet		vasayil abapashi	nam kashtakar meh ahval	nam malik meh ahval	nam tarf ya patti meh nam number dar	number khata	number khatvat
	kisam	number			kisam zamin	marala	kanal	hal	sabik						

1. Split rows in the pdf based on column 2 (number khata basis).
2. Split the column 5 (nam kashtakar meh ahval) in pdf into 5 columns i.e. Name, Parentage (W/o S/o D/o), Caste, Residential Village (sakin) and Remarks. This information can be coded based on the urdu words that exist in the cell.
 - a. For example, “kasht sahid v singh pisar attar singh kaum sukh sakindeh gair morosi” -> cultivator sahid v singh s/o attar singh caste sukh Resident of the village legal non heirs.
 - b. Term mapping for Record of Rights Translation

Original	Translation
----------	-------------



Kasht	cultivator
Pisaran	son of
Pisaran	sons of
Dukhtar	Daughter of
Dukhtaran	Daughters of
Zoja	Wife of
Byuh	Widow of
Kaum	caste
Namalum	unknown
Sakin	Resident
Sakindeh	Resident of the same village
Bayaan	seller
Mushtari	buyer
Wahib	Gifter
Mohoob allya	Receiver of the gift
Decree dehende	ordered to give by the court
Decree gerendaye	ordered to receive by the court

3. Each individual number in column 7 (khasra number) needs to be split into different rows, mapped to the same information in column 5 (nam kasht kar me ahval) . If a farmer in col 5 owns multiple land parcels (represented by khasra numbers), the last row is the sum of the sizes of each land parcel.

- a. PDF vs Excel comparison. Data to be in pdf format.

4	1	162			kasht sandhu singh pisar sain kaum rakwal sakindeh alati
8	1	166			
12	2	2	kitta		

Image 1 - Extract of Jamabandi of Gujral Village in the original PDF format

4 8 12	1 1 2	162 166 2
--------	-------	-----------

Image 2 - Excel output of the conversion of Image 1, ideally they should have been 2 separate rows which are getting added up in the third row, they are all mapped to the same farmer

- Given the scale of 6000+ villages, each having 50-500 pages of land records, automation of the data extraction process is essential to ensure efficiency, accuracy, and scalability.
- On Sheet 1 L8 and L9 represent the same person and the same land parcel but they are written in 2 different cells. This is because of duplicacy in khata numbers in col 0 in excel. We need to eliminate duplicacies in col 0.
- Each land record page is currently in different tabs. We essentially need to build a farmer registry that maps the demographic details of farmers (col 5) of each village to their land ownership information (in col 7 and 8).

How to read a Jamabandi Document

A jamabandi is an official government record maintained by the Revenue Department. It provides details about a piece of land including extent, type, Ownership, and cultivators. It is a vital legal document for determining revenue, assessing land taxes, and verifying land transactions, and is accessible online in many states.

2017-18 : sal		Jammu :zla		Jammu West :thsil		Gujral :babata mouza		jamabandi			
12	11	10	9	8	7	6	5	4	3	2	1
khafyat	havalat intakal	mutabba b khata masal va sevai	lagan jo majariya ada karta hai meh shara va tadad	raqba bakayad kizam		number khata v nam khata	vaasil abapaki	nam kashitkar meh ahval	nam malik meh ahval	nam tarf ya pati meh nam number dar	number khata
kiyam	number			kiyam zamin	marala	kanal	hal	shabli			number khevat
hemba nana	1343	0 masal 0 seva	gora nahri	16	0	156			kashir sahib v singh pitar amar singh kum vakh sakindeh gair morosi	de mohamme 20 kasa 8 mara 0 marabba soor pitar bagh ali kum gujar dasekri sakindeh	1
		0 masal 0 seva	gora nahri	4	1	162					2
		0 masal 0 seva	nahri	12	2	160	kita		rahwal sakindeh alati		3
		0 masal 0 seva	nahri	5	2	347/91			duwa singh pitar sa malum kum brahman vakh alati kashir bejariya mulezam		4
hemba nana	795	0 masal 0 seva	hil nahri	0	1	382/144			kashir haram singh pitar jamiat singh kum vakh sakindeh alati		5
		0 masal 0 seva	nahri	8	4	388/326			iqbal singh , harban singh , mamohan singh pitaran bachan singh kum namalun murenya 3		6
		0 masal 0 seva	nahri	16	3	92			alhaf badava badekar		7
vasist	795	0 masal 0 seva	gair mumakin kahi	2	0				iqbal singh 1/3 mamohan singh 1/3 harban singh 1/3 pitaran bachan singh kum namalun murenya 3		8
		0 masal 0 seva		18	3				alhaf wabhan balak singh 1/1 pitar sant singh kum namalun mahub aliya balak singh pitar sant singh kum namalun morosi 3alhaf		
wasit morosi dfa 3 alhaf	581	0 masal 0 seva	nahri	0	1	92			iqbal singh 1/3 mamohan singh 1/3 harban singh 1/3 pitaran bachan singh kum namalun murenya 3		
		0 masal 0 seva	hil nahri	12	1	382/144			alhaf wabhan jagdev singh 1/1 pitar balak singh kum namalun mahub aliya jagdev singh pitar balak singh kum namalun morosi 3alhaf		
		0 masal 0 seva	nahri	17	2				harban singh pitar chor singh kum namalun morosi 3alhaf		
		0 masal 0 seva	nahri	9	4						
				8	20		9	kita		mijan khevat	

Jamabandi in Jammu & Kashmir is written in Urdu and then transliterated into other Indian languages and for that reason they are to be read from left to right.

The format contains 12 columns, given page provides details of Jamabandi for the year 2017-2018 (Jamabandi is done on a given date and provides a snapshot of land ownership at that point) for Zla (District): Jammu, thsil (Block): Jammu West, Babata Mouza(Revenue Village): Gujral, all these details are mentioned in the header of the page that precedes the Jamabandi table.

Within the table the 1st (rightmost column) is that of a Khevat (a revenue village is further divided into subdivision) number 1, and all holdings in the Khevat are provided a Khata(account) number. Column 3 Nam tarf ya pati meh nambar dar (Name of the Nambardar, a government official who represents a group of land owners), Column 4 Nam malik me ahval (it represents the old owners of

the entire Khevat the land holdings of the old owners should add up to the entire Khevat's surveyed land as is highlighted in yellow), as an example consider Khata number 4 of Khevat number 1 Column 5 contains the details of the people who bear title to the land. Land belongs to "kasht sahid v singh pizar attar singh kaum sukh sakindeh gair morosi" which means Cultivator Sahid V Singh Son of Attar Singh Caste Sikh Resident of the village, legal non heirs. Column 6 conveys Vasail Abapashi (means of Irrigation). Column 7 has Khasra Number (survey number of the land) which contains hal - the present number of the land and the sabik - the previous survey number of the land in this example the Khasra number is 156. Column 8 contains Raqba Bakaid Kisam - Area along with type in this example the area of the land is 0 kanal 16 marala and land type is gora nahri. Column 9 provides lagan jo mujariya ada karta hai me sharah wa tadad (Description and amount of tax paid). Column 10 contains mutalba b tashari maal va sewai (Cess Levy on Goods and Services). Column 11 contains havala intequal (mutation reference) which implies the transactions that have taken place on the land in this case mutation number 1343 which is a hibba nama (gift deed) has also taken place on the land to enact the transfer of ownership. Column 12 is Kaifiyat (Remarks).

Key Asks

The conversion of the Record of Rights from PDF is facing a few problems that need to be resolved before the excel output can be ingested into the new LRIS system and also for the data to be consistent. Following are the problems in current conversion methods.

12	11		10	9	8			7		6	5	4	3	2	1
kaifiyat	havala intakal		mutalba b tashari maal va sewai	lagan jo mujariya ada karta hai meh shara va tadad	raqba bakiyad kisam			number khasra v nam khet		vasay il abapashi	nam kashtakar meh ahval	nam malik meh ahval	nam tarf ya patti meh nam number dar	numbe r khata	numbe r khevat
	kisam	number			kisam zamin	marala	kanal	hal	sabik						
	Rows need to be split to the corresponding entry in column 8	Rows need to be split to the corresponding entry in column 8			Column borders are misaligned	Column borders are misaligned	Column borders are misaligned	Column borders are misaligned	Column borders are misaligned		Comma is to be used as a delimiter and the row needs to be split into Name(string followed by Kasht), Parentage(string followed by pizar,pisaran, dukhtar,duktaran,zoja), Caste(string followed by kaum), Ownership Type				