

Putting the AI in Notarial Archiving

Author: **David Mizzi**

The Notarial Registers Archive, once the Cinderella of local archives, boasts over 20,000 documents of deeds spanning back to the 15th century. But trying to find anything is like looking for a needle in a haystack. What if we could use AI to streamline the process?



There's a joke about the legality of chopping my neighbour's tree if a branch crosses over my property. And if I can chop off an encroaching branch... well, anyway, the point is that property disputes are unsurprisingly common. After all, good fences make good neighbours. But how do we decide where the property line actually is? And how can I prove that he's secretly been edging the wall that divides our fields closer?

If you've ever had a tiff with your neighbour, you'll know that the first step is to pay a visit to a notary. After that, it's off to the Notarial Archives to try and find the original (or a copy) of the deeds which determined the exact property boundaries. But with over 2 km worth of historical documents, there's a chance your neighbour dies of old age before the dispute can actually be settled.

MORE THAN JUST NOTARIAL RECORDS

However, the notarial archives are so much more than just a means of settling squabbling neighbourly disputes. After all, the work of notaries goes beyond properties and buildings. Essentially, any legal or social contract throughout history needed to be drafted by a notary, including insurance, dowries, marriages, adoption, as well as inheritance. The notarial archives offer a snapshot of everyday life across time. 'They are historical documents, so they benefit our historical heritage. There is the *Kantilena*, the oldest Maltese poem. Then there is the research potential – a lot of information still needs to be tapped. Then there is current research, for example, you might want to learn about your property,' explains Dr Joan Abela, Director of Operations at the Notarial Registers Archive and Senior Lecturer at UM's Department of History. Whether you're a recent homeowner looking into how far you can extend your property, trying to find out about your family tree, or conducting research into 17th-century fashion trends, the Notarial Archives

are a treasure trove. But, with over 20,000 volumes written in legal Latin, the challenge lies with actually finding (and understanding) what you're looking for.

Hence, the NotaryPedia project. For Dr Abela, Prof. Charlie Abela (Lead Investigator of NotaryPedia and Associate Professor at UM's Department of Artificial Intelligence), and Charlene Ellul (Research Support Officer on NotaryPedia, and Chief Knowledge Officer at the Notarial Registers Archives), the idea is to digitise the archives and combine them with AI. 'We're trying to solve a number of problems, mainly, making the content accessible to the many researchers who visit the archives on a daily basis,' explains Prof. Abela.

'Usually, the first thing these researchers see when entering the archives is a catalogue,' explains Ellul. 'The catalogue gives an overview of the metadata pertaining to a document. Examples of metadata include information about who created the document, the dates, and the number of folios. However, to search for something specific, such as a specific area of land, you'd start with the notary and work your way manually through every single document until you find what you need,' she explains.

And if you thought that slugging through legal documents was a chore, remember that most of these documents are written in a mix of legal Latin (because Latin wasn't already hard enough), a sprinkle of Sicilian and a few Maltese-isms thrown in for good measure.

TEACHING LATIN TO YOUR AI

While ChatGPT can probably handle drafting a passive-aggressive note to your neighbour, having an AI understand the entire contents of the notarial archives requires a methodical approach. 'There is a pipeline,' explains Prof. Abela. 'First, we start from the availability of digitised manuscripts. These would be the manuscripts that have been photographed using special equipment ➤



Repository room at the Notarial Registers Archive, with shelves of bound historical registers.
Photo courtesy of the NotaryPedia Team

at the archives. From those digitised manuscripts, we extract the textual element within each document.' Yet, it is not just the text that makes these volumes valuable. 'Researchers might be interested in other aspects of

the manuscripts, such as the quality or material of the paper, or even the doodles found in the margins,' adds Prof. Abela. So this data also needs to be recorded.

Once the digitised documents are transcribed into a machine-readable format, the next step is to verify the transcription. Paleographers – persons who read ancient texts – comb through the transcribed document before passing it on to the next phase of the pipeline – training the AI to extract information. 'Notarypedia will be a growing knowledge structure. We are expecting more documents to enter this pipeline, so we need this kind of continuous learning mechanism in place,' says Prof. Abela. Training the AI to extract that information is Ellul's area of expertise. All of the extracted information is then fed into a knowledge graph, but what exactly is a knowledge graph?

'Wikipedia, for example, is a knowledge graph. All those hyperlinks connect various topics together,' explains Ellul. A knowledge graph has two elements – the schema and the data layers. The schema is how we understand the world, including its places, persons, and objects. The data layers, on the other hand, are the actual information. So the data 'Malta' would be mapped to a place, and a nobleman called 'Johannes' is mapped to a person. In this way, the AI understands that Malta is not a person or an object, but a place. For NotaryPedia, 'the schema is how we see the notarial world, but from an abstract perspective,' says Ellul, 'while the data is what we've extracted from the deeds. For example, a relation between two persons in a deed is mapped into the schema, the way the facts are established in the notarial world.'



A notarial register displayed at the folio of a deed, illustrating the archival structure that underpins research.
Photo courtesy of the NotaryPedia Team



A schema-level view of NotaryPedia's AI-driven knowledge graph, where deeds become structured, linked data.
Photo courtesy of the NotaryPedia Team

Think of the schema as the grammar rules or definitions. 'You have these vocabularies that define what an actor is, what a notary is, what a place is. All that information is kept separate,' says Ellul. It is a standardised semantic layer, and it helps the AI to make sense of the information it is being given. 'It's not enough to say "this person is a notary," but what does notary mean? Basically, everything needs to be defined and a meaning given.' However, there is an advantage to taking this methodical approach. By using internationally accepted vocabularies, NotaryPedia can be linked to other archives.

'Let's say I have Johannes the nobleman, and Johannes has been mentioned in a manuscript found in another archive. This archive is also using a graph, so we can easily create a link between the archives. Ultimately, it allows archives to be linked, rather than be siloed,' explains Ellul.

'By linking content between two archives, the knowledge graph is going to be much richer for researchers both locally and abroad. People who have been mentioned in notarial deeds would possibly also be mentioned in other documents, so we want to be able to link to other local archives, such as the Cathedral archives,' adds Prof. Abela.

DEFINING DEEDS

While AI is the ideal candidate for processing large swathes of data, it turns out even AI struggles to read certain handwriting. Notaries often employed a number of scribes so that while two documents might be signed by the same notary, they would have completely different handwriting. As a result, the AI

might be able to recognise one deed, but not the other.

In the past, notaries were also extremely economical with their paper. 'Paper was an added expense, especially since it needed to be imported, so the notary tried to use all the available space. Apart from being bound by law to leave wide side margins, notaries were also required not to leave empty spaces,' explains Dr Abela. Deeds might have pages where one deed ends and another, completely unrelated deed begins. 'There was even a case when one notary treated a deed as his own personal diary, describing an earthquake that occurred while he was writing the deed,' adds Ellul.

Besides these exceptional cases, there is the sheer volume of texts that need to be processed. However, the team is both optimistic and determined. 'We are creating a model, so others can build upon it. We are investing a lot of time establishing the foundations of the model,' says Dr Abela. While the pipeline can be adapted to different types of archives, Ellul points out, 'it is also important to have the guts to do something that works for us.'

Prof. Abela concludes, 'This is a project close to my heart. Even though I come from the realm of AI, whenever I speak to Charlene or Joan, I'm always learning a lot. It's important that researchers aren't just limited to their area. Expanding into different fields allows you to apply what you know and grow and learn.' **T**

NotaryPedia is a 3-year project which started in January 2025 and is a collaboration between the Notarial Archives Foundation and the University of Malta. The project is funded by Malta's Ministry for Culture, Lands and Local Government.