



Historians want to build a virtual archive for Venice, pictured here in the eighteenth century.

VENICE 'TIME MACHINE' PROJECT SUSPENDED AMID DATA ROW

Disagreements between international partners leave plans to digitize the Italian city's history in limbo.

By Davide Castelvecchi

Like the city itself, an ambitious effort to digitize ten centuries' worth of documents that record the history of Venice is at risk of sinking. Two key partners have suspended the Venice Time Machine project after reaching an impasse over issues surrounding open data and methodology. The State Archive of Venice and the Swiss Federal Institute of Technology in Lausanne (EPFL) say they have had to pause data collection, and the archive's director has raised questions about the usability of the information that has already been collected.

The project sought to digitize documents that stretch over 80 kilometres of shelves in the state archive. These record the minutiae of the city's administration – from financial transactions to citizens' addresses and family connections – during its heyday in the Middle Ages and the Renaissance as a republic that dominated trade in the eastern Mediterranean for centuries. Many are written in Latin or the Venetian dialect, and have never been read by modern historians.

The goal was to make this information freely available online to researchers worldwide. The project also aimed to push the state-of-the-art in text-recognition technology for handwritten documents, using machine learning to automatically read millions of pages and tag their contents so that historians could perform quick searches.

The project was launched as a collaboration between EPFL, the State Archive of Venice and the Ca' Foscari University of Venice. In 2014, all three organizations signed a non-binding memorandum of understanding on how the work would be conducted.

However, the original agreement left out crucial details on the research protocols, according to a 19 September press release from the archive announcing the suspension. In particular, it didn't specify the type of licensing that would regulate researchers' use of the digitized data – which must also comply with Italian law, says the archive's current director, Gianni Penzo Doria. He adds that he tried to jump-start negotiations for a detailed contract after he took up his post in August, but the two sides came to an impasse. The decision to halt

the project was inevitable, he says, and mutual.

But on 23 September, EPFL issued its own sharply worded press release saying that the archive had suspended the project unilaterally, and that EPFL was surprised to learn of the decision from the archive's website. "I think it's essentially a misunderstanding," says Frédéric Kaplan, a computer scientist at EPFL and director of the Venice Time Machine. He adds that the disagreement could potentially have been resolved by face-to-face meetings between the collaborators, but that so far all discussion had been by teleconference.

'Useless' files

Meanwhile, the fate of the 8 terabytes of data accumulated over the past 5 years – from around 190,000 documents – is unclear. Penzo Doria claims that from the point of view of archival science, "these files are useless", because their digitization did not follow guidelines set by the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project.

These guidelines mandate the scrupulous recording of information that certifies the provenance of each document, and require that a record of such information be kept in the metadata that comes with each file. This serves as a sort of electronic signature that ensures the long-term preservation and validation of a digital file. According to Penzo Doria, the EPFL researchers who made the scans did not document how they collected such information – or, if they did, they didn't share such documentation with collaborators at the archive.

Kaplan says the researchers did collect metadata, but based their methodology on a the International Standard Archival Description guidelines from the International Council on Archives. He says that EPFL researchers followed procedures established by the state archive's own staff. Kaplan also says that he provided documentation on the metadata in an e-mail to Penzo Doria's predecessor, Giovanna Giubbini, in February 2019. Penzo Doria and Giubbini both told *Nature* that they never received this documentation.

Raffaele Santoro, who was director of the State Archive of Venice in 2014 when the memorandum on the Time Machine project was signed, says that he doesn't know precisely how workers collected the metadata, but he assumes that they are scientifically valid because the archive's own staff were closely involved in the process. To make the documents that have already been digitized compliant with additional standards, one could simply add more information to the metadata, he says, "without any need to do it all over again".

Kaplan says he is hopeful the project can get back on track if the two sides meet to discuss new terms in person. "EPFL sincerely hopes the meeting will happen soon," he says.