Time Machine in the running to become a FET Flagship

2018 will be a pivotal year for the Venice Time Machine project. Not only does it mark the halfway point of the initiative launched in 2012, but it will also include two major steps forward: a drive to expand the project to cities across Europe, and the release of over two million documents that have already been digitized to historians and the general public.

“Any records that were kept before 2000 basically don’t exist, because we have no means of viewing them,” says Frédéric Kaplan, head of EPFL’s Digital Humanities Laboratory. This radical statement reflects Kaplan’s concern that in the not-too-distant future, only information recorded in electronic form will be accessible – meaning information in all other formats will fall by the wayside. “We urgently need to bring our archives into the digital age. We mustn’t lose contact with the past.”

Bringing the past into the digital age is exactly what the Venice Time Machine (VTM) project sets out to do. It aims to build a multidimensional model of the city that spans the past millennium, using millions of historical documents stored in a variety of formats. Launched in 2012, the project is now reaching its halfway point in a year that will include two major steps forward.

Mapping 5,000 years of European history

The first is a drive to expand it to cities across Europe under a broader Time Machine initiative. In a manifesto published in 2016, Kaplan called for universities across the continent to join forces to share the region’s cultural heritage. His paper received widespread acclaim, prompting several EU commissioners to champion his idea. Kaplan’s manifesto became the cornerstone of a proposal for the Time Machine project to become part of the EU’s FET Flagship program, which provides €1 billion in funding over ten years to large-scale research initiatives.
Time Machine aims to create a large-scale simulator for mapping 5,000 years of history. 165 partner institutions from 32 countries have formed a consortium to develop new technology for digitizing, analyzing, accessing, preserving and sharing historical and cultural documents. The FET Flagship application was sent in this month with a response expected next year.

Several Time Machine projects already under way
Some cities decided not to wait for the EU’s green light to launch their Time Machines. Amsterdam, Nuremberg, Paris, Jerusalem, Budapest and Naples have all decided to dig through their archives and create a huge database linking their history on a national and European level. Why? We spoke with Julia Noordegraaf, Professor of Heritage and Digital Culture at the University of Amsterdam, and Sander Münster, who heads the Department of Media Design at Dresden Technical University, to find out. Read the interview in the press release.

With the application submitted, the next step will be to create the online interfaces that will allow people to browse the over two million documents that have already been scanned. In the five years since the project began – or two since the researchers started digitizing documents – they have scanned 190,000 records from the Venice State Archive, 720,000 photographs from Fondazione Giorgio Cini, and 3,000 books on Venice’s history held in the city’s main libraries. “That’s less than 1% of everything that’s out there. It’s enormous!” says Kaplan. It also led to the discovery of an art trafficking ring between the Netherlands and Italy.

Tens of thousands of annotations
To scan this unprecedented volume of documents, the researchers first had to develop technology that could quickly digitize kilometers of archives, some dating back 500 years. The devices they built include a semi-automated scanner for fragile manuscripts and documents in large formats, a robotic arm to turn the pages of books and journals, and a high-speed scanner, developed in conjunction with Adam Lowe at Factum Arte, that can digitize a photographic record front-and-back in just 4 seconds.

In addition to the digitization effort, historians, paleographers and archivists hand-wrote over 160,000 notes on the documents indicating names, places and keywords. Their annotations were used to create an automatic handwriting recognition system that could extract information from the documents. “Our system had to be able to handle many different types of writing and abbreviations, since they weren’t standardized at the time. We showed the computer an array of examples so it could learn how to learn,” says Sofia Ares Oliveira, a researcher at the Digital Humanities Laboratory. The system she and her colleagues developed can transcribe unannotated words with an error rate of 4% per character. “It’s like when you enter a word with a typo into a search engine. Our system is smart enough to come up with fairly accurate results and give you a good idea of what the text is about,” she adds.

Three programs for delving into digital data
The project team will launch their search engine, called Canvas, by this summer. It will let people view original versions of scanned records from the State Archives and search the entire database of digitized documents by keyword, such as a name, place or date. It works like Wikipedia in that authorized users can correct scanning errors they find in the files, and thereby continually improve the service.

A second task was to digitize the vast photo library owned by Fondazione Giorgio Cini. Some 370,000 of the Foundation’s million photographs have been scanned so far. And here again, researchers added thousands of notes by hand to train the computer. That resulted in a search engine called Replica that can search not only text, but also images by establishing connections between them. “The machine learned to form links between the photographs’ composition, the position of the
people in them, and any other items or patterns. That allows the machine to spot similarities among the images – paving the way for art historians to flesh out their history," says Benoît Seguin, the EPFL student who developed Replica as part of his thesis, which he will defend this summer.

And a third task was to tackle the scientific books and journals stored away in Venice’s main libraries. This collection includes nearly 3,000 books from the 19th to 21st centuries and over 100 years of scientific publications, all on Venice. They were digitized under the Linked Books project – a project funded by the Swiss National Science Foundation and spearheaded by researchers Giovanni Colavizza and Matteo Romanello working in association with Kaplan. An ad hoc program called Venice Scholar was also developed to automatically extract and search citations, making it easier for historians to use them as they navigate through the city’s history. This is an open system, like the other two, and new entries can be easily added to the database. These secondary sources are also linked to the primary sources in Canvas.

Press Release:
https://go.epfl.ch/VTMFetFlagship

Contacts :
Anne-Muriel Brouet Press Service EPFL
+41 21 69 32442
anne-muriel.brouet@epfl.ch

Sandy Evangelista Press Service EPFL
+41 21 69 32185
+41 79 502 81 06
sandy.evangelista@epfl.ch

Frederic Kaplan Digital Humanities Laboratory EPFL
+41 21 69 30253
+41 21 69 31901
frederic.kaplan@epfl.ch

Isabella di Lenardo Digital Humanities Laboratory EPFL
+41 21 69 33522
isabella.dilenardo@epfl.ch

Julia Noordegraaf Department of Media Studies University of Amsterdam
+31 20 525 72 79
+31 6 172 50 595
J.J.Noordegraaf@uva.nl

Sander Münster Media Center of the TU Dresden
+49-(0)351-463-32530
+49-(0)178-13-525-14
sander.muenster@tu-dresden.de