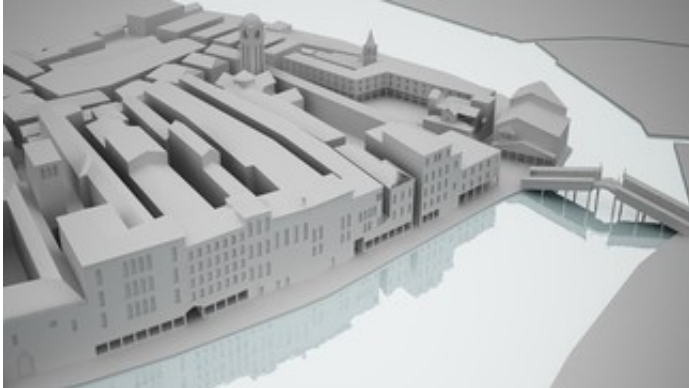# Blueprints finalized for digital archive

June 23 2014

3D renderings of the Rialto. Credit: EPFL / Venice Time Machine

One of the oldest and most complete historical archives in the world is a step closer to being developed into an open digital archive. The Lombard Odier Foundation is joining as a funding partner to take the Venice Time Machine project to the next level.

More than a millennium of cultural heritage will be converted into an open digital archive, which will both preserve it and provide a virtual tool for navigating Venetian history. In a digital humanities project called the 'Venice Time Machine', EPFL researchers are developing the infrastructure and technology needed to convert the enormous amount of administrative manuscripts from Venice's State Archives into a gigantic information system, in collaboration with Ca' Foscari University. Today in Venice, Fondation Lombard Odier is joining the project as a funding partner. It has selected the Venice Time Machine for its exemplary

combination of art, history, and science, and will support the researchers in their effort to develop the project's digitalization and data exploitation workflows.

Since the launch of the Venice Time Machine a year ago as part of a broad partnership in Digital Humanities with Ca' Foscari University and Archivio di Stato, a team of over a hundred researchers and students has put tremendous effort into developing the project's overall infrastructure, sorting and grouping the State Archives by format, and defining the unprecedented workflow for their mass digitization. The project is now supported by an international board of renowned scholars from Stanford, Columbia, Princeton, and Oxford. This board supervises the efforts of the international community to turn the Venice Time Machine into a fundamental resource for studying Venice and its historical mediterranean empire.

THE HERITAGE: 80 km of Interconnected Documents

The State Archives of Venice contain a massive amount of hand-written documentation in languages evolving from medieval times to the 20th century. An estimated 80 km of shelves are filled with over a thousand years of administrative documents, from birth registrations, death certificates and tax statements, all the way to maps and urban planning designs. These documents are often very delicate and are occasionally in a fragile state of conservation.

The diversity, amount and accuracy of the Venetian administrative documents are unique in Western history. By combining this mass of information, it is possible to reconstruct large segments of the city's past: complete biographies, political dynamics, or even the appearance of buildings and entire neighborhoods. "The documents are intricately

interweaved, telling a much richer story when they are cross-referenced", explains Raffaele Santoro, director of the Venice State Archives: "All of the documents here are somehow interconnected".

THE VISION: Accessing the Venice Time Machine from the Cloud

At the moment, historians from around the world have to physically travel to Venice in order to consult a select number of documents in the State Archives. But could these documents be accessed from any computer in the world? Could we search these historical archives using automatically extracted keywords, especially considering that the documents themselves are hand-written? Could we cross-reference architecture plans across the ages and reconstruct Venetian streets and canals as they stood hundreds of years ago?

The Venice Time Machine will give the archives a new, virtual existence on the Web. It will reanimate Venice's past life from them by re-creating social networks and family trees, and visualizing urban development and design. An open digital archive of Venetian treasures will provide an entirely new research environment. "With the Venice Time Machine, we will be able to research topics that were practically impossible before", says historian Dorit Raines at Ca'Foscari University. Working on the Venice Time Machine, she would like to cross-reference and navigate through an estimated 300,000 inheritance wills to reconstruct the ownership of valuable assets throughout the centuries.

THE TECHNOLOGY: Big Data for the Archives

There are multiple technical challenges associated with converting this unique and vast cultural heritage into a digital archive. An optimal workflow must be established for the accurate text recognition of

millions of manuscripts. Mass digitization not only requires the systematic scanning of ancient manuscripts, but also the automatic processing of different hand-writing styles, as well as the analysis of Latin and several other languages as they evolve through time.

Given the immense quantity of data, mass digitization also requires the management, mining and classification of big data. The Venice Time Machine is built for this. For example, algorithms for automated text recognition continuously adapt and improve as data increase and databases of Venetian people and places grow larger.

"The goal is to transform all of these records into a digital information database of the past", says Frédéric Kaplan, an EPFL professor leading the Venice Time Machine. "The good news is that Big Data is not a problem; it is part of the solution."

As part of the "Fondation Lombard Odier – EPFL venture fund" – dedicated to initiate, fund and develop ambitious EPFL strategic initiatives – the Fondation Lombard Odier is proud to join the Venice Time Machine Project as a funding partner. "This initiative resonates with our history, legacy and care for future generations as an institution established in 1796; it breaks new ground in the realm of Big Data and opens tremendous opportunities that are meaningful to our civilization", says Thierry Lombard, President of the foundation. According to EPFL President Patrick Aebischer: "This collaboration began as a shared wish to celebrate past and future memories by capitalizing on the digitalization and exploitation of big data."

Provided by Ecole Polytechnique Federale de Lausanne