Time Machine: Big Data of the Past for the Future of Europe

Deliverable D3.2
TM Operation (Pillar 2) Roadmap

Preamble:

This deliverable is part of a structured set of outputs produced and built upon during the 12-month TM CSA project. As such, the content contained within was further refined, synthesised and improved throughout the project and in particular when combined with material from other deliverables during the production of the full D8.5 TM LSRI Strategy and Implementation Proposal. Please be advised that the most up-to-date version of any information found in this document will be found in D8.5, where it can also be viewed in proper context as part of the entire TM LSRI proposal.

Abstract

A road map is presented for the Time Machine Operation, which is one of the main pillars (Pillar 2) of the Time Machine LSRI. The objective is to define the set of actions that will shape the constituent parts of the Time Machine infrastructure and the principles and processes for managing an ecosystem extending across the EU.

The main ideas that are used to develop the roadmap comprise:

- A governance scheme around a Time Machine Organisation (TMO) that sets out the global rules for the organisation and operation of the Time Machine communities.
- A Time Machine processing infrastructure, composed of a digital content processor and three simulation engines: a 4D simulator, a large-scale inference engine and a universal representation engine.
- Local Time Machine projects in specific geographic locations by partnerships of local stakeholders aiming to develop zones of higher “rebuilding the past activities” density.
- The use of Requests for Comments (TM RFC) to develop the Time Machine infrastructure and operations in an iterative and incremental process. Based on the methodology used by the Internet Society to define international standards, this approach will ensure a smooth consultation of the proposed developments and serve as a basic for the assessment and development phase to come.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 820323

Project Identification

Project Full Title	Time Machine: Big Data of the Past for the Future of Europe
Project Acronym	TM
Grant Agreement	820323
Starting Date	1 March 2019
Duration	12 months

Document Identification

Deliverable Number	D3.2
Deliverable Title	TM Operation (Pillar 2) Roadmap
Work Package	WP3
Delivery Due Date	31 October 2019 (Month 8)
Actual Submission	03 June 2020 (Version 1.3)
Leading Partner	ICARUS
Main Author	Frédéric Kaplan, Daniel Jeller, Camille Besse
Contributions	Allan Vestergaard, Bénédicte Bucher, Elsa Margiun-Hamon, Eva Vodochodska, Karl Heinz, Victor-Jan Vos, Yolène Jahard, Isabella di Lenardo, François Ballaud
Reviewer(s)	Gustavo Riva, Didier Richard, Hervé Déjean, Julia Fallon

History of Changes

Date	Version	Author	Comments
31.10.2019	1.0	Frédéric Kaplan, Daniel Jeller, Camille Besse	Document ready for submission, after integration of comments by internal reviewers.
17.12.2019	1.1	Frédéric Kaplan, Daniel Jeller	Reviewer comments addressed, Annex C added
05.03.2020	1.2	François Ballaud, Frédéric Kaplan	Minor changes implemented (alignment with D8.5). Annex C removed
03.06.2020	1.3	Kevin Baumer	Added preamble to title page which points to D8.5, per final review meeting.

Disclaimer

Definitions

4D Simulator	One of 3 TM Simulation Engines. The 4D Simulator manages a continuous spatiotemporal simulation of all possible pasts and futures that are compatible with the data. the 4D Simulator includes a multiscale hierarchical architecture for dividing space and time into discrete volumes with a unique identifier: a simulation engine for producing new datasets based on the information stored. Each possible spatiotemporal multiscale simulation corresponds to a **multidimensional representation** in the 4D computing infrastructure. When a sufficient spatiotemporal density of data is reached, it can produce a 3D representation of the place at a chosen moment in European history. In navigating the representation space, one can also navigate in alternative past and future **simulations**. **Uncertainty and incoherence are managed at each stage of the process and directly associated with the corresponding reconstructions of the past and the future.**
Big Data of the Past	A huge distributed digital information system mapping the social, cultural and geographical evolution. A key objective of Time Machine is that such system brings together dense, interoperable, standardised (linked data, preferably open) and localised (marked up with spatial-temporal information) social, cultural and geographical heritage resources.
Communities	Group of users, self-organised by territorial or transversal interests, offering various voluntary works and favours (annotation, digitisation, bibliographic research, development), according to the standards in place, to the partners. Those communities can elect a representative.
Digital Content Processor	Automatic process extracting information from documents (images, video, sound, etc.). Digital Content Processor of Level 1 just label mentions of entities. Digital Content Processor of Level 2 label relations between entities. Digital Content Processor of Level 3 label Rules. Each processing is fully traceable and reversible. The results of the processing constitute the core dataset of the Big Data of the Past and are integrated in the TM Data Graph.
Large-Scale Inference engine	One of 3 TM Simulation Engines. The Large-Scale Inference Engine is capable of inferring the consequences of chaining any information in the database. This permits to induce new logical consequences of existing data. The Large-Scale Inference Engine is used to shape and to assess the coherence of the 4D simulations based on human-understandable concepts and constraints. Its origin comes from more traditional logic-based AI technology, slightly overlooked since the recent success of the deep learning architecture, that can, nevertheless, play a key role in an initiative like TM.
Local Time Machine	Zone of higher "*rebuilding the past activities*" density. Constituted of a group of local partners and communities bound by a common territorial focus and a declaration of intent, which respect both graphical and values charters. Any institution who meets eligible criteria can integrate a Local Time Machine. The declaration of intent is reviewed on an annually basis (time for new partners to integrate the TM)
Project with Time Machine label (PWTML)	Project respecting the technical charter, whose tasks are documented - modelled within the Time Machine graph. All the partners of a PWTML must have signed the declaration of intent of the related Local Time Machine.
Technical Charter	Should contain information about infrastructure standards required within any project with Time Machine label. The Technical Charter defines the Time Machines Rules, Recommendations, Metrics and Official software. The document is revised periodically.

Time Machine Box	Servers that allow partners to store their documents and metadata and integrate easily the Time Machine Network and be appropriately documented in the Time Machine Graph. The Time Machine Box is part of the Time Machine Official Components.
Time Machine Data Graph	Formal representation of knowledge extracted by human or automatic process, represented with semantic web technology
Time Machine Index	The TM index is a global system indexing different type of objects: e.g. documents; iconography; 3D geometries. It gathers all information regarding documents and their contents. Could be used as a basis for other search engine infrastructures (allows backups).
Time Machine Infrastructure Alliance	Coalition of TM's partners regrouping in-kind donators for infrastructure components (server's space and computing power).
Time Machine Mirror World	One of the API of the Time Machine using the processing of the 3 TM Simulation Engines to produce a continuous representation model that can be accessed as information stratum overlaying the real world.
Time Machine Network	Set of all the partners *actually* interacting in the Time Machine. Each member of the Time Machine Network must have signed the Value and Technical Charter
Time Machine Official Components	Pieces of software (e.g. Time Machine Box) that help partners conforming to the Time Machine rules as they are directly embedded in the software.
Time Machine Operation Graph	Formal representation of the past, on-going and future operations of the partners in the Time Machine Network and the data pipelines.
Time Machine Organisation	Association regrouping the Time Machine Partners. Some maybe active and other not. Not all may have signed the Values and Technical Charters.
Time Machine Recommendations	Recommendation on technology which are not obligatory at this stage for the development of the Time Machine (e.g. choice of a particular IIIF image server).
Time Machine Request for Comments	Main document for the progressive design of the Time Machine infrastructures, standards, recommendations and rules, inspired by the process used for 50 years for the development of Internet Technology, today administrated by the Internet Engineering Task Force (IETF) as part of Internet Society (ISOC).
Time Machine Rules	Standard and rules that need to be followed to be acceptable in the Time Machine Network and become a Time Machine operators. Any entity not following these rules are out.
Time Machine Standard Contracts	Set of standard contracts to facilitate the interaction between Time Machine partners.
Time Machine Standard Metrics	Measures helping partners of the Time Machine Network coordinate with one another to compare performance (for quotes of services, but not only, there are also use for research performances, etc.).
Time Machine Super Computing Architecture and Simulation Engines	TM Super Computing Architecture composed of distributed computing resources from the TM Network provided by the TM Infrastructure Alliance. On this distributed architecture, different typologies of computing process can run. For instance, Digital Content Processors are intrinsically easier to run in parallel, whereas Simulation engines, which allow users to generate possible pasts and futures from the TM Data Graph need for more specific computing architecture.

Universal Representation Engine	One of 3 TM Simulation Engines. The Universal Representation Engine manages a multidimensional representation space resulting from the integration of the pattern of extremely diverse types of digital cultural artefacts (text, images, videos, 3D), and permitting new types of data generation based on transmodal pattern understanding. In such a space, the surface structure of any complex cultural artefact, landscape or situation is seen as a point in a multidimensional vector space. On this basis, it could generate a statue or a building, produce a piece of music or a painting, based only on its description, geographical origins and age.
Values Charter	Conform to the principle of openness in EU law

List of abbreviations

AI	Artificial Intelligence
CH	Cultural Heritage
GLAM	Galleries, Libraries, Archives, Museums
LTM	Local Time Machine
PWTML	Project with Time Machine Label
RFC	Request for Comments
SSH	Social Sciences and Humanities
TM	Time Machine
TMO	Time Machine Organisation

Table of Contents

1 Introduction

Time Machine (TM) is a Large-Scale Research Initiative (LRSI), pushing the frontiers of scientific research in Information and Communication Technologies (ICT), Artificial Intelligence (AI) and the Social Sciences and Humanities (SSH).

TM is built around the vision to develop the Big Data of the Past, a huge distributed digital information system mapping the European social, cultural and geographical evolution. This large-scale digitisation and computing infrastructure will enable Europe to turn its long history, as well as its multilingualism and multiculturalism, into a living social and economic resource for co-creating a common future. The proposed LRSI will use space and time as shared references across domains, disciplines and cultures, to understand and give value to constructions, artefacts, observations and data produced over centuries, enabling Europeans to better appropriate their heritage and strengthen the feeling of European belonging.

This document is the formal deliverable D3.2 presenting the roadmap for Pillar 2. Following this short introduction, the deliverable is organised in the following sections:

- Section 2 starts with an overview of the TM LSRI and then discusses the main aspects for the design of the TM operation, including the key concepts used and the links with the other TM pillars.
- Section 3 focuses on the research and innovation plans for Pillar 2, presenting the state of the art, the targeted achievements and the methodologies to obtain them.
- Section 4 discusses the funding resources that can support the Pillar 2 actions.
- Section 5 presents the stakeholders to be involved in and/or that are directly concerned by these actions.
- Section 6 examines the framework conditions that relate to the implementation of Pillar 2.
- Section 7 reviews the risks and barriers related to Pillar 2 and the mitigation strategies that are foreseen to address them.

2 Design of Pillar 2 – Time Machine Operation

2.1 Overview of the Time Machine LSRI

Rational

Over the centuries, the national, regional and local identities of Europe have evolved in relation to one another, through large swathes of transnational mobility and through dense exchanges that have shaped European languages, traditions, arts and many other aspects of human activity. These processes have largely contributed to the creation of a European culture characterised by diverse historical memories, which have laid the foundations to values and ideas harmonised by pluralistic and democratic dialogue.

To-date, however, increased globalisation, changing demographics and their threat against the idea of a shared past, as well as the resurgence of unresolved conflicts deep-seated in European memory are key drivers of a 'localisation backlash' that places local and personal interests above any other. These growing trends present a clear threat to the cohesiveness of European cultural identity and sense of belonging.

Pluralistic and democratic dialogue in Europe has traditionally been facilitated by important intermediaries, such as cultural media and institutions acting as cornerstones of our shared values, principles and memories. Today, the dialogue between different actors and the historical visions they embody is complicated by the rise of private digital platforms that have created a new space of opinion-leadership, as well as new forms of political expression and participation.

Managed by proprietary algorithms, such platforms may prioritise popularity and personal agendas over historical and cultural data, opening the way to fake news. In the resulting crisis of authority that affects journalism, academia and politics, many people do not trust anymore the information received from these institutions.

These unprecedented transformations create a vital need for Europe to restore and intensify its engagement with its past as a means of facilitating an evidence-based dialogue between diverse historical memories, their values and mutual interdependencies and building a common path across generations.

Time Machine responds to this need by building the required infrastructure, and an operational environment for developing the "Big Data of the Past" that will transform and enhance the role of history and culture across Europe, opening the way for scientific and technological progress to become a powerful ally to safeguarding European identity and democratic values.

For Time Machine, digitisation is only the first step of a long series of extraction processes, including document segmentation and understanding, alignment of named entities and simulation of hypothetical spatiotemporal 4D reconstructions. The hypothesis pursed by Time Machine is that such computational models with an extended temporal horizon are key resources for developing new approaches to policy making and to offering services to European citizens and consumers.

Still, there is one more crucial reason supporting the cause of Time Machine. After the creation of the web that digitised information and knowledge and the social media that digitised people and characteristics of human behaviour, a third technology platform is being created, digitising all other aspects of our world, giving birth to a digital information "overlay" over the physical world, a "mirror-world"[1]. The mirror-world will aim to be an up-to-date model of the world as it is, as it was and as it

[1] The term was first coined by Yale computer scientist David Gelernter in 1991 in its book "Mirror Worlds: Or the Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean" (Oxford University Press, 1991)

will be. All objects (including representations of landscapes) of the mirror-world will be machine-readable, and, therefore, searchable, traceable and subject to be part of simulations by powerful algorithms. In the mirror world, time will be a fourth dimension, as it will be very easy to go back to the past, at any location, reverting to a previous version kept in the log. One may also travel in the other direction, as future versions of a place can be artificially created based on all information that can be anticipated about the predictable future. Such time-trips will have an increased sense of reality, as they will be based on a full-scale representation of the present world. Time Machine is today the most advanced concrete proposal to build the first version of a European mirror-world.

Like the other two platforms, the mirror-world will disrupt most forms of human activity, as we know them today, giving birth to an unimaginable number of new ideas (and many problems) and creating new forms of prosperity from new forms of economic and social activity that will shape new behaviours and ecosystems. In this scenario that is currently unfolding, Time Machine will enable Europe to be one of the leading players, shaping the mirror-world according to its democratic values and fundamental ethics (open standards, interoperability). With Time Machine, while it will have a powerful tool to strengthen its cohesion and sense of belonging, Europe has, moreover, an opportunity to impose its own terms against the multinational technology giants that will fight for dominating this new technology platform, just as those who now govern the first two platforms have done in the past.

Expected impact

- A strong boost in EU competitiveness in AI and ICT:
  - An AI trained on Big Data of the Past will offer a strong competitive advantage for Europe in the global AI race.
  - Disruptive technologies in machine vision, linguistic and knowledge systems, multimodal (4D) simulation, HPC and long-term data storage will strengthen the competitive position of EU industry in these fields.

- New disruptive business models in key economic sectors:
  - Cultural Heritage is a unique asset for European businesses. Time Machine will act as an economic motor for new services and products, impacting key sectors of European economy (ICT, creative industries and tourism).
  - Time Machine will develop a paradigm to follow for cities that wish to make a creative use of their historical past.

- A transformational impact on Social Sciences and Humanities (SSH):
  - With Time Machine, SSH will evolve to address bigger issues, allowing new interpretative models that can smoothly transition between the micro-analysis of single artefacts and the large-scale complex networks of European history and culture.

- Moreover, Time Machine will:
  - Be a driver of open science, as well as open (public) access to public resources.
  -