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

Deliverable D3.1
TM Operation (Pillar 2) Roadmap - draft

Abstract

A draft 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	D3.1
Deliverable Title	TM Operation (Pillar 2) Roadmap - draft
Work Package	WP3 - Roadmap for TM Operation
Delivery Due Date	30 June 2019 (Month 4)
Actual Submission	1 July 2019
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
Reviewer(s)	Gustavo Riva, Didier Richard, Hervé Déjean, Julia Fallon

History of Changes

Date	Version	Author	Comments
28.06.19	1.0	Frédéric Kaplan, Daniel Jeller, Camille Besse	Document ready for submission, after integration of comments by internal reviewers.

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.

The key objective of the TM CSA project is to develop a full LSRI proposal around this TM vision. Detailed roadmaps will be prepared, organised around four pillars, namely science and technology, TM operation, exploitation avenues and framework conditions. The roadmap development methodology foresees the elaboration of draft roadmaps for each pillar by working groups composed of Consortium experts, followed by a round of consultations with relevant external stakeholders. These consultations will enable the Consortium to finalise the pillar roadmaps in a way that reflects the needs and expectations of a pan-European ecosystem that has been built around Time Machine and is currently expanding at fast rate.

The roadmap for the TM operation pillar is developed in WP2. This document is the formal deliverable D3.1 presenting the draft roadmap for Pillar 2. The emphasis is on describing the qualitative aspects of the proposed research and innovation actions in a sufficient level of detail, enabling informed feedback to be received during the consultations that will follow. The final roadmap is planned for Month 8 (October 2019).

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.
  - Provide a constant flux of knowledge that will have a profound effect on education, encouraging reflection on long trends and sharpening critical thinking.
  - Render education for Europeans more accessible, interactive and diversified.
  - Develop new or updated legislation or guidelines in the field of AI, including ethical norms and ethical standards in areas such as access to and re-use of digital data, harmonised rules on data-sharing arrangements, especially in business-to-business and business-to-government situations, as well as clarified concepts in data ownership.

- Create new jobs for digital and traditional humanists and social scientists, while offering clear opportunities for talented humanities graduates with increased digital skills, by demonstrating the benefits of the new profession “Digital Humanities expert”.

- Having confirmed itself as one of the pioneers, Europe will make meaningful contributions to the foundation and use of the mirror-world, in line with its values and ethics.

LSRI Structure

The Time Machine LSRI is articulated around four pillars, each defining a specific objective of the initiative:

- Pillar 1 – Science and Technology for the Big data of the Past: Addressing the scientific and technological challenges in AI, Robotics and ICT for social interaction, for developing the Big Data of the Past, while boosting these key enabling technologies in Europe.
- Pillar 2 – Time Machine Operation: Building the TM infrastructure for digitisation, processing and simulation, in order to develop a sustainable management and operational model (“TM franchise”), as well as to create the basis for and engagement with the TM communities participating in the development and use of Time Machine.
- Pillar 3 – Exploitation Avenues: Creating innovation platforms in promising application areas, by bringing together developers and users for the exploitation of scientific and technological achievements, and therefore leveraging the cultural, societal and economic impact of Time Machine.
- Pillar 4 – Outreach and innovation: Developing favourable framework conditions for the outreach to all critical target groups, and for guiding and facilitating the uptake of research results produced in the course of the LRSI.

Each pillar comprises thematic areas, as shown in Figure 2-1.

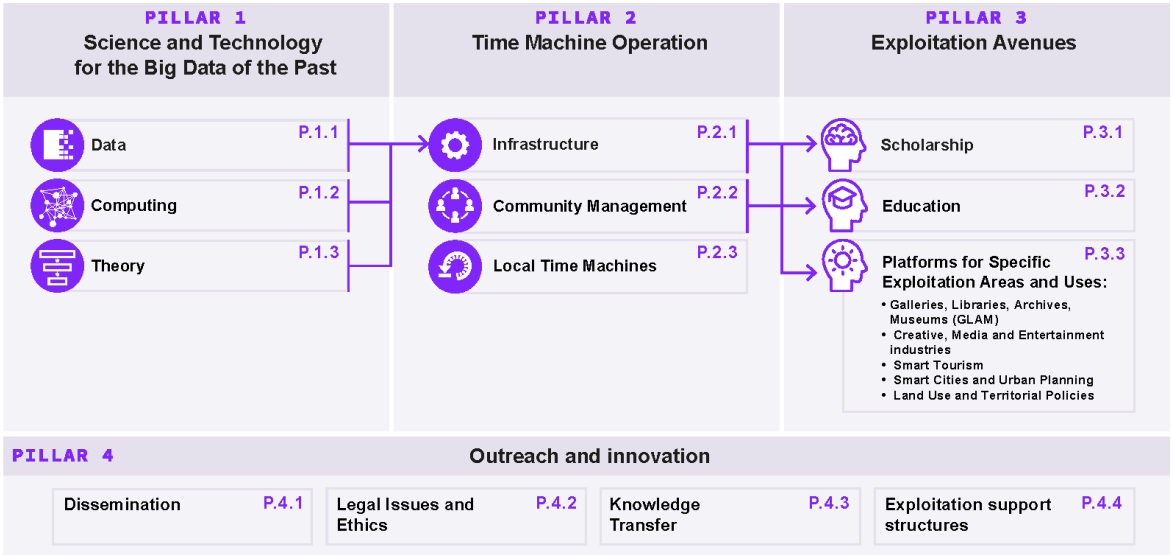

Figure 2-1: Time Machine Pillars & Thematic Areas and their interrelations

2.2 Pillar 2 key concepts

Pillar 2 aims to put in place the constituent parts of the Time Machine infrastructure and the management principles and processes for an ecosystem of Time Machine contributors and users extending across the EU. The roadmap is developed using the main ideas presented below.

Requests for Comments

Reaching consensus on the technology options to follow in a programme as large as Time Machine is a complex issue. To ensure the open development and evaluation of work, a process inspired by the Request for Comments (RFC) publication mechanism for negotiating the standards and protocols used by the development of Internet protocol will be adapted to the needs of Time Machine (https://en.wikipedia.org/wiki/Request_for_Comments).

Time Machine Request for Comments will be a freely accessible publication, identified with a unique ID. Like for the development of the Internet, the Request for Comments aims to be the main process for establishing rules, recommendations, core architectural choices for the Time Machine Infrastructure. The pillar 2 road map specifies a first (non-limitative) list of the TM RFCs to be prepared.

The development of RFCs will be managed by organisational structures that will be elaborated in WP6 – Governance Scheme.

Time Machine Organisation

The whole governance of Tim Machine is based on a **Time Machine Organisation (TMO)** that sets the global rules for all actions and operations related to Time Machine, including definition of processes, labelling system and recognised infrastructure.

The organisational scheme and details of the TMO governance will be specified in the WP dedicated to governance (WP6), based on requirements formulated in each pillar. The pillar 2 road-map already indicates a series of important functions, like the Time Machine Infrastructure Alliance formed by the institutions sharing in-kind resources for storage and computing and the RFC Editorial board, responsible for the editing process of the TM Requests for Comments mentioned above.

Time Machine processing infrastructure

The TM digitisation infrastructure will be composed of a network of digitisation hubs and will be organised on a European scale. A peer-to-peer platform will be in charge of managing and optimising digitisation strategies at European level, and will also be tasked with the development of generic solutions for archiving, directly documenting the digitisation processes, and swiftly putting the digitised documents online. The hubs will cover regional digitisation needs with standardised hardware for digitisation, storage, information exchanges and on-demand scanning, based on results of Pillar 1 and existing metadata standards, like the one developed by Europeana.

The peer-to-peer platform will federate system integrators at European level, facilitating the deployment of this equipment. The effort will build upon existing EU Research Infrastructures (DARIAH, CLARIN) and infrastructures providing access to CH (Europeana, Archive Portal Europe, etc.). TM will introduce new processing pipelines for transforming and integrating CH data in such infrastructures.

A typology of digitisation interventions will be established, separating collections that can be moved and processed in digital hubs (large, non-fragile collections), collections or objects that need local intervention (e.g. very fragile document, statues, buildings), process that can be performed by volunteers using mobile technology (e.g. scanning campaign across cities, on-the-fly digitisation in reading rooms), processes that can be performed using robots and drones, etc.

TM will establish both top-down digitisation pipelines optimised for efficiency and bottom-up on-demand digitisation services driven by scholarship questions or the establishment of particular

projects. X-ray, multispectral and very high resolution scanning will initially be used for on-demand test cases (e.g. analysis of a drawing under the surface of a painting). They may integrate the massive pipelines if their costs become affordable and their relevance demonstrated. In more general terms, the overall infrastructure will be designed to progressively incorporate the progress of imaging technology and be able to deal with the coexistence of documents with increasing levels of resolution and information density.

Documents are digitised using different kinds of acquisition machines and are treated separately depending on their nature (textual and audio-visual documents, iconographic elements, maps, 3D objects and environments). Information is extracted progressively, either manually or automatically, to produce elementary historical units, connected with one another. This progressive decomposition and refinement need to be seen not as a mere automatic process but as a collective negotiation. Each intervention, either algorithmic or human, is fully traceable and reversible. The results of the processing constitute the core dataset of the Big Data of the Past.

The infrastructure is shown in Figure 2-2. 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:

- The 4D Simulator manages a continuous spatiotemporal simulation of all possible pasts and futures that are compatible with the data. Therefore, 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.

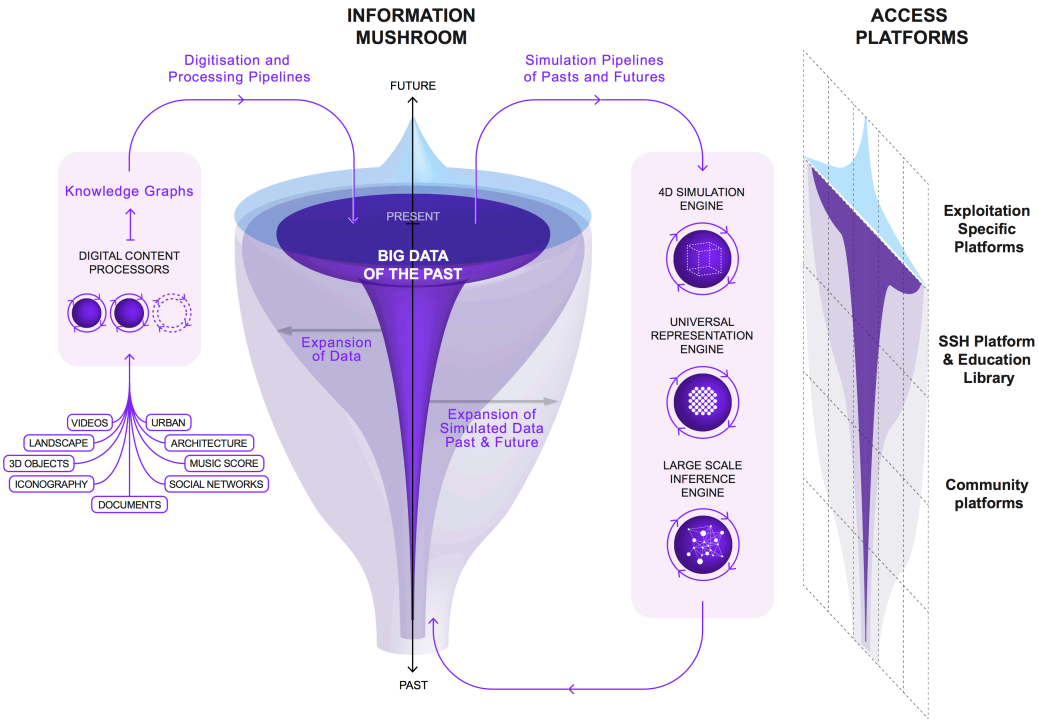

Figure 2-2: Schematic representation of the TM Digital Content Processor and the three simulation engines

- 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. Such an engine is the conceptual extension of current trends in machine learning, targeting new abstraction levels in natural language understanding (e.g. multilingual representation space used for translation).
- 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.

All functions of the different components can be deployed through a fully distributed solution using a storage and computation architecture aimed at an integrated, long-term and sustainable storage of the processed content. This solution embodies our strategy for the long-term availability of processed content, even beyond the lifetime of the organisations hosting it, through predefined and legally binding agreements on licensing, redundant storage, automatic hand-over policies and long-term self-supporting investment initiatives to indefinitely extend the availability of the digitised content of TM. The solution is highly redundant for a chosen set of documents and data, includes full transparency of all operations performed, and is fully version-aware.

Local Time Machines

The Time Machine Network is organised as an unlimited amount of **Local Time Machines** (LTMs). Each LTM is anchored in the space of a city, a region, around which various partnerships can form, aiming to transform it into a zone of higher *"rebuilding the past activities*" density. The Local Time Machines follow the rules of the Time Machine Organisation, elaborated as a series of RFCs. The TMO provides help in their launch and growth, which resulted in the identification of several *offices*.

In the course of time, Local Time Machine pass through different maturity phases (preparatory phase, submission phase, operation phase, level of operations etc.). Each maturity phase permits to envision specific exploitation strategies. For instance, only extremely dense Local Time Machine can launch Mirror World interfaces.

Projects with Time Machine Label (PWTML)

PWTML are conducted within a Local Time Machine. In order to increase the activities density of the zone, and once funding is secured by the partners with, when needed, the help of the TMO, partners of a Local Time Machine can decide to gather around a common goal and create and finance a new PWTML. There are no obligations for partners of a LTM to be involved in each PWTML, they're all free to gather around specific and mutual affinities and be part of the PWTML (s) they wished for.

A PWTML is composed of various *bricks,* such as:

- **Redocumentation brick**: when a stabilised project wishes to enter the TM network, several actions might be requested to fulfil the minimum requirements.
- **Digitisation brick**: when digitisation and 3D processes are requested within a project.
- **Community brick**: when the project engages a community around a specific task.

- **Research brick**: when the project is meant to answer a specific research subject, involving scholars on a higher level.
- **Valorisation brick**: when a project aims to valorise or expose works already conducted within a particular LTM.

Interaction with TMO

The TMO will supervise the following services leading to the growth of a LTM:

LTM/Framework	• Supervision, project tracking office • Know-How, Guidance office
LTM/Labelling System	• Supervision, certification office
LTM/Legal setting	• Legal office
LTM/Financial system	• Finance, Economic intelligence and Watch services office
LTM/Incorporation, Collaboration and enhancement	• Solidarity and Collaboration office • Research Data Management office
LTM/Smart Cluster	• Smart Collaboration

The eligible partners and external communities.

As shown in Figure 2-3, Local Time Machine, various typologies of partners are central players for the development of both the Time Machine Organisation and the Local Time Machines. We can however distinguish two players' typologies, the ones actual partners within the TM network (meaning they did ratify and comply to TM's rules and recommendations and a specific franchise system), and the external communities that could be engaged with voluntary actions (in the context of a Local Time Machine or identified projects at the TMO level), or provide guidance and advices on specific topics (e.g. professionals communities or cultural-heritage networks).

Some already identified players are here shortly presented for contextual understandings, along with their specific tasks within the TM network. One of the duties of the Local Time Machines TM RFC on Local Time Machine Framework's is to further develop this list of eligible partners and provide means to ensure those partner's involvement. As several of those players can also be considered as exploitation avenues, the relevant outputs can be linked with the ones from pillar 3.

a. *The eligible partners*

Patrimonial Institutions This group is mostly composed of GLAM institutions (Galleries, Libraries, Archives, Museums). As keeper-owners of data or occupants of places of interest, they are bringing in a PWTML or as an LTM partners, either materials to be digitised or digitised collections. GLAMs are also constituting one of the identified exploitation avenues.

Operators This group is composed of actors involved within a digitisation or 3D process. Once the digitisation hub is fully in place (it has been tested and approved by the users and meets all TMO quality criteria), they can be considered as official TMO scanning, logistic or development operators. Operators mostly collaborate with Patrimonial and Research Institutions.

Research Institutions This group is principally composed of private or public units (such as universities, research centres related to relevant technical or humanities fields …). It mostly collaborates with the Operators (regarding technological research) group but could also be a direct user of the data exposed in the TM Data and Operation Graphs, used as basics for their scientific research. On the one hand, papers and raw data they create, can, once transformed and their content information properly extracted, contribute to the growth of the Big Data of the Past Graph and on the other hand, their research can be fed thanks to the data of the Big Data of the Past Graph. This group is therefore also part of the identified exploitation avenues as scholarship. As technological centres, they can directly help for the relevant infrastructure needs of an LTM and help to improve or create new processes.

Funding institutions This group is composed of Research institutions, Cultural institutions, Territorial players, Industrial partners, Civil society organisation, Associated programmes, philanthropist, private sponsors etc. They do, in their specific ways, support funding either to the TMO or to an LTM or PWTML. Those funding can happen once, be made on an annual frequency or follow any regular frequency. Future dedicated tasks will work for the development of such financial partnerships (*PILLAR 4, task 4.4: co-financing models with private money, WP6, task 6.3: money coordination and management structures and WP7 task 7.5: support from public money*).

Local authorities A Local Time Machine is anchored within a specific region; therefore, local authorities and policy makers will be invited to play an active role in the creation and definition of the shape of the Local Time Machine. Both Local Time Machine and local authorities share common interests and should work closely to ensure that Local Time Machine are adapted to the local specific features. These diverse forms of collaboration and associated shared interests are meant to be strengthened over time and should set the LTMs as strategic partners in development initiatives. Local Authorities can draw real benefits from the Big Data of the Past, with regards to smart tourism, smart cities, land use and urban planning. In this context, the LTMs can become trustworthy partners for the sustainable development of a region.

a. External communities

This group is composed of individuals or networks - associations wishing to voluntarily engage within the context of either the TMO or an LTM. They may form several sub-communities: the developers, the educators, the professionals, the scholars, and the more *generic mass* of volunteers. Three targeted achievements will be detailed within the present roadmap, aiming to design suitable infrastructures, dedicated communication means (in close relation with the outputs of *WP7, task 7.4: Development of TM citizens*) and corresponding operations monitoring: 1. Community Interface, 2. Community Inclusion, 3. Community statistics. External communities are directly related to the impact O.1 and O.5 (see section 2.4 below).

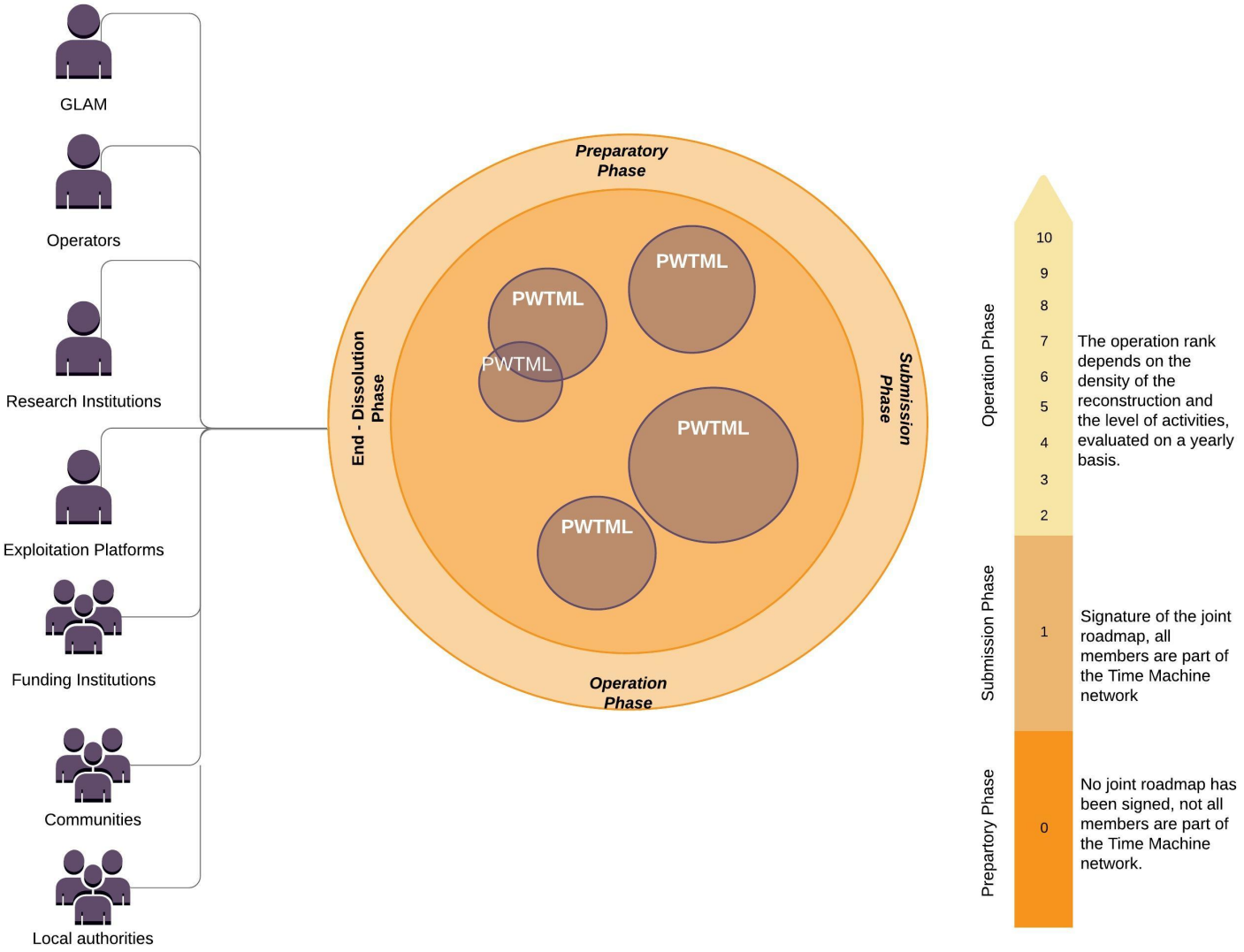

Figure 2-3: Local Time Machine

2.3 Pillar 2 scope of work

Pillar 2 has been divided into three thematic areas, Infrastructure (P.2.1), Community Management (P.2.2) and Local Time Machines (P.2.3) that are further discussed below.

Infrastructure

This thematic area focuses on the design of:

- A digitisation-hubs network that will facilitate and standardise digitisation processes at European scale
- A distributed super computing infrastructure for processing the Big Data of the Past
- A distributed storage system.

The core concepts developed in the Time Machine CSA will be collectively extended and interconnected using RFCs as indicated below.

For Digitalisation Hubs:

- TM RFC on definition of typologies
- TM RFC on standardisation and homologation
- TM RFC on Open Hardware
- TM RFC on synergy and interaction in EU Research Infrastructure
- TM RFC on on-demand digitisation
- TM RFC on global optimisation of digitisation process

For Processing and Simulation

- TM RFC on General Standards for the Super Computing Architecture
- TM RFC on Digital Content Processor (DCP)
- TM RFC on TM Data Graph
- TM RFC on Large-Scale Inference Engine
- TM RFC on 4D Simulator
- TM RFC on Universal Representation Engine

For Distributed Storage

- TM RFC on Distributed storage system for Public Data
- TM RFC on Distributed storage system for Private Data
- TM RFC on Content Filtering

Community Management

The massive involvement of scholars, developers, CH professionals, service providers and citizens is a key success criterion for the Time Machine initiative. The main idea is that a system of platforms will connect Time Machine with such external communities that will benefit from and can provide input in various forms to the Time Machine. The aim of the thematic areas is, therefore, to build a strategy for Community management and an associated Community Management System, responding to well identified requirements for staff and processes leading to mutually beneficial and sustainable interactions with the TM communities.

As mentioned, the TM mutualised platforms, including those corresponding to exploitation avenues referred to in Pillar 3, will enable the organisation of the following community dynamics:

- **Developers:** the community of developers, private and public sector institutions or volunteers, working on each open source component of TM infrastructure will organise and document each conceptual and technical choice made during the evolving design of the infrastructure.

- **Educators:** educators, both on the school and university level will be able to rely on the Big Data of the Past as a source for their daily work as well as a platform to share common materials and workflows in developing more efficient courses and lectures.
- **Professionals:** the community of professionals consisting of people that work in archives, museums, libraries, documentation centres and so forth will be responsible for curating the general catalogue of material objects (primary, secondary sources, CH objects) and documenting the history of their movement in time and space.
- **Scholars:** scholars conducting research using the Big Data of the Past and simulation technologies will have dedicated community networks. Their debate can be followed live and will allow room for competing views.
- **Volunteers:** a continuously evolving group of users who participate in the transcription and interpretation of the machine extracted information and actively engage with the Big Data of the Past by contributing knowledge, expertise and their personal heritage, both online and at offline events to strengthen personal, inter-generational as well as communal ties at local, regional, national and/or transnational levels.

Local Time Machines

The LTMs are meant to be at the centre of the TM network, rallying different partners behind the TM objectives and helping the growth of both TM Data and Operation Graph. Shaped as independent networks of projects, they're still active within a common global TM network, and as such benefit from the TM infrastructure and contribute to its development. Finding the proper balance between TM network governance and the LTM activities is an important factor in the TM development.

As one of the pillar 2 objectives is to provide as soon as possible a starter-kit for minimum LTM functionality, draft versions of supporting documents, clarifying minimal technical, legal and values requirements, have been created and presented in Annexes A, B and C. The TM Requests for Comments method is also used to develop the following outputs:

For LTM/TM's rules and recommendations
- TM RFC on data lifecycle
- TM RFC on Vision Mission and Values Charter
- TM RFC on Technical Charter

For LTM/Framework
- TM RFC on LTM/Framework
- TM RFC on Training

For LTM/Labelling System
- TM RFC on Value Scale

For LTM/Legal setting
- TM RFC on Intellectual property rights and licences

For LTM/Financial system
- TM RFC on Franchise

For LTM/Incorporation, collaboration and enhancement
- TM RFC on Enhancing collaboration
- TM RFC on Knowledge Transfer
- TM RFC on Solidarity
- TM RFC on Top-down initiatives

For LTM/Smart Cluster
- TM RFC on Smart Cluster

2.4 Interactions of pillar 2 with pillars 1 and 3

Pillar 2 plays an intermediary role between pillar 1 and pillar 3 as shown in Figure 2-1. Of particular importance are the impact-facilitating objectives identified in the Pillar 3 roadmap, listed below.

O1: *Cheap Digitisation*: Enable the provision of cheap and cost-efficient solutions for the further digitisation of resources through standardized offers and services and easily replicable open hardware technologies

O2: *Generic Automation*: Enable the automation of the mark-up of these resources tagging concepts, named-entities, relations and rules.

O3: *Connection*: Facilitate the intelligent connection of existing fragmented data resources using, adopting and building on existing legal frameworks and developing standards for distributed storage solution.

O4: ***Simulation*:** Transform sparse data into continuous 4D representations cable of representing multiworlds.

O5: ***Experience***: Enable new paradigms for the restitution of the data to the end-user including spatio-temporal search engines, geo-historical services and Mirror Worlds.

The *Infrastructure* thematic area specifically describes how to develop the underlying hardware and computing structure that will support these objectives. It is built on research objectives identified within pillar 1, in the shape of specific documents called *TM Request for Comments,* that will serve to set the basis for the operation and infrastructure decision of pillar 2.

The *Community Management* thematic area focuses on developing a system connecting external communities to the Time Machine it contributes also to the development of pillar 3 impact objectives (in particular O1 and O5).

The *Local Time Machine* thematic area is related to the governance scheme of the Local Time Machine and by extension to the overall governance scheme of the TMO. It constitutes the core of the Time Machine sustainability model and ensure the growth of the entire project.

The successful design of the Time Machine Data Graph is underlying all the Time Machines objectives (O1 – O5). Its mains characteristics are as follow:

a. It is sparse, as high density is only obtained where a Local Time Machine is active.
b. It contains some Bright Data (curated by a collective of humans) and Dark data (compiled and analysed by machines). Each data types need specific processes in terms of alignment and connections (O3), Simulation (O4) and Experience (O5).
c. It contains some Public and Private Data. Each data types needs specific processes in terms of digitisation (O1), Connection (O2), Simulation (O4) and Experience (O5).

The pillar 2 roadmap directly links the *Impact Objectives* of pillar 3 with a series of RFCs. The table below gives first mapping between some RFCs developed in the following PILLAR and their corresponding objectives in pillar 3. The list will be improved during the consultation phase of the roadmap.

PILLAR 2 tasks	Outputs - TM RFC	PILLAR 3 impacts
2.1 Infrastructure	Definition of Typologies	O1 – Cheap digitisation
	Standardisation and Homologation	
	Open Hardware	

PILLAR 2 tasks	Outputs - TM RFC	PILLAR 3 impacts
	Synergy and Interaction in EU Research Infrastructure	
	On-demand digitisation	
	Time Machine Box	
	Global optimisation of digitisation process	
	General Standards for the Super Computing Architecture	O2 – Generic Automation O4 – Simulation
	Digital Content Processor	O2 – Generic Automation
	Large-Scale Inference Engine	O4 – Simulation
	4D Simulator	O4 – Simulation
	Universal Representation Engine	O4 – Simulation
	Data Graph	O2 – Generic Automation O3 – Connection O4 – Simulation
	Virtual/Augmented Reality and Discovery	O5 - Experience
	Distributed Storage system for Public Data	O3 – Connection
	Distributed Storage system for Private Data	O3 – Connection
	Content Filtering	O3 – Connection
2.2 Community Management	Community Interface	O1 – Cheap digitisation O5 - Experience
	Community Inclusion	
	Community statistics	
2.3 Local Time Machine	Data Lifecycle	O1 – Cheap Digitisation
	Vision Mission and Values Charter	O2 – Generic Automation
	Technical Charter	
	Local Time Machine's Framework	O3 – Connection
	Training	O5 - Experience
	Value Scale	
	Smart Cluster	
	Intellectual Property Rights and Licenses	O3 – Connection
	Franchise	
	Enhancing Collaboration	
	Knowledge Transfer	
	Solidarity	
	Top-Down Initiatives	

Pillar 1 contributes to Pillar 2 through the establishment of key infrastructure RFCs. The *Infrastructure* Thematic Area is based on the delivery of breakthrough technologies, research will play a decisive role in the building of the TM network. The *Local Time Machine* Thematic Area will contribute to shape sustaining operation systems. Several TM RFCs will be written based on the already identified taxonomies domains stated in Pillar 1 roadmap. The table below specifies these links.

<table>
<tr><th colspan="2">Pillar 1
Taxonomy</th><th>Pillar 2 Thematic Area Infrastructure</th><th>Pillar 2 Thematic Area Local Time Machine</th></tr>
<tr><td rowspan="3">Data</td><td>Data modelling</td><td rowspan="2">TM RFC on Definition of typologies
TM RFC on standardisation and homologation
TM RFC on Time Machine Box
TM RFC on Synergy and interaction in EU Research Infrastructure</td><td rowspan="3">TM RFC on data lifecycle
TM RFC on Technical Charter
TM RFC on Knowledge Transfer</td></tr>
<tr><td>Data acquisition</td></tr>
<tr><td rowspan="2">Long Term Preservation</td><td>TM RFC on Distributed storage system for Public Data</td></tr>
<tr><td></td><td>TM RFC on Distributed storage system for Private Data</td><td></td></tr>
<tr><td rowspan="6">Computing & AI</td><td>Computer Vision and Pattern Recognition</td><td rowspan="3">TM RFC on Digital Content Processor (DCP)
TM RFC on TM Data Graph</td><td></td></tr>
<tr><td>Natural Language Processing</td><td></td></tr>
<tr><td>Machine Learning and Artificial Intelligence</td><td></td></tr>
<tr><td>Human-Computer Interaction and Visualization</td><td>TM RFC on Universal Representation Engine
TM RFC on Virtual/Augmented Reality and Discovery</td><td></td></tr>
<tr><td>Computer Graphics</td><td>TM RFC on 4D Simulator
TM RFC on Virtual/Augmented Reality and Discovery
TM RFC on TM Data Graph
TM RFC on Large-Scale Inference Engine</td><td></td></tr>
<tr><td>Super Computing</td><td>TM RFC on General Standards for the Super Computing Architecture</td><td></td></tr>
<tr><td rowspan="2">Humanities and Social Sciences</td><td>Theory</td><td rowspan="2">TM RFC on Large-Scale Inference Engine</td><td rowspan="2">TM RFC on LTM Framework's
TM RFC on Training
TM RFC on Enhancing collaboration
TM RFC on Knowledge Transfer</td></tr>
<tr><td>Disciplines</td></tr>
</table>

3 Research and Innovation Plans

3.1 State of the art

Infrastructure

Mapping of Cultural Heritage resources
Cultural Heritage (CH) mapping remains often limited to the macro-spatial level of towns/municipalities and to present-day maps and landscapes. There is a Lack of updated, well-organized and freely accessible multi-purpose databases. Digital assets are most often connected to each other, linked and interrelated. Some initiative exists. CKAN is an example for a registry providing tools to streamline publishing, sharing, finding and using data. European initiatives such as ICEDIG – "Innovation and consolidation for large scale digitisation of natural heritage", currently supporting the implementation phase of the new Research Infrastructure DISSCo aiming to mobilise, unify and deliver information at the scale, form and precision required by scientific communities, or Archives Portal Europe, are proposing aggregation and curation platforms dedicated to a specific GLAM sector. However, we are still far from a global entry point to access in a uniform manner European Cultural Heritage resource.

Digitisation infrastructure and technologies
Some guidelines exist. The FADGI articulates common sustainable practices and guidelines for digitised and born digital historical, archival and cultural content, this collaborative effort is divided in two working groups: Still Image (developing guidelines for still image materials) and Audio-Visual (working on sustainable technical guidelines methods and practices for digitised and born digital sound recordings and moving images).
Digitisation companies. Several actors are evolving on the digitisation field such as:

- Picturae (specialised in 2D & negatives, Audio-visual and Herbarium)
- Fraunhofer IGD (specialised in 3D scanning)
- Iconem (specialised in the digitisation of endangered cultural heritage sites in 3D)

Open-source resources. A small number open-source bricks exist. For instance, Goobi is an open-source software application for digitisation projects, it allows modelling, management and supervision of production processes, handling all steps involved in creating a digital library.
Nonstandard archival record. One main Challenge is the mass of non-standardized archival records. Digitisation remains concentrated on 'easy' formats, there is little knowledge of the existence of infrastructure for digitization of 2D or 3D materials and their linked costs. Multispectral, X-ray, or robotic digitisation are seldom used.
Costs of small institutions. Most small CH institutions with local items have no means to digitise their assets

Storage infrastructure and technologies
Cultural Heritage institutions are doing small-scale efforts to keep their data accessible and safe. They need to analyse and define preservations policy and workflows, including parameters like number of replicas, technology behind, accordance with archiving models etc. Today there are several actors active in storage market and CH institutions need to clearly set their needs before choosing an operator. Various initiatives have resulted in the creation of data management tools or general-purpose open-access repositories offering short-term data sustainability mostly for research data in the context of open science such as: figshare, zenodo, Dataverse, EOSC (European open science cloud) ; or long-term storage facilities such as: archivematica, islandora, preservica. Those services are often offered cloud-hosted.

Linked Open Repository
A triplestore or RDF store is a purpose-build database for the storage and retrieval of triples, subject-predicate-object datastructure. For instance Druid (CLARIAH) is an example of a triplestore.
Linked Data is composed of connected triplestores. The archiving of Linked Data was first addressed within the Prelida EU Project (2013-2014) establishing that this community has not been traditionally

targeted by the Digital Preservation community. Most of the above storage repositories can handle triples delivered in RDF and / or JSON-LD formats. The remaining challenge is to be able to connect every triple to the proper ontology and link authorities to triples. Several efforts have been made in this direction including for instance SeCO (Semantic Computing Research Group) or OntoMe (Ontology Management Environment),

Generic document processing
This section and the following talk mostly about technology and issues covered by the pillar 1 roadmap. Indeed, results of pillar 1 research will lead to the definition of infrastructure choices through the establishment of corresponding RFCs.

In recent years, there have been multiple successful attempts tackling document processing problems separately by designing task specific hand-tune strategies. The digitisation workflows differ depending on how advanced the institution is. The diversity of historical document processing tasks prohibits to solve them one at a time and shows a need for designing generic approaches to better handle the variability of historical series. We distinguish different stages within a document process, involving firstly the automation of text localization on an image, document understanding, natural language processing and optical Character Recognition (OCR: a technology that enables the transformation of fixed files into editable and searchable data). Various open source engines are available supporting some of the listed processes, such as tesseract, kraken, transkribus. All current solutions use machine learning. Recently released, dhSegment offers a generic approach to segment regions and extract content from different type of documents, it was developed within the context of the Venice Time Machine in need of a more generic tool for dealing with huge amount of document.

High performance Computing
Data processing in the context of digitisation is done in small server rooms at CH Institutions, resulting in difficulties of scaling those processes. Existing means of processing large datasets: Spark, GPU cluster (allowing very fast calculation performances).

Deep Learning Frameworks
There is currently a limited number of Deep Learning applications for multimedia data related to CH. In the very recent years, a growing community in Art History and Natural History collections has been exploring deep-learning based machine vision approach to help search through large collection, but these efforts are still essentially done as part of academic research programs. Google and Facebooks have released highly popular deep learning programming environment which tends to be used by a large majority of researchers in the academic and industrial world.

4D Technologies
Although 3D technology is progressing fast, in particular photogrammetric approaches, there is a lack of effective data management practices and tools, which constitutes a barrier to create efficient 4D data models. Time is essentially not included in most management tools. In addition, most tools are still based on a desktop approach and intrinsically limited for scaling to larger spatial models.

Inference Engines
Inference engines were an extremely popular Symbolic Artificial Intelligence technique and many development framework and programming language existed. The success of highly scalable Deep Learning approaches has somehow put less emphasis on these work. Privately own system like IBM's Watson program and Wolfram Alpha are still in operation but not directly usable as open code for building larger infrastructure.

Communities

Due to the diverse nature of the various communities the following section will detail each type of community separately.

Developers

The international community of developers is very diverse and only loosely organized around certain topics, technologies or areas of interest. There is no distinction in general between professional developers and hobbyists as both have equal access to the general means of communication and individual communities.

Generally, there are no big formalized organizations with some notable exceptions. For instance, the "Chaos Computer Club", a mainly German organization of "hackers" has around 7000 members according to Wikipedia.

Due to the interconnected nature of working on and with technologies, the main interaction method is the participation in virtual communication platforms, starting with *UseNet* and *IRC* in the 1990s and early 2000s, online forums, email newsgroups and nowadays systems like Slack, Discord, particular areas of the virtual Forum and Social Media Site Reddit and similar services as well as the most widely used site to get answers to technical questions, Stack Overflow. Most of these platforms allow the formation of groups focussed on specific topics like a programming language or a technology. Developers frequent these sites to get help with a problem, to talk about areas of interest and just to socialize. Of course, like other communities, developers also use platforms like Twitter and Facebook.

Apart from virtual communication platforms, there has been a number of physical communities as well. On the one hand, there are platforms that are focussed on the creation of small localized physical meetings (for instance the platform Meetup). On the other hand, there are a wide number of so-called *Hackerspaces* that are physical locations, often managed by a local association. For instance, the Metalab in Vienna is organized by an association called *Verein zur Förderung der Erforschung und Bildung sozialer und technischer Innovationen - metalab*.

Apart from communication tools and physical places, developer communities often form around a specific technology, often an open source software product that is collaboratively created by a community of volunteer developers. The main platforms for this interaction are the Source Code Management Platforms GitHub and Gitlab. These repositories are often the starting point for the development of wider communities that focus on the usage and development of this specific software.

Educators

Educators from all different levels of education, most often grouped together in primary, secondary, higher and adult education are mostly employed by the public services and are regulated by various laws as to what topics they focus on, what technologies they use and what materials are available for education. The official materials are on the one hand created by educational writers and on the other hand accompanied by primary and secondary sources of various kinds.

Professionals

The communities that people working in archives, museums, libraries, documentation centres and similar institutions are organised in are either focusing on general aims of their domains, or on very specific aspects of the field. Most people are, therefore, members of more than one community, even though they might not realise it. Most of the communities are organised by a foundation or an association that is funded by membership fees of the participating organisations or personal members. On a European level, we estimate that there are about 25 organisations that formally organise members and institutions in the cultural heritage field.

Within the (digital) cultural heritage domain the only cross-domain network is the Europeana Network Association. This organisation consists of over 2,000 heritage professionals (in personal capacity)

that are members to promote the cross-domain interoperability of Europeana. Members can influence the policies and progress of Europeana by joining the communities that focus on several specific cross-domain topics and more specifically by proposing task forces to solve problems on a given topic. The Europeana Network Association also engages in advocacy on behalf of its members towards the European Union and its member states.

To better interact with national, and regional existing aggregation platforms and associations (*an aggregator is an organisation providing a service that acquires, collects, validates, harmonises, stores, and often enriches data from cultural heritage institutions which have digitised collections and then deliver this data for showcasing in Europeana.eu*), Europeana has created the Europeana Aggregator forum, organised locally and thematically, offering a place for professional exchanges and ensure the involvement of this strategic partners in operational decisions. Aggregators offer their partner institutions advice on support in digitisation, content description, licensing, media formats accessibility, multilingualism and domain and subject vocabularies.

Scholars

The different parts of the community are mostly not well interconnected. There are plenty of projects, corpora and different type of databases that are internal and available for a certain institution or closed community. There are certainly also open resources with free access.

Volunteers

Most projects are rather closed around themselves. Typically, a project is based in one institution, involving one or few types of materials, and recruiting volunteers for only that projects. There are most probably several hundreds, of such projects throughout Europe today, which basically each one, has invested in making online indexing-platform which are more or less identical. There are also relatively few platforms more generic in their scope. One example of this is the Dutch website VeleHanden, hosting projects from many institutions on one platform – allowing volunteers to choose and switch between different projects, or the Danish crowdsourcingportal, aimed at being able to set up bespoke transcription projects for every form of digitized documents in the Danish National Archives collections quickly, and thereby being able to also facilitate user-driven initiatives, supporting a very “outside-in” way of thinking user-involvement. A third form of more generic form of projects, not necessarily involving volunteers, are projects like the Norwegian Historical Population Register, or the British populationpast-project. The idea behind these projects is to link together information of persons from different sources (e.g. censuses and parish records), thereby being able to quickly overview a given person´s appearance in different sources, making these registers very potent tools for educational purposes, genealogy and for historical, demographic and even medical and social health research.

User involvement related to cultural heritage materials are by no means only hosted by public or semi-public cultural-heritage institution. In Utah, the Church of Jesus Christ of Latter-day Saints has included indexing-projects as an integrated part of FamilySearch.com and in Australia, the Nicolaysen-family, by adapting an extremely user-involving strategy, has been able to make Danishfamilysearch.com maybe the most widely used resource for Danish genealogists. A major commercial player on the genealogical scene like Ancestry, has opened up for volunteers contributing to quality control of their online materials.

Cultural Heritage institutions have realised, that using volunteer communities is no free lunch, and that volunteer is not a synonym for non-expert. Volunteer communities as e.g. people organized in genealogical societies, railway-history societies, medical-history societies and so on, will typically be retired professionals with a deep knowledge of whatever their interest is and lots of time, dedication and energy to put into their hobby. In many cases, it would be appropriate to name them volunteer-experts. Many of the organizations e.g. The Danish National Society with approximately 8.000 members, are organized within the boundaries of one country, and in many cases with only limited experience in cross-country cooperation, but are within these limits still very capable organizations

with well-functioning leading boards, able to formulate effective strategies and accumulating impressive economic and human resources and thereby being very attractive partners in large projects. As an example, 50 volunteer scanning-operators working in the 4 reading-rooms of The Danish National Archives, and organized in a collaboration between the Archive, the genealogical society and others, are expected to deliver more than 4 million pictures to the National Archives webservices in 2019.

To muster these cooperations and projects, cultural heritage institutions are setting up regular volunteer-handling organisation, headed by dedicated officers with title as – in Denmark and Norway – crowdsourcing coordinators. Among the tasks of these institutions is the recruiting of volunteers, daily communication and troubleshooting, establishing of rules and procedures, planning new projects and, not the least, negotiations with both individual volunteers and their organizations.
An example that shows that the activities of volunteers are not necessarily restricted on creating transcriptions or producing digital images is the Topotheque. The focus of volunteer work within this very wide enterprise, covering presently 12 countries, is to deliver primary sources regarding the historical development the lowest level of social and stately conglomerations – the communities. This material (photos, document, films, audio files etc.) are mostly private owned and very highly endangered to be lost forever in case of the change of generations. Volunteer groups in this context are understood in a very wide sense since almost every household possesses relevant material. The volunteers are divided into two groups: on one hand, we have the almost endlessly big community of content providers and on the other hand the also volunteer persons uploading the material and providing it with metadata.

In this case, the term of the above-mentioned *volunteer-expert* is accurate, because they are the only people to know anything relevant about the local history of their village, family or house and professional experts depend on a high degree to the knowledge of the volunteers.

Local Time Machine

Data standardisation

Data standardisation is today an issue for GLAM, research and private industries engaged in digitisation processes and aiming to publish and valorise those data, as needs of standardisation have always existed and with the raise of linked open data, models have been adapted through time. Several efforts of standardisation have conducted to different metadata models and standards depending on each GLAM's industry's needs such as:

- CIDOC Conceptual Reference Model (CRM) intended to promote a shared understanding of cultural heritage information, which provides a common framework that any cultural heritage information can be mapped to (official ISO standard since 2006).
    - LIDO: An application of the model, aiming to provide an explicit format to deliver museum's object information in a standardised way (an XML harvesting schema).
- IFLA Library Reference Model (LRM), a high-level conceptual reference model developed within an entity-relationship modelling framework, designed to be used in linked data environments.
    - The METS schema is a standard for encoding descriptive, administrative and structural metadata regarding objects within a digital library, expressed in XML.
- ISAD(G) General International Standard Archival Description defines the element that should be included in an archival finding aid.
    - Encoded Archival Description (EAD), an XML standard for encoding archival finding aids.
- The Dublin Core Metadata Initiatives (DCMI), initially created as an answer to the silos situation, with the goal to better develop metadata for **online and networked resources**. By virtue of being a lowest common denominator, it is low enough to also describe physical resources. The Dublin Core Shema is an original set of 15 metadata terms.

If in the above listed schema are framing the structure (terms) of metadata, controlled vocabularies are playing a crucial role in clarifying the respective value of each field. In the GLAM industries, several are being used such as:

- LCSH (Library of Congress Subject Headings)
- Getty Vocabularies (Getty Research Institute)
- GeoNames

Europeana confronted to the difficulty of gathering those bibliographic information's coming in different standards, and establishing the lack of authoritative data from the cultural heritage community in the Linked Open Data environment, created their own specific model, the Europeana Data Model, meant to ease exchange protocols and foster interoperability. It uses Dublin Core and related controlled vocabularies among other technical means.

Data interoperability

The multiple data standards based on a domain-model approach has led to incompatible standards that make it really difficult to share data between libraries, museums and archives. Cultural heritage institutions are mostly sustaining their data in silos, isolated from one another and isolated from the wider ecosystem of the web.

Protocol exist to enhance federated search, with their own limitations such as:

- OAI-PMH: Data Providers are repositories that expose structured metadata via OAI-PMH. Service Providers then make OAI-PMH service requests to harvest that metadata.
- IIIF the International Image Interoperability Framework defines several application programming interfaces that provides a standardised method of describing and delivering images over the web.

The Linked Data approach is designed to support heterogeneous descriptions models and built following a bottom-up approach, allowing each institution visibility and ownership of their own data. It is grounded by four principles:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI provide useful information
4. Include links to other URIs so that they can discover more things.

Even if means exist, and new technologies are evolving quickly, there is today no global model for data publishing and interoperability in GLAM industries. Institutions are implementing one or several of those tools often depending on their own context.
In addition, URIs being location-based, proved to be not necessarily the most efficient approach to give a stable referent to resources.

Selection and connection of sources

Documentary traditions are greatly diverging within Europe. Sources are often scattered in many different institutions, with general view of their content. It is therefore difficult to access requested information and evidences granted are often bad. Once the sources identified, they're often not digitised as it is closely related to individual institutions and research project funding. Europeana has conducted an overage study through the ENUMERATE Observatory, which offers a reliable baseline of statistical data about digitisation, digital preservation and online access to cultural heritage in Europe. Different aggregation platforms aim to valorise and inform about existing digitised datasets, such as: Archives Portal Europe.

Legal framework

The legal framework surrounding digitisation of data is linked with nationals or regionals law related to copyright. A WIPO study shows that libraries and archives work under a patchwork of provisions

that differ in scope and effect from country to country, making cross-border working unnecessarily complicated. Licences such as Creative Commons are one of the tools allowing institution to take care of copyright issues, but surely is not a cure-all solution either. They are today one of the solution proposed to cultural institution to answer those legal issues.

Europeana proposes a series of solutions to GLAM institutions: A core element of the Europeana operating model is the licensing framework which sets the conditions for the exchange of metadata so that they are interoperable and clearly labelled with rights and reuse information. It is developed in an integrated manner with the Europeana data model mentioned above. Together it means that for each of the 58 million objects currently published and conform to the data model, the metadata are available via CC0 (*Creative Commons*) and each object is labelled with rights info.

As a member of the RightsStatements.org Consortium together with the Digital Public Library of America (DPLA) and several other institutions from Australia, New Zealand, India, US and Canada, Europeana proposes a set of standardised rights statements, aiming to propose ways to communicate the copyright status of a digital object, easing its use and re-use.

Financing model

They are several financing models for digitisation, driven either by external funds (on-demand platforms, sponsoring, grant's opportunities) or internal. They might result in potential conflict between a) funding from scientific research b) funding from heritage policies and c) open access requirements.

Labelling system

The UNESCO is curating a World Heritage List aiming to identify cultural and natural heritage of "Outstanding Universal Value" and therefore worthy of special protection against the dangers which threaten them. A list of ten criteria has been established aiming to identify those cultural and natural heritage.

The RightsStatements.org Consortium is currently undertaking research into how indigenous rights could be acknowledge, but there isn't any online publication available yet.

Local Time Machine Initiatives

Following the example of the Venice Time Machine, several initiatives have raised prior and during the CSA phase. As they still lack a common infrastructure and framework, they operate more on a national or regional level rather than at a European scale.

Once the TM network settled, their respective operation will be accurately mapped and monitored. Below a brief glance of some of this LTM initiatives and their current on-going activities.

LTM	Goals and methodologies	Achievements
Venice	The project started in 2012 with the goal to analyse 1,000 years of maps and manuscripts from the city, by scanning documents including maps, monographs, manuscripts and sheet music, enabling researchers to search and cross-reference the information, thanks to advance in machine-learning technologies.	Most tools are now incorporated in diamond.timemachine.eu (including morphological image search, historical geographical information system and handwritten text recognition). Generic structure that can easily be scaled for the other TMs has been put in place.
Amsterdam	The Amsterdam Time Machine (ATM) is a hub for linked historical data on Amsterdam. Born in 2017, it brings together efforts in the fields of academia, cultural heritage and computer science to digitally unlock Amsterdam's past. Ultimately, the web of information on people,	The Amsterdam Time Machine is built upon linked data infrastructures from key academic and cultural heritage institutions in the Netherlands, including CLARIAH and Adamnet. It benefits from funded research and heritage projects

LTM	Goals and methodologies	Achievements
	places, relationships, events, and objects will unfold in time and space through geographical and 3D representations. While we're working on that, we provide access to the three building blocks of the Time Machine: a Linked Data cloud visualisation called ALiDa; historical Maps and other geo information; and 3D reconstructions..	that digitise, explore and remodel historical data (AdamLink, Golden Agents, Virtual Interiors, Alle Amsterdamse Akten, CLARIAH ATM project) and develops applications in monthly or bi-monthly data sprints, which have so far developed a historical navigation interface ("Years"), map demos (ATM Cesium), linked datasets and search engine prototypes.[2]
Antwerp	The Antwerp Time Machine grew out of setting-up an Historical GIS (Gistorical Antwerp) from 2011 onwards, aimed at reconstructing five centuries of urban development at the level of individual houses and households. Its focus is on spatial humanities; long-term developments; connecting sources; and making an interface which connects micro developments with macro urban change and evolution. However, efforts of geo-spatialization, digitization and data-integration are still costly, time-consuming and fragmented. Several technologies – such as automated transcription of handwritten sources and A.I. enhanced deep-learning methods of maps – are still in an experimental stage.	Most tools are being developed in the context of https://www.uantwerpen.be/en/projects/antwerp-time-machine/ (including image searching; historical geographical information systems and handwritten text recognition). It basically follows the same methods of connecting sources & data through space; and developing automatized forms of text and image recognition that are also being developed elsewhere. Cadastral sources are used as template for HisGIS ince end of sixteenth-century onwards; data are linked to this system derived from urban archives and other CT partners.
Nuremberg	Reconstruction of the largely destroyed historical city of Nuremberg in 3 time layers: 1620, 1811, 1910 (and additionally 2016). Integration of digitized materials from archives and museums. All elements are georeferenced. Current limitations: size of the WebGL 3D model takes very long to load and only runs in Chrome.	The project is planning to cover in the next three years 50% of its urban space on 3 time layers, digitising 10 000 objects and referencing 5000 external resources. Focus will start on the 19th century and progressively include the 18th (2023-2025) and 16-17th (2025-2028)
Frankfurt	Reconstruction of destroyed or no longer existing buildings; reconstruction of life paths based on the biographies of people starting from stumbling blocks.	
Sofia	Aiming to build a unified and well-structured coordinating system between different archives, libraries and archives. Creating an excellence cluster for Cyrillic written cultural heritage. This will require to develop the current state-of-the-art: Partially digitisation of the documents of National Library, National Archives, The Archives of the Academy of sciences and University Library. There is also a Digital Archive of primary sources from several monasteries and local archives situated at the Library of Sofia University. The information is sporadic; the metadata is basic and not unified. Partially digitization at National Historical Museum and National Archaeological museum.	

[2] These outputs can be found in http://amsterdamtimemachine.nl/category/projects/ or http://amsterdamtimemachine.nl/category/interfaces

LTM	Goals and methodologies	Achievements
Madrid / Iberian node	The Iberian node, which will be able to become the physical access point for all Spanish and Portuguese entities, (at present: Council of Santiago; Diputación Provincial de Ourense; SPE Gran Canarias, Universidad de Salamanca; CTA Technological Corporation of Andalucía). The key idea consists of this node being accessible through Cloud Technology and Semantic Services. Any other Local Time Machine will also be able to connect to this infrastructure via an external interface.	

3.2 Targeted achievements

As the PILLAR 2 outcomes imply various aspects of the TM, this present section has been split accordingly. Each sub-section contains specific targeted achievements and a brief description. Please note that as this is a draft document, all listed proposition could be improved and modified by other PILLAR results.

Overarching achievement: TM Request for comments

> ***TM Request for Comments.*** *Publication and reviewing process for TM infrastructure and Local Time Machine is incrementally defined*

As the development of the TM operations is an iterative process, the entire step-by-step negotiation will be done through the writing, editing and publication of **TM Request for Comments**, inspired by the process by which the Internet Society establish its international standards. Establishing this process is one of most urgent achievement as it organises the process for all the others infrastructure developments.

Infrastructure

> ***Digitisation hubs.*** *Facilitate, standardise and optimise digitisation process at European Scale.*

The TM digitisation infrastructure will be composed of a network of digitisation hubs and will be organised on a European scale. A peer-to-peer platform will be in charge of managing and optimising digitisation strategies at European level, and will also be tasked with the development of generic solutions for archiving, directly documenting the digitisation processes, and swiftly putting the digitised documents online.

> ***Processing and simulation****. Development of distributed super computing infrastructure for processing Big Data of the Past.*

TM Super Computing Architecture is composed of distributed computing resources for Digital Content Processor and at least three simulation engines, will allow users to generate possible pasts and futures from the TM Data Graph.

> ***Virtual/Augmented Reality and Discovery.*** *Development of innovative infrastructure grounding for new experiences based on the TM Data Graph.*

The data generated by Time Machine, will allow creation of new experience for the users, both in term of content and discovery interfaces.

Distributed storage *Development of the distributed storage infrastructure for both public and private data*

TM Distributed storage infrastructure will be a fully decentralised highly redundant architecture based on the shared resources of TM Infrastructure Alliance partners. Its main aim will be to store the core public datasets of the project (The Great Commons) and possibly private datasets associated to the project.

Communities

Community interface *– Design TM infrastructure in a way that enables interfacing with existing communities*

The different parts of the TM infrastructure, for instance the source code powering the various platforms, the design of software portals for crowdsourcing or the reference mechanism for scholarly articles will be designed in a way that enables the connection to existing community infrastructures. For example, source code will be published on GitHub to enable direct contact to open source developers and user accounts for crowdsourcing volunteers can be connected to other social media or community accounts. Data can be shared.

Community inclusion *– Tailor communication to various communities*

In cooperation with WP7 the TM includes external communities in communication efforts of various milestones and designs specific communication activities tailored to different communities. For instance, descriptions of new software releases, internal architecture or the selection of various technologies is communicated via articles, forum posts or similar means to developer communities. Hackathons with the TM APIs will be organized. Success or post-mortem reports will be produced on crowdsourcing activities. Guest articles by scholars involved in the TM are invited and published in the appropriate communication channels created with WP7.

Community statistics *– Enable tracking of community involvement*

To enable tracking of and reporting on community efforts, mechanisms to collect statistical data on individual participation, impact on the TM data and similar metrics are included in appropriate parts of the TM infrastructure, mainly user accounts and contribution transparency.

Local Time Machines

TM's rules and recommendations*. Ensure cohesion and network's operation by the implementation of general values and technical standards.*

Gathering all data generated by a digitisation process or already existent ones, will not be possible without agreement on minimal technical specifications, bounded with technical needs of the TM infrastructure. Defining a common framework and a proper starter kit implies also to set clearly some values and perimeters criteria aiming to support common objectives and regulate data acquisition, data sharing and data publishing.

Framework. *Set up a support structure aimed at launching an LTM and ensuring a regular and smooth development of the TM network.*

Any partners wishing to integrate or launch an LTM should be redirected to a clear path and find requested guidance during the whole process. As active partners, they would integrate a network and should therefore respect the TM rules and TM recommendations.

***Labelling system**. Set up a value scale in order to precisely evaluate the progression of an LTM and support the TM network's dynamics.*

How to encourage any LTM to enter in the operational phase and contribute to the growth of the TM Data and Operations Graphs, and how best guarantee member's commitments towards TM's objectives? The labelling system should contribute to answer such issues.

***Legal setting**. Guarantee the respect of national and European policies and laws in order to build a coherent and standardised contractual and licensing system for TM's network operations aiming to regulate data acquisition, data sharing and data publishing.*

Many trade agreements are willing to take place within the TM's network. Standardisation will prevent malicious use and ease those processes, as well as guaranteeing conformity with national and European requirements.

***Financial system**. Foster financial independence of all TM's instances.*

Funding sustainability plays a central role in the building of any PWTML, LTM's partners would need to conduct efficient financial research, and the services offered by the TMO (infrastructure and coordination) will also require support. A franchise model will be shaped consequently.

***Incorporation, collaboration and enhancement**. Enabling existing initiatives to be aligned and consolidate the global network by encouraging connections amongst TM's partners. Contribute to the future development of the network by supporting the creation of new LTMs (top-down process). Ensure knowledge transfer amongst the partners.*

Existent LTM and PWTML initiatives will be included in the TM's network. Global heritage networks (e.g. Europeana) are partners of the TMO and therefore should be invited to contribute in the network, and share their expertise and best-practices, enhancing a global European collaboration system. Top-down initiatives will also contribute to the system's growth, and as it does, knowledge and technical achievements will be shared amongst network's partners.

***Smart cluster**. Ensure the identification and future creation of local competitive assets*.

Once data gathered within the *Big Data of the Past Graph*, how to encourage future exploitation avenues to identify opportunities, build on their regional particularities and shape innovative cultural-heritage experience platforms and services?

3.3 Proposed methodologies

Overarching methodology: TM Request for comments

Motivation. To ensure the open development and evaluation of the Time Machine Operation, a process inspired by the Request for Comments (RFC) publication mechanism for negotiation the standards and protocols will be established. Standard themselves may come and go, but the open way of establishing them and negotiating them should persist.

Principles. The basic principles of the TM Request for comments will be the following:

1. **Accessibility.** TM RFC are freely accessible, free of charge.
2. **Openness.** Anybody can write a TM RFC,
3. **Identification.** Each TM RFC, once published as a unique ID and no changes are allowed after publication. Any important changes result in a subsequent TM RFC. For this reason, some TM RFC could be tagged as obsolete.

4. **Incrementalism.** Each TM RFC should be useful for its own right and act as a building block to others. Each TM RFC must be aimed as a contribution, extension or revision of the TM infrastructure.
5. **Standardisation and linguistic diversity.** TM RFC should aim to make use of standardised terms to improve the clarity' level of its recommendation but can be written in any language. Once published they should be translated in a maximum of language.
6. **Scope.** TM RFC are designed contribution and implementation solutions solving practical problems. TM RFC are not research papers and may not necessary contain experimental evidence.
7. **Self-defining process**. Like for the development of the Internet, TM RFC could be the main process for establishing TM Rules, TM Recommendations, TM Standard Metrics but also the processes and roles for managing TM RFC themselves

Publication Process. To bootstrap the publication process, the initial publication pipeline will follow the following stages

a. Submission of RFC text and figures
b. Handling by RFC editors naming RFC reviewers.
c. Open Review Process
d. Attribution of TIME RFC ID and a DOI publication with names of reviewers disclosed and possible additional comments by them.
e. Translation in several other languages.

This publication process, constitution of the review committee could be established through the RFC process itself.

On-going redaction process. In order to smooth the coordination of writing of RFC, authors are invited to announced when they are currently working on a giving RFC, giving possibility of other contributors to collaborate.
Examples of possible TM RFC are given in the following pages.

TM RFC Editorial Committee: In order to bootstrap the process, a small group will be created, the TM RFC committee. The committee will have to decide on:

- **RFC Reviewing and Publication Platforms.** It will be crucial to decide on the technology and platform for managing the publication process of RFC. White-labelled version of existing publication platform may be considered.
- **RFC Committee's rules of appointment and renewal**

Infrastructure / Digitisation Hubs

The Digitisation Hubs will cover regional digitisation needs with standardised hardware for digitisation, storage, information exchanges and on-demand scanning, based on results of PILLAR 1 and existing metadata standards, like the one developed by Europeana. The peer-to-peer platform will federate system integrators at European level, facilitating the deployment of this equipment.
TM will establish both top-down digitisation pipelines optimised for efficiency and bottom-up on-demand digitisation services driven by scholarship questions or the establishment of particular projects. X-ray, multispectral and very high resolution scanning will initially be used for on-demand test cases (e.g. analysis of a drawing under the surface of a painting). They may integrate the massive pipelines if their costs become affordable and their relevance demonstrated. All these experiments will be monitored in the **TM Operation Graph**.

The development of the digitisation hubs will be done through the publication of several TM RFC

TM RFC on Definition of typologies: A typology of digitisation interventions will be established, separating

a. collections that can be moved and processed in digital hubs (large, non-fragile collections), collections or objects that need local intervention (e.g. very fragile document, statues, buildings),
b. process that can be performed by volunteers using mobile technology (e.g. scanning campaign across cities, on-the-fly digitisation in reading rooms),
c. processes that can be performed using robots and drones, etc.

TM RFC on standardisation and homologation: Definition of the terms and contracts enabling digitisation partners to become part of the Time Machine network.

TM RFC on Open Hardware: Definition of the open hardware strategy for Time Machine including licensing terms, catalogues,

TM RFC on Time Machine Box: Also meant to cover storage needs for data, the Box should help partners involved in a data acquisition, sharing or publishing process to conform to the metadata specifications and delivery – harvesting protocols as stated by the data model. One of its goal is to smooth and contribute to the automatization of the digitisation process (offering for instance a way of monitoring the digitisation tasks). This hardware is part of the Time Machine Official Components. A first prototype (see the Time Machine Box website) has already been put in place and is currently tested in Venice (part of the Venice Time Machine project outcomes). The RFC will define how the production of the Time Machine Box should be manage in the long run.

TM RFC on Synergy and interaction in EU Research Infrastructure: The TM digitisation network will build upon existing EU Research Infrastructures (DARIAH, CLARIN) and infrastructures providing access to CH (Europeana, Archive Portal Europe, etc.). TM will introduce new processing pipelines for transforming and integrating CH data in such infrastructures.

Other documents may include:

- **TM RFC on on-demand digitisation**
- **TM RFC on Global optimisation of digitisation process**

Infrastructure / Processing and simulation

TM Super Computing Architecture is composed of distributed computing resources for Digital Content Processor, intrinsically easier to run in parallel, and three simulation engines, will allow users to generate possible pasts and futures from the TM Data Graph and needs for centralised resources. Partners in the TM Network offer Computing Resources that forms the Time Machine Super Computing Architecture. Several TM RFC will define the requirement for the incremental development of this Super Computing Architecture. The different RFC correspond to the different component of Figure 2.

TM RFC on General Standards for the Super Computing Architecture: This document will define the general rules that the TM Network partners have to follow to integrate their computing resources in the TM Super Computing Architecture and the routing processes managing the data pipelines. This document will particularly define

- The software and hardware standards that the computing resources will follow across the entire distributed Super Computing Architecture.
- The routing protocols of the TM Operation Graphs.
- The processes for naming and renewing the administrators of the administrators
- The role of the TMO for managing of the infrastructure.

TM RFC on TM Data Graph: The Time Machine Data Graph contains all the information modelled in the Time Machine. The graph is constructed both manually and automatically through the processing of the Digital Content Processor. The Graph is intrinsically composed of two subparts

a. The bright graph composed of information that have been manually mapped and integrated with other large database or used in a publication. This information is integrated with the current sum of digital human knowledge. It can be considered actual.
b. The dark graph composed of information extracted automatically from (massive) documentation which has been used so far apart as individual historic items. It can be considered virtual.

The Dark Graph is likely to be quantitatively much bigger that the Bright Graph. The role of this document will be to establish the implementation strategy for both graphs and their interconnections. One may use for instance Wikidata as a shared ontological reference for the Bright Graph and UUID systems for naming entities in the Dark Graph.
The document will also specify the use of both the Bright and Dark Graphs for indexation of searching engines and distant viewing analysis.

Bright Data Graph	Mapped manually with other database or cited in human-written publication, part of the Digital Human Knowledge, Actual
Dark Data Graph	Processed automatically from document or deduced by simulation engines, part of the Machine Knowledge, Virtual (waiting to become actual)

TM RFC on Digital Content Processor (DCP): Digital Content Processor are Automatic processes for 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. The document should define:

a. The technical condition for implementing DCP that can be inserted in the Time Machine Operation Graph.
b. The requirement for hosting DCP in an TM Super Computing Infrastructure.
c. The process by which DCP are developed, tested, labelled, published and puts in operations

Concerning point c the following pipeline may be envisioned.

1. Development of a DCP in dedicated “Sandbox” (a place where trial and errors can be made without compromising the entire functioning of the Time Machine architecture). Training will be done on existing labelled documents.
2. Submission of the DCP to the Time Machine Organisation dedicated service.
3. After some benchmark and assessment of performances acceptation or rejection of the DCP. The DCP becomes a TM Official Component.
4. Deployment of the DCL on the Super Computing Architecture and integration in the routine services.

TM RFC on Large-Scale Inference Engine: 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 derives 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. The document will specify the various kinds of rules that the Large-Scale Inference Engine can process including

- Rules extracted from documents by DCP
- Implicit Rules made explicit
- Rules (statistical or not) induced from the data.

The document should define

- the process by which rules are submitted, tested and integrated in the engine
- the processes for managing conflicting rules or results from various rules.

The document will also motivate implementation solution in relation with existing deployed systems like Wolfram Alpha or IBM Watson and standards like OWL, SKOS.

TM RFC on 4D Simulator: 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 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.** The document will specify two key elements of the 4D Simulator:

The 4D Grid	Through a hierarchical division of space (3D) and time, the system organises a multi-resolution 4D grid which serves as a general spatiotemporal index. Each “cube” in the grid indexes all the information relevant for these particular spatiotemporal elements. It offers an efficient perspective for organising the large datasets and performing collective curation through manual and automatic processes. Each element of the grid will also be potentially labelled according to other various multidimensional criteria, some of them being AI-based descriptors (e.g. descriptors for architectural style detection in images).
The 4D Simulations	The 4D grid is sparse as many places/times in the world that are not directly associated with existent archival data. A central research challenge is to develop AI-systems capable of extending the information of the data grid in space and time through continuous extrapolation and interpolation, and developing new ways of visualizing which part of the content is anchored in sourced data, simulated or unknown. Extensions of current deep-learning generative methods, originally developed for 2D imaging, can be envisioned to deal with the richness of the 4D datasets. Many 4D simulations can be associated with the same 4D grid and one central challenge is to manage this multiplicity of worlds and their specific resolutions level for various services of the Time Machine (e.g. entertainment, policy planning).

The document should specify the interaction of the 4D simulator with the rest of the architecture answering questions like:

- How can an entity of the TM Data Graph be associated to a particular element of the 4D Grid
- How can 4D simulations be run and cached for future use
- How can the system be used directly in exploitation platform

TM RFC on Universal Representation Engine: 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 time), 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. The document will specify the integration of the URE in the global architecture, outlining for example how a give node in the TM Data Graph can be associated with a parametric representation space.

Infrastructure / Virtual/ Augmented Reality and Discovery

TM RFC on Virtual/Augmented Reality and Discovery: Based on PILLAR 1 research outputs, this RFC will focus on creating means to enable innovative and enriching experience. The RFC will define the standards that should be adopted in order to enable the development of Virtual/Augmented Reality services and Discovery module on top of the TM Data Graph.

Infrastructure / Distributed storage

The TM distributed storage system will be aimed at offering an alternative solution to current HTTP based storage. Although in a first phase, most document and data will be stored on specific servers accessible through standard protocol (e.g. images on IIIF servers), the aim of the project is to develop a storage solution that would have the following objective

1. Giving access to high volume of data with high efficiency
2. Optimising storage to store more data
3. Implementing long-term preservation of data, preventing accidental or deliberate data deletion and keeping a fully versioned history of the data stored
4. Guarantying authenticity of the data stored and preventing the inclusion of fake sources.

Like for the other part of the architecture, the incremental development of this distributed data storage system will be done through a series of TM RFC.

The Time Machine deals with two family of datasets: Public and Private. Public data corresponds to all the datasets issued by Public administration that can be globally shared under creative common licences. Private data includes personal data (personal document, scans of interiors, etc.), data from companies and data from public administration that cannot be released publicly for ownership or privacy reasons. Independently of the public and private nature of the data, some data may be associated with restricted viewing condition because of content (e.g. creative common document containing violent, shocking, illicit content that may not be adapted to all audiences). The way to handle this issue will be dealt through a specific TM RFC.

TM RFC on Distributed storage system for Public Data: This document will define the infrastructure principle for a decentralized solution of public datasets based on Creative Commons licences like CC-0, CC-BY, CC-BY-NC and the Europeana rights declarations. Storage will be done on the resources shared by the partners of the TM infrastructure alliance.

A distributed system like IPFS (Interplanetary File System) and the work done by the IPFS Consortium for persistence of IPFS object may be a good starting point. Such kind of file system do not identity a resource by its location but by a unique identification number. Routing algorithms optimise through P2P algorithm the most efficient ways to bring the data to the visualization or computing processes. This also speed up process when the host is a region with low connectivity. Redundancy and long-term resilience can be guaranteed. This means that the system can be designed to make in practice undeletable any public data content that it starts to store, making it hard to censor content. For this reason, it is especially well adapted for public data associated with creative common licenses. Systems like IPFS also give the possibility for each node of the network to choose the categories of data they accept to replicate. This gives some flexibility in the negotiation of common strategy by the Time Machine Infrastructure Alliance.

For ensuring the authenticity of the data stored, a blockchain type solution could be the solution. The interaction between the distributed file system and the authentication solution will be defined by the RFC.

TM RFC on Distributed storage system for Private Data:

Private datasets could be stored either

- In specific layer of the distributed storage system provided reliable cryptographic and authentication systems are in place
- In "fenced" location otherwise offered by partners of the Time Machine Network as storage solution.

In both cases, the RFC should define how such closed dataset could use the services of the Time Machine infrastructure and under which condition.

TM RFC on Content Filtering

Content filtering may be necessary to control the exposure of users of the Time Machine services to unsolicited content. Finding the right technology to allow such a control without giving the possibly of abusive censorship operation will be the challenge of this RFC.

Box 3.3-1: TM Requests for Comments (RFC) / Infrastructure

- For Digital Hubs:
  - TM RFC on Definition of typologies
  - TM RFC on standardisation and homologation
  - TM RFC on Open Hardware
  - TM RFC on Time Machine Box
  - TM RFC on Synergy and interaction in EU Research Infrastructure
  - TM RFC on on-demand digitisation
  - TM RFC on Global optimisation of digitisation process
- For Processing and Simulation
  - TM RFC on General Standards for the Super Computing Architecture
  - TM RFC on Digital Content Processor (DCP)
  - TM RFC on TM Data Graph
  - TM RFC on Large-Scale Inference Engine
  - TM RFC on 4D Simulator
  - TM RFC on Universal Representation Engine
- For Virtual Reality and Discovery
  - TM RFC on Virtua/Augmented Reality and Discovery
- For Distributed Storage
  - TM RFC on Distributed storage system for Public Data
  - TM RFC on Distributed storage system for Private Data
  - TM RFC on Content Filtering

LTM / TM's rules and recommendations

As the TM infrastructures are currently on a construction phase, there isn't any common rule or recommendation followed by the partners of existent LTM or PWTML initiatives. Once values and technical requirements settled, they are meant to be the basic structure of the TM networks and play an essential part in the development of the TM Operation and Data Graphs. As technical means are due to evolve over time, those rules and recommendations will be flexible enough to allow future modifications.

TM RFC on data lifecycle (Data selection model, Data acquisition model, Data sharing, Data publishing):

TM deals with data on two levels: "raw historical data" (provided by research or GLAM partners to be processed, enriched and inferred to sustain the Big data of the past Graph and research data (models, scripts, methodologies, workflows, standard operating procedures, protocols etc.)

necessary to validate research findings. This RFC focuses only on raw historical data as the below **TM RFC on Knowledge transfer** will address the question of research data. As raw data are also part of research data, both RFC will share similarities and complement one another on some issues (specifically regarding sharing opportunities), they should prevent contradiction in terms.

Documents - Data selection: The data lifecycle within the TM starts with the concept of documents and data selection. The goal is to help partners select proper documents or collections to be processed by the TM pipelines and then the data within the documents. The data selection is closely related to the LTM or PWTML's perimeters (see LTM/Framework) and should be stated prior to any projects launch. Criteria such as intellectual property rights, obtaining copyright permissions, digitisations, OCR processing or metadata creation costs are also to be taken into concern.

A Data selection model based on the identification of performance criterion, will help partners focus on those aspects and take according decisions. The National Information Standards Organisation (NISO) also proposed a Framework for guidance for Building Good Digital Collections.

Example of performance criterion:

- Significance of Content to Internal Stakeholders (degree to which a collection, once digitised supports the immediate and long-term research and teaching needs of the institution)
- Significance of Content to External Stakeholders (a highly successful digital collection is of interest to researchers and users outside of the university)
- Uniqueness (many unique institutional resources such as original photographs, archival materials, grey literature such as university technical reports and conference proceedings remain to be digitised).
- Exposure (degree to which the digital collection garners the institution positive recognition and press and assesses the potential for the digital availability of the collection to result in grants and other funding).

Data acquisition: The goal is to precisely define what are the technical requirements for the data to be processed and ultimately being added to the TM Data Graph and published. What involves manual processing and what can be automatically created or calculated need to be precisely defined. Once first assumptions are established with the help of partners already involved in massive digitisation projects and used to deal with provenance, structural, and technical metadata issues (e.g. Europeana), they should be first compared with technical needs as defined by the **task 2.1 and PILLAR 1**, then gathered with partner's political and internal requirements to help the creation of a satisfying model which will finally be stated within the Technical Charter. What need to be further determined for the **Data acquisition model**:

- The need of a metadata schema and therefore the use of specific thesaurus, controlled vocabularies, syntax encoding, named entities and ontologies (if considered relevant).
- Prerequisite regarding harvesting protocol (OAI-PMH for metadata, IPFS, ResourceSync, SWORD) or use of delivery frameworks (such as IIIF) to enhance data extraction. The first assumption is based on the use of IIIF (Annex A – Technical Charter) and should be further developed.

Data sharing: Parallel to the data acquisition and simulations processes it might be relevant for TM partners to share specifics sets of their data for contextual purposes, such as research or training. Even if those sets are not yet published, data sharing will be supported amongst TM partners, with the idea to smooth such exchanges internal to the TM network. A **Data sharing model** should focus on proposing and identifying relevant methodologies (such as contractual forms, license agreement) and their prerequisite technologies. On a future perspective, such sharing processes could help evaluate cohesion and coordination within the TM network and be displayed within the TM Operation Graph and published on the TimeMachine.eu. Many process and prerequisite similarities link data sharing with the TM Data Management Plan (see below **TM RFC on Knowledge transfer),** both RFC should therefore complement one another and prevent contradiction in terms.

Develop a set of proposed methodologies to smooth the process (e.g. relevant data format, data models (e.g. Europeana Data Model).

- Propose technical means
- Display those interactions within the TM Operation Graph.

Data publishing: Once the data displayed within the TM Data Graph (after being processed by the Simulations Engines: Digital Content Processor, Large-Scale Inference engine, 4D Simulator, Universal Representation Engine), there should be made available for public or private uses. The **Data publishing model** should both ensure respect of licenses and therefore specific publications requirements and develop facilitating interoperable technologies and frameworks for data delivery. Such technologies should aim at encouraging the creation of new services based on the TM Data Graph.

- Develop a set of proposed methodologies (e.g. required metadata) to ensure respect of institutions licenses (e.g. machine-readable rights statements).
- Develop technologies for data deliveries (APIs etc.) and find suitable ways to implement them.
- Enhance interoperability through alignment of data models and accessible visual entry points

Once all models specified, this task should be dedicated to implement this standardised network, enhance a better identification of existing sources and develop monitoring and measuring means for relevant TM Operation activities and impacts of this global standardisation.

Documentation - Action:

TM RFC on Vision Mission and Values Charter: In order to protect overarching purpose, fundamental values and ethical principles, a common Charter will be created. Its duty will be to protect the core of the TM and sustain its future. Becoming a TM network member implies to ratify the Vision, mission and values Charter (Annex B – Vision, mission and Values Charter).
TM RFC on Technical Charter: The goal of the Charter is to guarantee a first level of standardisation for data and processes, in order to remain light and useable by the most, the charter also encourages the use of universal and open interfaces and references that do not need central coordination. The first attached version will be discussed and improved with the outcomes of PILLAR 1, task 2.1 Infrastructure and expertise from cultural-heritage network *(e.g. Europeana).* (Annex A – Technical Charter).

LTM / Framework

Existing projects and local initiatives have already been launched prior to the building of a common framework and several projects continue to be set up. In order to guarantee automated mapping of LTM's activities within the TM Operation Graph, advertising of on-going activities and efficient TM network's follow-up, a first proposal will shortly be put in place.

TM RFC on LTM/Framework's:

Partners typologies and routines – By its very nature an LTM is composed of various partner's typologies, each of them concerned by specific routines. The ones established within the schema (Key concepts and global overview – *Eligible partners*) will be optimised and further developed and their specific organisation and processes coordination needs identified and answered. Some of the partners (GLAM, Scholars) are also exploitations avenues and therefore those documents will be adapted and further developed with the outcomes of **PILLAR 3**.

Perimeter: In order to guarantee a smooth development of the LTM, a perimeter will be clearly settled for both LTM and PWTML. An LTM must be geographically rooted and PWTML contribute to

the space-time density' expansion of that same geographical location (Key concepts and global overview – *Local Time Machines)*. Those first assumptions will be tested and further discussed.
The main task of the RFC, once the definition settled, will focus on designing means to map those operation and routines within the TM network in order to ensure activities monitoring at all-time and build related guidance (as described below in Specific offices proposals).

Documentation - Action:

Welcome Guide - Outlining the global proceedings and detailing at the most each required step per partners' typologies, as well as advancing the benefit outputs per role. As the TM networks embrace several communities, the brochures will be multilingual. This outcome will be closely related to the **WP7** Dissemination and Promotion tasks.

Starter Kit – Step-by-step instructions defining what it takes to enter or launch an LTM or a PWTML. The already drafted scenarios (Annex C), will be stabilised and their scope extended to the predefined partners typologies. Proposed scenarios will also be tested during interview's sessions. Those scenarios will be turned into detailed instructions later published in the welcome guide.

LTM and PWTML forms: Various documents (accession form for a new member of the TMO, Joint declaration and roadmap form when launching an LTM, PWTML's perimeter form when launching a PWTML) punctuate TM's network operation. Templates "ready to be fulfilled" will be proposed by the TMO, meant both to guarantee minimal records of the network's activities and smooth launching processes for the partners. An exhaustive list of requested templates will be proposed, elaborate with the help of those already involved in LTM or PWTML initiatives.

TM RFC on Training: Complying with the TM Rules and Recommendations and the legal settings, using the TM components and understanding TM infrastructures, will require specific trainings. A proper set of documentation, tutorials, videos, online courses will be offered to the partners. This outcome will be closely related to the **WP7** Dissemination and Promotion tasks.

Specific organisational requirement:

Know-How, Guidance – Becoming an active partner within the TM network and getting involved in an already-existent or launching a new LTM or PWTML, surely would raise some questions. This will be taken into consideration and a specific office within the TMO should be dedicated to provide guidance upon request (this will be further developed according to **WP6** outcomes). This office will also be in charge of training's supervision and contribute to adapt the documentation through time.
Supervision, project tracking - Integrating the TM network implies the respect of its rules and goals, administrative follow-up will also ensure the respect of framework's proceedings (TM rules and recommendations), the signatures of Charters and requested projects documents, and the smooth development of the network. A specific office within the TMO should be dedicated to those tasks. (this will be further developed according to **WP6** outcomes)

LTM / Labelling system

The already in place initiatives aren't built around the same framework and pursue different goals. There isn't any existing label to inform about specific projects 'orientations' in the "rebuilding the past activities" field. Both a mean of identification and a mean of regularisation, the labelling system once in place will contribute to the global network quality, transparency and dynamism.

TM RFC on Value scale: The proposed LTM's value scale (Key concepts and global overview – *Local Time Machines),* based on density criteria, will be discussed and its relevance and organisation further developed. Value scale concerning PWTMLs will be created based on the already defined bricks and other criteria need to be proposed in order to foster the development of PWTMLs (e.g.

collaboration, cooperation metrics). As some of the measure will relate to qualitative process, a dedicated TM RFC on collaboration indicators, will focus on creating a suitable metrics system. The labelling system might require a third-party certification to assess its efficiency and respect.

Specific organisational requirement:

Supervision, certification – Obtaining a label, implies the accomplishment and respect of predefined tasks. There will be preferably a neutral office (*external to the TMO might be a solution*) in charge of label delivering and certifications' control. (This will be further developed according to **WP6** outcomes)

LTM / Legal setting

National rules do regulate commercial transactions. Heritage datasets and use metadata are also often linked with national policies and copyright protection laws. Actual LTM and PWTML have been taken care of those aspects so far, without receiving any legal advice or guidance by the TMO. As the network is meant to grow, it will be more and more necessary to ensure minimal contractual requirements and regulate data acquisition, data sharing and data publishing. The goal is both to support every partner in their commercial gestures and ensure their conformance with the TM rules and recommendations. The below methodologies are not due to answer to each local and national specificities but to guarantee the overall quality and legality of TM operations. The below assumptions should be further developed with results of **PILLAR 4, task 5.2 (Policy and legal issues and ethics)**.

Documentation - Action:

Contractual documents: The proper shape of required documentation (contracts template or paragraph to insert in a legal document) will be established with the help of legal advisors and representative of LTM and PWTML initiatives already engaged in digitisation processes. Franchise's contracts are included in contractual documents.

TM RFC on Intellectual property rights and licenses: Defining licenses to preserve intellectual property rights (regulating data acquisition, sharing and publishing) and sustain the interoperability and accessibility of the TM. The proposed solutions (Annex A – Technical Charter). are based on the Creative Commons copyright licenses and should be further developed with the help of cultural-heritage network *(e.g. Europeana),* used to tackle such issues and already proposing solutions. Means of monitoring the openness process of data should also be taken into concern.

Specific organisational requirement:

Legal – Providing help and guidance to PWTMLs and LTM in their network's operations, as well as ensuring the legality of actions conducted within the network and by extension the legality of TM activities. This office will take responsibility for TM contractual documents and coordinate contracts ratification and the respect of licences. The legal office doesn't have to be an instance of the TMO and can be tied to an external firm. (This will be further developed according to **WP6** outcomes).

LTM / Financial system

The existent TML and PWTML initiatives have already ensure their project's support, thanks to their own researches. This system should be strengthened in some extends. The TMO won't support project's specific needs but the TM label might help for grant's application, and coordination will be put in place as a service. All of the below proposed methods will be updated according to the output of the **PILLAR 4 task 4.4** (co-financing models with private money), **WP6 task 6.3** (money coordination and management structures**)** and **WP7 task 7.5** (support from public money).

TM RFC on Franchise: A franchise model clarifying financial relations between LTM and TMO's services will be put in place, according to the financial needs and costs of the technical and coordination infrastructures of the TM. The franchise system is meant to be evolving, as the network will grow along with its reputation in the public's eye, and the LTM's financial benefit generated. This franchise system is meant to be complementary to the one established for the TMO partners (**WP6**). To enter an LTM an institution should at least become a member of the TMO. One of the subtasks of this RFC is also to assess and further design the role of the TMO as a "Finance, Economic, Intelligence and Watch services office" (as described below).

Documentation - Action:

Supportive letters: Grant's application from TM's partners will be enhanced, upon request and depending on the context, with a TM supportive letter.

Specific organisational requirement:

Finance, Economic intelligence and Watch services– In charge of the collection of franchises' fees and the TMO's sustainability as well as acting as a watch service for call for projects and other relevant competitions. When relevant and to foster LTM's low governance structure, this office will serve as a bank, centralizing money transfers linked with a specific LTM's PWTML (Key concepts and global overview – The Time Machine Organisation). This office doesn't have to be an instance of the TMO and can be tied to an external firm (this will be further developed according to **WP6** outcomes). The existent national and regional funding schemes for digitisation will be scouted and mapped in **WP7 Task 7.5.**

LTM / Incorporation, Collaboration and enhancement

Existent initiatives LTM and PWTML are not part of a common framework yet, as the TM network does only exist in theory. Alignment of those initiatives will be one of the first mean to help the growth of the TM Data and Operation Graphs and will be done in the first years of the project. There are other PWTML's perimeter conformant projects, previously conducted but remaining on a stand-by phase for a couple of years or active initiatives supported by national or international funds, one of the duty of the TM network will be to propose ways to incorporate them within the TM network and contribute to their respective growth. The systematic process for contacting patrimonial institutions already engaged in digitisation programs will be developed by **tasks PILLAR 4 4.1 and WP7 7.1**.

As TM regroups some of the biggest networks related to Culture and Heritage, collaboration between them and TM network should be taken into concern in order to reach a win-win balance. Solutions to encourage partners to take part in other networks, or the creation of a professional hub to exchange best-practices, favours the share of contents and sustain digitisation infrastructures might be both suitable solutions and should be further investigated.

Sustaining information transfer and knowledge sharing throughout the network by defining both management and infrastructure means will de facto requires the definition of **Research Data Management (RDM) Strategy** and **Data Management Plan (DMP**) for the research data. This core documents will address issues related to data description, data quality and documentation, storage and backup during the research process, legal and ethical requirements, data sharing and long-term preservation, data management responsibilities and resources and ensure respect to EU regulations and directives.

Top-down initiatives with the goal to create new LTMs will help fostering the TM network.

TM RFC on Enhancing collaboration: Investigating how to supports partnerships across networks members of the Time Machine (e.g. Europeana, Icarus) or external cultural-heritage networks, and

what means would be a suitable answer. The reflexion should focus on different aspects: how to enhance collaboration internally and externally both at the level of the LTM and the TMO (e.g. with other LTM or partner's networks), how to enhance best-practices exchanges, share of content and collaboration with already existing cultural-heritage networks and associations. This RFC will imply coordination with the task 2.1 Infrastructure (*relations with the TM Infrastructure Alliance and support for the digitisation hubs*) and task 2.2 Community Management *(definition of platform for professionals*).

TM RFC on Knowledge transfer: Investigating how to support achievements and knowledge transfer inside the TM network. Ensuring a global research collaboration at a European scale. This task will be further developed according to **PILLAR 4 task 4.3** outcomes, but implies among other tasks the definition of a Research Data Management (RDM) Strategy (at the TMO level) and guidelines for Data Management Plan (DMP) (at the LTM level). There are rare examples of large-scale research data management models, dealing with similar complexity level as the TM, however some guidelines can be found such as: "Guidance Document Presenting a Framework for Discipline-specific Research Data Management" (*Science Europe, January 2018*), "Practical Guide to the International Alignment of Research Data Management" (*Science Europe, November 2018*).

Research Data Management (RDM) Strategy:

TM as a network of several LTM supporting diversified PWTML will not benefit from a single DMP as processes and technologies are closely bounded with each LTM's perimeter and associated PWTML. The RDM Strategy aims to fix standards for the global network, finding proper balance between LTM's autonomy and fixed criteria, ensuring success for the implementation of the RDM Strategy and support the data management life cycle for all research data that will be collected, processed or generated by the TM. The RDM Strategy must not conflict with the different public or private LTM's partners already existing data management or sharing policies.
The elements to be described within the RDM Strategy are yet to be decided, however some sub-tasks could be focusing on:

- Data management organisation and budget (*Who will be responsible for data management operations? What resources will you need? What standards, methodologies or quality assurance processes will you use? How will you organise your files and handle versioning?)*

Aiming to clarify coordination of data management between LTM's instance and the TMO, and define what specific role will need to be undertaken in the process. Need for specific training will be answered and organisation costs determined.
Standardised models for data formats, file naming, or metadata schema will be developed, aiming to ensure global consistency.

Existing suitable metadata standards in the relevant discipline shall be adhered to and provide guidance for:

- Descriptive and provenance metadata (e.g. Dublin Core, PREMIS)
- Substantive metadata: guidance on how data should be read or interpreted (same or similar of a README file).

Non-proprietary or open standard file formats will be used to ensure accessibility, reuse and long-term preservation such as:

- For images: TIFF (.tif, .tiff), or PNG (.png)
- For maps: ESRI shapefiles (.shp, .shx, .dbf)

Research Compendium and data packaging: the compendium is a tool allowing research to be gathered within one virtual place (Marwick et al., 2018), deployment of a similar tool within the TM should be further assessed.

- Internal data sharing policy

This document should specify access to data at any time (prior to public publication) and provide information regarding when an LTM can access to others LTM's research data and how to respect existing practices and regulations. How to deal with sensitive data and pertains data security measures in the context of the TM should be addressed. Each LTM would be responsible for writing their own DMP.

- *Public data sharing (How will potential users find out about your data? Are there any necessary limitations to protect sensitive data? On which repository, do you plan to share your data?)*

  Definition of the global strategy undertaken by the various LTMs. This sub-task will be precisely and contextually answered in each LTM's DMP.

- Intellectual property rights (IPR) and copyright (*Which licenses will be applied to the data? What restrictions apply to the reuse of third-party data? How to cite datasets?)*

It might be relevant to consider four levels of access, even if creatives commons licences should be applied to the most datasets, research data might be produced by private partners dealing with copyright or required specific embargo period, or their access might be limited due to sensitive data or competitive issues:

1. Confidential datasets (only accessible by a certain group of partners)
2. Restricted datasets (period of embargo or sensitive data)
3. Internal (data is shared only amongst TM's partners)
4. Open datasets (CC-0, CC-BY)

Need of specific IPR ownership agreement will be taken care of.

- *Data storage and preservation (What are the storage capacity and where will the data be stored? What are the back-up procedures? What procedures would be used to select data to be preserved? What file formats will be used for preservation?)*

The strategy will propose a set of minimal criteria that repositories have to fulfil to conform with the FAIR data principles (*set of guiding principles in order to make data findable, accessible, interoperable and reusable*) and ensure these are respected by the storage-relevant TM Infrastructure Components. This sub-task will be precisely answered within each LTM's DMP, as it is closely related to contextual element and choices. Examples of criteria to be further developed (also available on the "Practical Guide to the International Alignment of Research Data Management" (*Science Europe, November 2018*):

1. Datasets and single files within it should be given globally unique and persistent identifiers (e.g. DOI, ARK, etc.)
2. Upload of descriptive and project-specific metadata
3. Information regarding licence should be clearly stated and conform with TM guidelines (the user – uploader should be able to choose the one related to his/her dataset.
4. Citation information and metadata should be publicly accessible event in the case of datasets with restricted access.
5. Structured and machine-readable metadata.
6. Long-term preservation plan for archived data.

LTM Data Management Plan (DMP):

First assumption states that each LTM should provide a specific DMP, but as several PWTMLs might take place within an LTM, this first assumption's granularity should be further assessed. The DMP is

a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Building on the previously approved by each partner's RDM strategy, the elements to be described within a DMP are yet to be decided, however some sub-tasks could be focusing on:

- *Data collection and documentation (describing what data will be collected, observed, generated and reuses in the LTM context and the documentation associated with.)*
- *Ethics and security issues (describing potential ethical issues and how these will be addressed, introduce sensitive data and their corresponding risks)*
- *Data storage and preservation (describe particular storage infrastructure, which data will be archived for long-term and how)*
- *Open data sharing (explain where data will be openly available and when restriction applies, describe them).*

Implementation scale

The definition of the global strategy and its application by LTMs, will be managed through time. Therefore, several conformance stages with FAIR principles and research data guidelines could be used to monitor the process, e.g. scale defining six steps towards FAIRness (*Mons, 2018*).

A	B	C	D	E	F
Identification of re-useable data.	*Findable* Dataset identified with a persistent and persistent linked to the dataset identifying.	*FAIR metadata* Addition of metadata.	*FAIR data - restricted access* Restriction access are take care of.	*FAIR data - Open Access* Data made available under well-defined conditions for reuse.	*FAIR data – Open Access / Functionally Linked* Linked with other FAIR data with proper FAIR metadata.

TM RFC on Solidarity – How to select, align and finance the "redocumentation" project, for potential projects compatible with TM goals but left on a stand-by stage for a while, will be the main tasks of this RFC. What would it take (training, formation) to ensure the project scalability (this will be further developed according to **WP6** outcomes).

TM RFC on Top-down initiatives: Defining what local and national measures might contribute to the creation of LTMs, this would be further developed with outcomes from **WP7 Dissemination and promotion**.

Documentation - Action:

Incorporation Kit: Based on the starter kit, specific documentation will accompany LTM and PWTML during the incorporation process, with the goals to guarantee previous works integrity and conformance with TM rules and recommendations.

Specific organisational requirement:

Solidarity and Collaboration - In charge of the supervision of the collaboration protocols and solidarity of top-down initiatives. This organisational structure will organise events to foster internal and external collaboration amongst network's partners and external cultural-heritage networks and contribute to build relevant infrastructure for both ensuring knowledge transfer across LTM's, and collaboration with already existing networks. This office will also work closely with the **Know-How, Guidance office**, to ensure the alignment of previous initiatives within the TM network.

Research Data Management – Dedicated office to supervise implementation of Data Management Strategy, in charge of trainings and quality assurance of the whole process.

LTM / Smart Cluster

The European Commission conceived the 3S approach (*Smart: identify the region's own strengths and comparative assets, Specialised: prioritise research and innovation investment in competitive area, Strategic: define a shared vision for regional innovation*), characterised by its geographical place-based process. In this context several tools have been put in place, enabling willingly partners to gather and shape together the best innovative outputs to meet their needs. The process is described by the Commission as an "*inclusive process of stakeholder's involvement cantered on 'entrepreneurial discovery*'".

For LTM, it is therefore crucial to help future exploitation avenues to build on those principles, and therefore foster exchange of knowledge and collaboration amongst them. One of the duties of the LTM should be to ensure the creation of such a regional smart cluster, and support to the most its activities and needs. This achievement implies a close collaboration with PILLAR 3, dedicated to build relationships with Exploitations Avenues.

As the platforms and outputs produced aim to be related to regional or territorial particularities, regional and political stakeholders will be invited to collaborate within the *smart cluster*, helping participants to set up priorities, and setting in return the LTM as a key and strategic advisor for those decision makers.

A key benefit of the *smart cluster* is its strong potentiality to become a job's and opportunities' creator for both exploitation avenues and local digital industries, as most of its activities will engage practical outcomes.

TM RFC on Smart Cluster: Defining what would be the rules to be followed by the future *smart clusters* (for instance compliance with *LTM rules and recommendations*), what means will ensure the creation of such a space for creativity, support inter-disciplinary exchanges, political involvement and jobs creation, what relations could be built between the participants of the *smart clusters* and the partners of the LTM, how to monitor, evaluate and revise or update the process.

Specific offices:

Smart collaboration – Dedicated to the supervision of the creation of *smart cluster hubs* within the context of LTM, aiming to contribute to its reliability by ensuring management quality and compliance with the overall rules of the TMO, and sustain its growth by supervising contact with local authorities when requested.

Box 3.3-2: TM Requests for Comments (RFC) / Local Time Machine

- For LTM/TM's rules and recommendations:
  - TM RFC on data lifecycle
  - TM RFC on Vision Mission and Values Charter
  - TM RFC on Technical Charter
- For LTM/Framework
  - TM RFC on LTM/Framework
  - TM RFC on Training
- For LTM/Labelling System
  - TM RFC on Value Scale
- For LTM/Legal setting
  - TM RFC on Intellectual property rights and licences
- For LTM/Financial system
  - TM RFC on Franchise

- For LTM/Incorporation, collaboration and enhancement
  - TM RFC on Enhancing collaboration
  - TM RFC on Knowledge Transfer
  - TM RFC on Solidarity
  - TM RFC on Top-down initiatives
- For LTM / Smart Cluster
  - TM RFC on Smart Cluster

Box 3.3-3: Identified Offices / Local Time Machine

- For LTM/Framework
  - Supervision, project tracking
  - Know-How, Guidance
- For LTM/Labelling System
  - Supervision, certification
- For LTM/Legal setting
  - Legal
- For LTM/Financial system
  - Finance, Economic intelligence and Watch services
- For LTM/Incorporation, Collaboration and enhancement
  - Solidarity and Collaboration
  - Research Data Management
- For LTM/Smart Cluster
  Smart Collaboration

3.4 Milestones

Pillar 2 outputs are core to the development of the Time Machine, as they mostly set the basis for the global network. The four first years of the project have been identified as a necessary period to put the minimal components of the system in place. The table below aims to indicate when and in which order the identified RFC should take place during this "bootstrapping" period. The RFC method as grounding for all processes, will be defined firstly.

Number	RFC Title	Year Due
RFC1	TM Request for comments	1
RFC2	TM RFC on LTM/Framework	1
RFC3	TM RFC on Value Scale	1
RFC4	TM RFC on Definition of typologies	1
RFC5	TM RFC on Standardisation and homologation	1
RFC6	TM RFC on Open Hardware	1
RFC7	TM RFC on General Standards for the Super Computing Architecture	1
RFC8	TM RFC on Technical Charter	1
RFC9	TM RFC on data lifecycle	1
RFC10	TM RFC on Intellectual property rights and licenses	1
RFC11	TM RFC on Vision Mission and Values Charter	1
RFC12	TM RFC on Time Machine Box	1
RFC13	TM RFC on Synergy and interaction in EU Research Infrastructure	2
RFC14	TM RFC on Franchise System	2
RFC15	TM RFC on Training	2
RFC16	TM RFC on Distributed storage system for Public Data	2
RFC17	TM RFC on Distributed Storage system for Private Data	2
RFC18	TM RFC on Content Filtering	2
RFC19	TM RFC on on-demand digitisation	2
RFC20	TM RFC on Global optimization of digitisation process	2

Number	RFC Title	Year Due
RFC21	TM RFC on Digital Content Processor (DCP)	2
RFC22	TM RFC on TM Data Graph	3
RFC23	TM RFC on Large-Scale Inference Engine	3
RFC24	TM RFC on 4D Simulator	3
RFC25	TM RFC on Universal Representation Engine	3
RFC26	TM RFC on Virtual/Augmented Reality and Discovery	3
RFC27	TM RFC on Solidarity	4
RFC28	TM RFC on Enhancing Collaboration	4
RFC29	TM RFC on Knowledge transfer	4
RFC30	TM RFC on Top-Down initiatives	4
RFC31	TM RFC on Smart Cluster	4
RFC32	TM RFC on Collaboration indicators	4

Several time periods have been previously identified for Time Machine:

- Bootstrapping – 2021-2023
- Scaling – 2023-2025
- Sustaining – 2025-2027
- Globalising – 2027-2030

As the first period will mostly focus on the development and implementation of all RFCs, scaling will see the network and relevant operations taking larger proportions and sustaining mark the period when Time Machine should have reached its cruising speed. Over the 10-year perspective of this roadmap, several milestones will contribute to monitor the processes and ensure of their component's relevance, efficiency, conformance with TM objectives and contribute to required updates.

The below table provides an overview of what could be the milestones for pillar 2 roadmap (with a strong focus on *the Local Time Machine* Thematic Area) and should be adapted in regards of the other pillar outputs.

Phase	Milestone number	Milestone title	Due date (in year)	Due date (in month)	Means of verification
Bootstrapping	MS?	RFC Coordination	1	1	Each identified RFC have been assigned a specific coordinator by the RFC Committee.
	MS?	Framework: definition	1	6	Close of the TM RFC on LTM/Framework (e.g. typologies and routines, perimeter).
	MS?	Labelling system: definition	1	6	Close of the TM RFC on Value scale systems. V1 of the value scale systems for both LTM and PWTML.
	MS?	TM's Rules and Recommendations: Time Machine Box, development	1	8	Close of the TM RFC on Time Machine Box. Prototype V1.
	MS?	Framework: documentation	1	10	Close of the documentation and trainings phase for the implementation of the framework, including welcome guide, starter kit, LTM and PWTML's forms.

Phase	Milestone number	Milestone title	Due date (in year)	Due date (in month)	Means of verification
	MS?	TM's Rules and Recommendations and Legal settings: models	1	12	Close of TM RFC on Vision Mission and Values Charter, TM RFC on Technical Charter, TM RFC on data lifecycle, TM RFC on intellectual property rights and licenses, and contractual documents phase. V1 of the respective documents.
	MS?	TM's Rules and Recommendations: Time Machine Box, assessment	2	14	End of the evaluation of the technologies. When needed, new RFC.
	MS?	TM's Rules and Recommendations and Legal settings: roll-out	2	18	Close of the test phase for implementation with a representative percent of the TM partners aligning with its processes and prerequisites and making feedback to the development team.
	MS?	Labelling system: assessment	2	20	End of the evaluation of the system. When needed, new RFC.
	MS?	Financial system: definition	2	22	Close of the TM RFC on Franchise system, franchise V1. And close of the TM RFC on Training.
	MS?	TM's Rules and Recommendations and legal settings: adaptation	2	24	According to the result of the roll-out phase: Data selection Model, Data acquisition Model, Data Sharing Model, Vision Mission and Values Charter, Technical Charter, contractual documents, decision on opening a new RFC.
	MS?	Financial system: assessment	2	28	End of the evaluation of the system. When needed, new RFC.
	MS?	Standardisation of the TM network	3	30	Launch of the first 10 LTM officially complying with the TM's Framework and TM's Rules and Recommendations and Legal settings, advertisement and tracking on the TM's Operation Graph and on the TM' networks landing pages (use of the proper labels).
	MS?	Standardisation of the PWTMLs	3	34	Launch of the first 10 PWTMLs officially complying with the TM's Framework and TM's Rules and Recommendations and Legal settings, advertisement and tracking on the TM's Operation Graph and on the TM networks

Phase	Milestone number	Milestone title	Due date (in year)	Due date (in month)	Means of verification
					landing pages (use of the proper labels), payment of the proper franchise fee.
Scaling	MS?	Alignment	4	40	All TM partners now comply with TM's Rules and Recommendations and Legal settings (comprises alignment of previous initiatives).
	MS?	TM Cooperation	4	42	Close of the TM RFC on Solidarity, TM RFC on Enhancing collaboration, TM RFC on Knowledge transfer, TM RFC on Top-Down initiatives, TM RFC on Smart Cluster. V1 of the respective models.
	MS?	Community - Density growth	4	44	Close of the assessment of the network's growth performances in terms of partners actively involved within the TM, number of LTMs and PWTMLs launched and density of "rebuilding the past activities" in a geographical area. Definition of supportive top-down initiatives or new RFCs.
	MS?	Assessment	5	54	Close of the evaluation of TM's Framework, Labelling system, Rules and Recommendation, Legal settings, Financial system accuracy and suitability with TM needs. Decision on updating initiatives or new RFCs.
	MS?	Community - Density growth	5	56	Close of the assessment of the network's growth performances in terms of partners actively involved within the TM, number of LTMs and PWTMLs launched and density of "rebuilding the past activities" in a geographical area. Definition of supportive top-down initiatives or new RFCs.
	MS?	Assessment	6	66	Close of the evaluation of TM's Framework, Labelling system, Rules and Recommendation, Legal settings, Financial system accuracy and suitability with TM needs. Decision on updating initiatives or new RFCs.
	MS?	Community - Density growth	6	68	Close of the assessment of the network's growth performances in terms of partners actively involved within the TM, number of LTMs

Phase	Milestone number	Milestone title	Due date (in year)	Due date (in month)	Means of verification
					and PWTMLs launched and density of “rebuilding the past activities” in a geographical area. Definition of supportive top-down initiatives or new RFCs.
Sustaining	MS?	Assessment	7	78	Close of the evaluation of TM’s Framework, Labelling system, Rules and Recommendation, Legal settings, Financial system accuracy and suitability with TM needs. Decision on updating initiatives or new RFCs.
	MS?	Community - Density growth	7	80	Close of the assessment of the network’s growth performances in terms of partners actively involved within the TM, number of LTMs and PWTMLs launched and density of “rebuilding the past activities” in a geographical area. Definition of supportive top-down initiatives or new RFCs.
	MS?	Assessment	8	90	Close of the evaluation of TM’s Framework, Labelling system, Rules and Recommendation, Legal settings, Financial system accuracy and suitability with TM needs. Decision on updating initiatives or new RFCs.
	MS?	Community - Density growth	8	92	Close of the assessment of the network’s growth performances in terms of partners actively involved within the TM, number of LTMs and PWTMLs launched and density of “rebuilding the past activities” in a geographical area. Definition of supportive top-down initiatives or new RFCs.
	MS?	Assessment	9	102	Close of the evaluation of TM’s Framework, Labelling system, Rules and Recommendation, Legal settings, Financial system accuracy and suitability with TM needs. Decision on updating initiatives or new RFCs.
	MS?	Community - Density growth	9	104	Close of the assessment of the network’s growth performances in terms of partners actively involved within the TM, number of LTMs and PWTMLs launched and density of “rebuilding the past activities” in a geographical area.

Phase	Milestone number	Milestone title	Due date (in year)	Due date (in month)	Means of verification
					Definition of supportive top-down initiatives or new RFCs.
	MS?	Assessment	10	114	Close of the evaluation of TM's Framework, Labelling system, Rules and Recommendation, Legal settings, Financial system accuracy and suitability with TM needs. Decision on updating initiatives or new RFCs.
	MS?	Community - Density growth	10	116	Close of the assessment of the network's growth performances in terms of partners actively involved within the TM, number of LTMs and PWTMLs launched and density of "rebuilding the past activities" in a geographical area. Definition of supportive top-down initiatives or new RFCs.

3.5 Key Performance Indicators

As the targeted achievements of each task of pillar 2, although complementary, don't serve the same purposes, specific Key Performance Indicators will contribute to monitor each of them. However, one of the overall objectives of the Time Machine is to increase the amount of cultural-heritage digitised datasets available across EU, therefore previous monitoring studies conducted within the cultural-heritage sector (by Europeana) will be used and supported.

Europeana provides a baseline of statistical data indicators aiming to monitor the progress on digitisation of cultural heritage across Europe. Divided in four main themes (supply, demand, economics and sustainability), this global study "Enumerate" was last conducted in 2017 and would serve to precisely evaluate the impact of the TM, as it provides a clear state-of-the-art of the digitisation through Europa and defines key indicators such as:

- Presence of a digitisation strategy
- Presence of digital heritage collections
- Necessity to reproduce analogue heritage collection in digital format (per object type)
- Copyright conditions

infrastructure

- Number of units and relation in the TM Data Graph (measured by typologies in Bright and Dark Graphs).
- Number of rules in the Large-Scale Inference engines (measured by typologies of rules)
- Number of digitised images
- Number of digitised artefacts
- Number of digitised sites
- Number of linked resources

Communities

The different communities will have different impact on the TM and therefore will have different indicators for efficiency and impact in addition to a few general ones.

General

- Community outreach efforts (AMAs, Blogposts, Hackathons, Giveaways)

Developers

- Opened Issues/Commits/Forks/Pull Requests of Open Source Developers to TM Repositories
- Stars given to TM Repositories
- Known reuse of TM-Software and developed Standards

Educators

- Lectures / Courses given with use of TM Data

Professionals

- numbers of professional organisations that join the time machine organisation
- number of agreed policy measures about exchange of cultural heritage

Scholars

- Scientific articles using TM Data or about the TM

Volunteers

- Gamification of Crowdsourcing
- Members in the TM Portals

Local Time Machine

Collaboration:

- Average of partners involved per PWTML within an LTM

TM RFC on Collaboration indicators. Regular studies should be conducted to evaluate how collaboration is perceived by the different members and therefore measures to assess this qualitative process should be determined. Indicators will focus on evaluating following themes:

- Contentment of the various partners regarding collaboration
- Capacity of the collaboration in place to enhance the meet of project's goals
- Degree of partners' participation towards decision-making
- Perception of the different partners regarding collaboration
- Members participation's rate at project's stakeholders meeting
- Mix of line businesses involved in the project

Growth of the TM network:

- Number of new partners of the TMO (measured per membership's categories)
- Number of new LTM
- Number of new PWTML (in total)
- Activity of an LTM (number of PWTML launched over a period of time)
- Engagement rates: active contributors in the TM networks

Labelling system:

- Label adoption (number of labelled LTM and PWTML per label's categories)

Guidance and Training:

- How many new requests are coming in?
- Attendance: how many people are coming to special events-formations?
- Training completion, percentage rate.

4 Funding sources

Funding sources, as well as corresponding mechanisms and processes to be followed in the contractual relations with the different funders.

4.1 Infrastructure

Infrastructure is divided in two fields of activities:

- One dealing with public data (referred to as Great Commons)
- One dealing with private data (referred to as Private Time Machine)

Great Commons: **In-kind contributions** for both storage and computing powers. Institutions, industries, wishing to give servers' space running with the chosen distributed storage protocol or computing power for TMO technologies will be welcomed as members of the Time Machine Infrastructure Alliance. **Philanthropists** (e.g. states, private institutions) wishing to contribute to those Infrastructure components, will be given the opportunity to join the Time Machine Infrastructure Alliance through the TMO, in charge of acquiring such servers or computing powers in their names.

Private Time Machine: Industries, institutions that cannot expose all (some of them might still feed TM pipelines) their data under an open – creative commons licence yet, will benefit from TM infrastructures, on payment of the service's costs. As those data are probably more sensitive (e.g. private, financial, bank information), specific security means will need to be deployed (such as cryptography, authenticated access, erasing data right etc.). The TMO will encourage a dedicated and regional network of operators to provide required technical components and ensure security and quality criteria, and as the coordinator of this network, will perceived royalties.

Infrastructure will therefore be financed thanks to a collaboration amongst private and public partners.

4.2 Local Time Machine

Before becoming an official partner of an LTM, candidate institutions should become an official member of the TMO and therefore chose amongst memberships categories and their related fees (as defined in **WP6**) for the different services offered by the TMO. This financial system to take place between the TMO and the LTM will be further developed within the **TM RFC on Franchise** aiming to clarify usage infrastructure and services costs.

LTMs and PWTML will operate essentially with public and/or private funding that relates to regional development. An example of public funding is through the current European Structural and Investment Fund (ESIF) instruments, for which LTMs and PWTLMs offer concrete projects ideas with very focused development goals to be supported.

The PWTMLs are core to the operation phase of an LTM. All technical operations are meant to happened within a PWTML. When launching a PWTML all partners should ensure its financial sustainability and provide a financial plan, verified and validated by the TMO. They can be helped in the process by the TMO for identifying a funding strategy and in developing grant application.

The creation of *smart clusters* will enable the LTM to create job's opportunities within both local industries and exploitation avenues and contribute to the overall sustainability of the TM network.

5 Stakeholders to be involved

The large variety of stakeholders to be involved, their corresponding roles and, therefore, the needs in terms of management and coordination for the program.

Stakeholders	Areas of impact
External experts: Research Data Curator, Data architect Existing programmes partners of the TMO (Europeana, ICARUS) GLAM with on-going or finished digitisation process Partners of LTM involved in digitisation process -- External experts: Data architect, Data curator, metadata Specialists Representatives of partner's typologies (GLAM, research, operators, funding, communities)	**LTM: TM's Rules and Recommendations** Documents - Data selection model Data acquisition model Data publishing model -- Data sharing model
Representatives of TMO's partners	**LTM: TM's Rules and Recommendations** Vision Mission and Values Charter
Representatives of already existent LTM and PWTML initiatives Representatives of partner's typologies (GLAM, research, operators, funding, communities) TMO's experts: e.g. FIZ Karlsruhe, KNAW, Unibo	**LTM: TM's Rules and Recommendations** Technical Charter Time Machine Box
Representatives of already existent LTM and PWTML initiatives	**LTM: Framework** Partners typologies and routines Perimeter Starter Kit LTM and PWTML forms
Legal consultants from different countries Existing programmes partners of the TMO (Europeana, ICARUS) Representatives of already existent LTM and PWTML initiatives	**LTM: Legal setting** Contractual documents (e.g. Franchises) Intellectual property rights and licenses
Representatives of already existent LTM and PWTML initiatives External experts: Certification agencies	**LTM: Labelling system** Value Scale and associated labels for LTM and PWTML
Existing programmes partners of the TMO (DARIAH, ICARUS) External experts: Finance agencies Representatives of TMO's partners Representative of States, areas, municipalities (e.g. president, mayor, deputy, counsellor)	**LTM: Financial system** Franchise model **LTM: Smart Cluster** Shape of the future Cluster
Specialists of cognitive systems and large-scale inference engine. Ex: IBM Watson, Wolfram Alpha	**Infrastructure:** Large-Scale Inference Engine

5.1 Infrastructure

The large part of the infrastructure building will be done by the TM Infrastructure Alliance. Members of the alliances will share in-kind resources in computing and data storage. It is anticipated that the most important contribution will be done by universities and private companies. Large private company like Microsoft, Google or Amazon could participate in the Alliance and the writing of some core RFC. The current proposal offers a concrete way to approach them.

5.2 Community Management

The main problem is the fact that the various communities are either very diverse or disorganized or very country specific. Therefore, to reach the most important player in each field, a larger number of stakeholders will be needed than in other PILLARs. For instance, there are several important genealogy networks multiple European countries that will have to be contacted.

5.3 Local Time Machine

As one of the core sustainability strategy for TM is based on the LTM model, involved stakeholders will play a significant role in helping to precisely determine previously identified objectives and their related outcomes as listed in methodologies. Implication of stakeholders in each step, will help matching infrastructure constraints and partner's expectations and contribute to ease TM's developments and acceptation within diverse communities. Those stakeholders can be split in two categories: the larger composed of partners of TMO and the smaller regrouping external experts.

Partners of the TMO (e.g. representatives of LTM and PWTML initiatives, existing programmes partners, representatives of partner's typologies, TMO's experts) will be directly involved in the creation of listed documentation and the process of definition – clarification of identified needs and resolution of the TM RFCs. Each identified task will be conducted by a dedicated task-force. Coordination and follow-up should be carefully planned, and each task supervised by a chosen TMO member or a RFC's review board. Tools such as RFC, online shared documents or virtual meetings need to be carefully discussed and agreed by the task's partners.

External experts are invited to advise, and review documentation and decisions taken by each task-force, but final calls are not under their responsibilities.

As by definition an LTM is anchored within a geographical area, strategic stakeholders are their respective politic representatives (e.g. country, state, region, city, neighbourhood), who are likely to play an active part within an LTM as a member of the TMO, an external funding partner, or an advisor helping the LTM to growth and be shaped according to regional particularities. They will also be invited to play an active role within the creation and sustainability of the future *smart cluster*. Those diverse forms of collaboration and those shared interests, are meant to be strengthen over time and should set the LTM as a strategic partner for those decision makers. Politicians benefiting from the Big data of the Past to define their strategies regarding smart tourism, smart cities, land use and urban planning; and LTM becoming a trustworthy partner for the sustainability of a region.

6 Framework conditions

The framework conditions relating to policy, legal aspects and ethics that have to be considered.

Framework conditions	Proposed actions
FAIR Data management & Open access (European Commission H2020 Programme)	All partners should apply guidelines from the FAIR data management plan regarding: • the handling of research data during and after the end of the project • what data will be collected, processed and / or generated • which methodology and standards will be applied • whether data will be shared / made open access and • How data will be curated and preserved (including after the end of the project). The FAIR principle could be in some extend embedded in the TM components. Under Horizon 2020, each **beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results**. (*as stated in the Participant Portal H2020 Online Manual*)
EU General Data Protection Regulation (GDPR)	When PWTML requires to deal with private data, compliance with GDPR rules must be carefully put in place (*as stated in the 2018 reform of EU data protection rules)*. Privacy policies should be written in a clear straightforward language. The user will need to give an affirmative consent before his/her data can be used by a business. Silence is no consent.
Gender equality (European Commission H2020 Programme)	When writing a PWTML proposal, attention to gender equality need to be paid from different angles (*as stated in the Participant Portal H2020 Online Manual*): 1. Human resources: balance between women and men in the research teams who will implement the project 2. Content: analysing and taking into account the possible differences between men and women, boys and girls, or males and females, in the research and innovation content of the project.
Ethics Appraisal Procedure	Before launching a PWTML, an ethic self-assessment must be completed with extra care on the potential misuse of research results (research with a potential impact on human rights)
Landmark declaration of cooperation for advancing digitisation of cultural heritage	The declaration has three PILLARs of action: 1. A pan-European initiative for 3D digitisation of cultural heritage artefacts, monuments and sites; 2. Re-use of digitised cultural resources to foster citizen engagement, innovative use and spill-overs in other sectors; 3. Enhancing cross-sector and cross-border cooperation and capacity building in the sector of digitised cultural heritage. This Declaration coincide with TM goals and will help LTM and PWTML to find national supports.
European Framework for Action on Cultural Heritage	4 principles: 1. A **holistic approach**, looking at cultural heritage as a resource for the future and putting people at its heart;

Framework conditions	Proposed actions
	2. **Mainstreaming and integrated approach** across different EU policies; 3. **Evidence-based policy making**, including through cultural statistics; 4. **Multi-stakeholder cooperation**, encouraging the dialogue and exchange among a wide range of actors when designing and implementing cultural heritage policies and programmes. TM operations will be following the same principles as TM goals also implies multi-stakeholder cooperation, a holistic approach and is currently trying to find the best ways to answer to 2. and 3. needs.
Commission Decision (22.02.2019). Adopting Creative Commons as an open licence under the European Commission's reuse policy. Recast of the Public Sector Information (PSI) Directive (approved 04.04.2019)	The EU recently adopted CC BY 4.0 and CCO to share public documents including photos, videos, reports, peer-reviewed studies and data. The PSI regarding digitisation of cultural resources and licences: "That period should, however, be limited in time and as short as possible, in order to respect the principle that public domain material should stay in the public domain once it is digitised. The period of an exclusive right to digitise cultural resources should in general not exceed 10 years" "Any licences for the re-use of public sector information should in any event place as few restrictions on re-use as possible, for example limiting them to an indication of source." They are current studies aiming to foster use of open access licences (CCO or CC BY) amongst EU partners. The TMO should stay updated with those proposals as they serve the same goal as the TM: the use of open access licences within the TM network.
UNESCO Charter on digital heritage conservation (2004)	UNESCO offers some recommendations regarding digital long-term preservation issues, that will be tackled by TM infrastructure: (article 4) "Awareness-raising and advocacy is urgent […] sensitizing the general public to both the potential of the digital media and the practicalities of preservation". (article 5) "To preserve digital heritage, measures will need to be taken throughout the digital information life cycle, from creation to access." (article 7) " […] main criteria for deciding what digital material to keep would be their significance, and lasting cultural, scientific, evidential or other value." (article 11) "Preservation of the digital heritage requires sustained efforts on the part of governments, creators, publishers, relevant industries and heritage institutions".

7 Risks and Barriers

The approaches and measures that address any barriers to market entry and/or facilitate the commercial exploitation of research results.

Potential Risks and Barriers	Likeli-hood	Impact	Proposed risk-mitigation actions
There is a linguistic bias preventing some potential partners to join the network	Low	High	Connect with all linguistic communities and ensure the accessibility of project's material to the greatest number of partners. General exchanges should be made in English or in specific community's languages when requested by the project's needs. The multilingual nature of Europe and relevant European community's, needs to be taken into account when designing communication strategies. Some communications, for instance with developers, will be mainly in the English language, others will have to be organized/undertaken by native speakers.
Rules are too restrictive and prevent a great number of potential partners to join in	Medium	High	Rules and recommendations should be designed and approved by partner's representative with the aim of being the most inclusive. Their shape should be adaptable enough to prevent such issues and allow change over time.
Cooperation amongst LTMs partners is missed	Medium	Medium	Cooperation will be stated as a value within the Vision, Mission and Values Charter. Time and resources will be dedicated to team building: workshops, break-out sessions, excursions (to member's laboratory, libraries, archives). Mixed projects involving different typologies of partners will be encouraged with rewards (label or extra advertisement).
Lack of money during a PWTML process	Low	Medium	Funding plan is required to launch a PWTML and it will be closely monitored by the TMO to ensure sufficient funding. When the situation still happened, TMO with the help of the LTML would provide help for extra grants application.
An LTM is sued for non-respect of intellectual property rights	Low	Medium	Organise prevention sessions, and involve specific institutions used to deal with legal constraints in the process. Dedicate trainings materials. Offer external guidance (Legal Office).
Partners struggle to understand how to use TM technologies and TM components	Medium	Medium	Guidance and Trainings means adapted to member's need are developed, using different formats (e.g. tutorials, videos, online courses, guide). A specific guidance office replies to question.
Technical components on the partner's side are	Low	Low	A PWTML conformity-entrance brick focus on allowing conformity to technical TM's requirements. Time Machine Box provide for

Potential Risks and Barriers	Likeli-hood	Impact	Proposed risk-mitigation actions
preventing massive data extraction			data storing and encoding to the ones requesting it.
The quickly changing nature of various communities / communication technologies necessitates a constant engagement at the risk of not being aware of critical changes.	High	Medium	Special care needs to be taken to ensure up-to-date information on the most important partners/communities.
Ensure the activities are (and are perceived) as mutually beneficial to both the communities and the TM	Medium	Medium	Efforts, especially with volunteer communities need to be taken. The most dedicated contributors can be reached if the persons involved individually benefit from the engagement, especially in relation to professional communities. Contributions need to be made visible and transparent.

Annex A: Technical Charter – Draft

BACKGROUND

The Technical Charter aims to present the minimal requirements to be endorsed by each institution wishing to participate to the Time Machine operation.
Compliance with its content ensures a first level of standardisation for data and processes and sets the basics for the on-going development of the Time Machine.
In order to remain light and useable by the most, the charter encourages the use of universal and open interfaces and references that do not need central coordination.

DIGITAL IMAGES

IIIF

To be processed by the Time Machine distributed infrastructure and guarantee the greater interoperability, all digital images (already existing ones, or produced during a digitisation process) should be accessible through the IIIF protocol.
By supporting combination of content from diverse repositories and contributing to the growth of a market in compatible servers and viewing applications, this set of shared application programming interfaces (API), constitutes today's best answer to Time Machine interoperability and future accessibility needs.

Creative Commons license

To enable both future exploitation platforms to share, use and build upon those digital images and creators or owners of copyright- or database-protected images to waive interests in their works, CCO or CC-BY licences should be used.
CCO is equivalent to public domain and required by wikidata.
CC-BY forces users to give appropriate credit, provide a link to the license and indicate if changes were made.

REFERENCES

Wikidata

Named entities (person, location, organisation, product), topics, objects and concepts should be referred when possible using Wikidata entity ID for items (commonly known as WikidataID).
To prevent at the most ambiguity and clarify the context of each items, their respective relations should be precisely documented using Wikidata entity ID for properties.
When necessaries and according to the Wikidata rules, the Wikidata base can be manually extended.

UNIVERSALLY UNIQUE IDENTIFIER

For entities, massively extracted (e.g. names in almanacs, ID of places in cadastres, etc.) a UUID should be given. This means that every partner can generate its own UUID without central coordination. Online tools exist to generate these UUID. When possible Wikidata properties should be used to refer to relation between entities.

SOFTWARE CODE

Open Source code for Time Machine component is recommended but not obligatory at this stage of the project.

OTHER

Some elements are currently not treated in this version of the Technical Charter.

> Hardware
> 3D
> Sound
> Musical Score
> Web archiving

Annex B: Vision, Mission and Values Charter

(Not defined yet)

Annex C: Starter Kits for LTM and PWTML

Starter Kit for a Local Time Machine

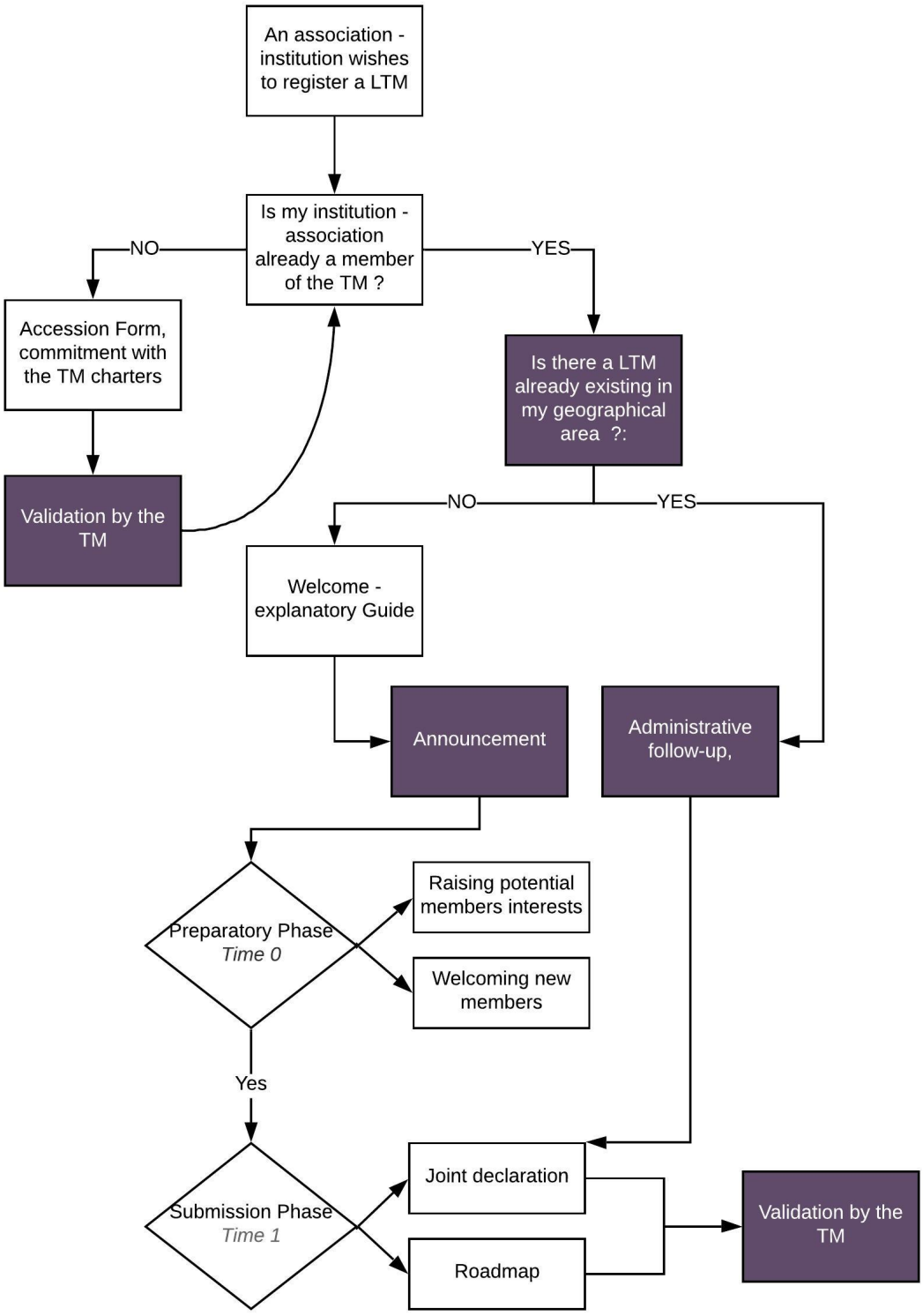

Figure C-1: Actions for the launch of an LTM

Starter Kit for a Project with Time Machine Label

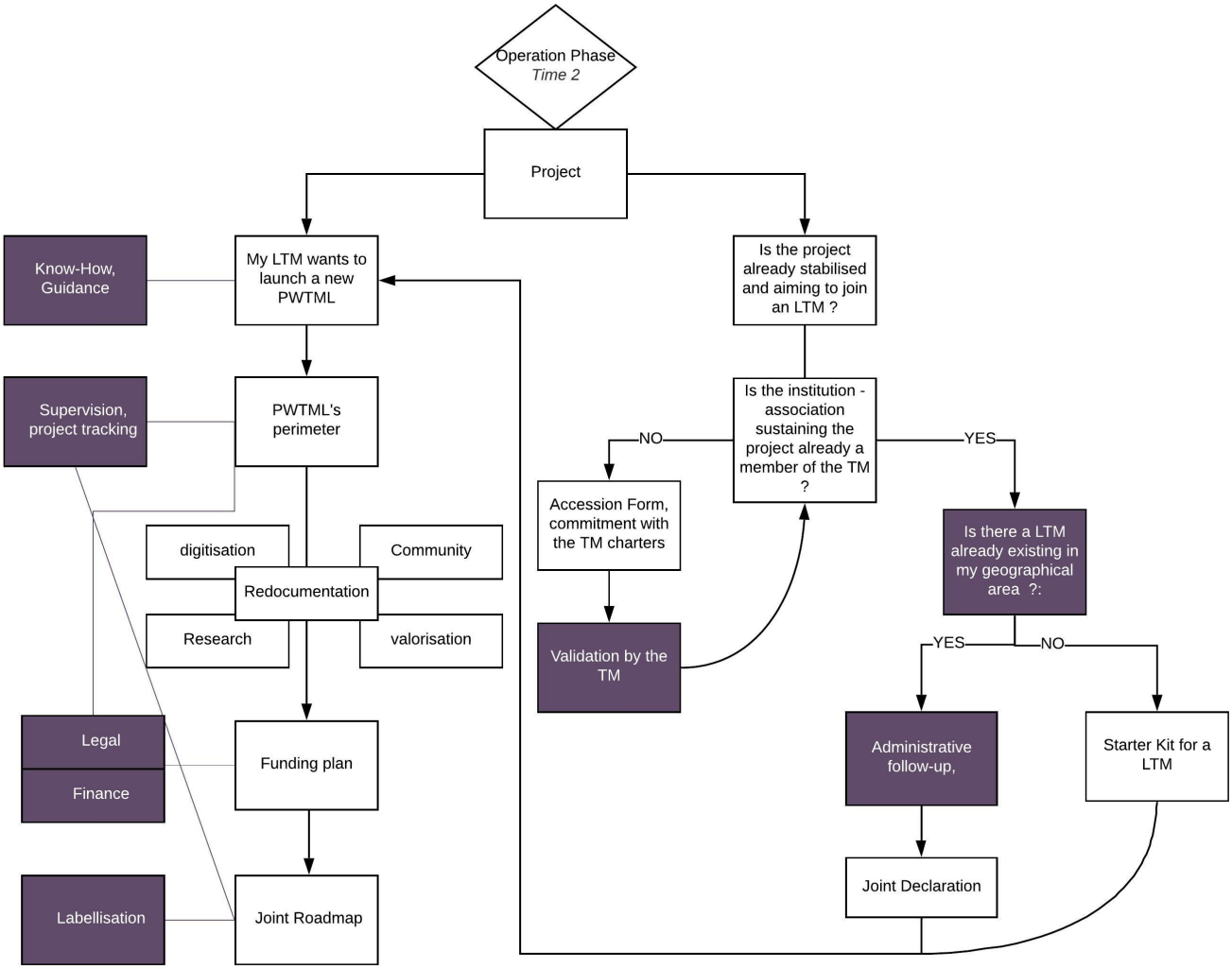

Figure C-2: Actions for the launch of a PWTML