TRAME. *Text and manuscript transmission of the Middle Ages in Europe.*

Evolving the system towards Horizon2020 and VCMS[1] challenges

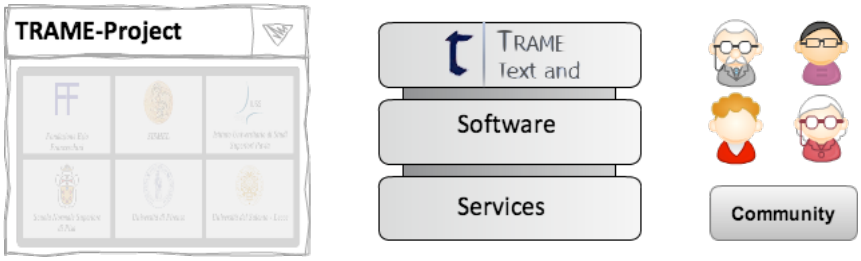
The TRAME project. A short description

TRAME[2] is a research infrastructure project focused on promoting interoperability among different digital resources available in the medieval digital ecosystem domain by connecting repositories of digitized images of medieval manuscripts, their codicological descriptions, their textual and philological interest, their cultural significance in the context of the european history. Currently it implements a number of features (including simple, shelfmark, advanced search modes etc.) on more than 80 selected scholarly digital resources around western medieval manuscripts, authors, and texts across EU and USA, including - digital libraries, research databases and other projects from leading institutions, such as:

University of Califonia Los Angeles - Catalogue of Digitized Medieval Manuscripts	Bodleian Library - Manuscripts Catalog
ICCU - Manus on line. Censimento dei manoscritti delle biblioteche italiane	Société des Bollandistes - BHL, Biblioteca Hagiographica Latina manuscripta
Scriptorium - Bulletin codicologique	AIMD - Manoscritti Datati d'Italia
Universitat de Barcelona - Ramon Llull Database	SISMEL - MIRABILE, Digital Archive for Medieval Culture
IRHT - Medium: gestion des reproductions de manuscrits	IRHT - Jonas : Répertoire des textes et des manuscrits médiévaux d'oc et d'oïl
University of California Berkeley - Digital Scriptorium	E-codices - Virtual Manuscript Library of Switzerland
Biblioteca Medicea Laurenziana - Plutei Digitali	Biblioteca Apostolica Vaticana - Catalogo dei manoscritti
Bibliothèque nationale de France - Mandragore, base des manuscrits enluminés	Bibliothèque nationale de France - Gallica, manuscrits
Biblioteca Apostolica Vaticana, Manoscritti Digitalizzati	

[1] *Virtual Center for Medieval Studies*, see infra
[2] http://trame.fefonlus.it

TRAME and the research community: COST Action IS1005 Medieval Europe



TRAME is more than just a piece of software: it is a research tool deeply rooted in the international medieval scholarly community, whose development is in line with the Memorandum of Understanding of the COST Action IS1005 "Medieval Europe - Medieval Cultures and Technological Resources"[3] , representing 260 researchers coming from 39 leading institutions (archives, libraries, universities and research centers) in 24 countries across the EU.

Medieval Europe - Medieval Cultures and Technological Resources - IS1005 coordinates existing research on the use of ICT for Medieval Studies, a field where ICT has potential to improve and substantiate research, and specifically on interoperability between databases (DBs) used for storage and research in this area. *Medieval Europe - Medieval Cultures and Technological Resources* aims to link modern technology to medieval studies and combines technological development with intensive training of researchers in order to connect the wealth of scientific material available, while ensuring interoperability, data security, and suitability for differing academic contexts.

TRAME has been selected by the *Medieval Europe - Medieval Cultures and Technological Resources* community as one of the cores for the upcoming VCMS, *Virtual Center for Medieval Studies*, a digital environment for scholarly research designed by COST Action IS1005 Work Package 4 (focused on the VCMS design) according to Horizon2020 principles (cfr. topic: *e-Infrastructures for virtual research environments*, VRE) to *support capacity building in interdisciplinary research communities and empower researchers through development of service-driven digital research environments, services and tools tailored to their specific needs, foster cross-disciplinary data interoperability, allowing data citation, promoting data sharing and trust.*

[3] http://www.cost.eu/domains_actions/isch/Actions/IS1005

Project goals: enhance TRAME to be ready for Horizon2020 challenges

According to EINFRA 9 description[4], the VRE (VCMS) *will be composed of generic services delivered by e-infrastructures and domain specific services co-developed and co-operated by researchers, technology and e-infrastructure providers and should clearly identify and build on requirements from real use case:*

- *integration of heterogeneous data from multiple sources*
- *value-added services for computing*
- *data exploration*
- *mining and visualisation*

Where data are concerned, projects will define:

- *the semantics, ontologies, the 'what' metadata*
- *the best computing models and levels of abstraction (e.g. by means of open web services) to process the rich semantics at machine level (the so called 'how' metadata)*

The EINFRA 9 description suggest to ***re-use tools and services*** *from existing infrastructures and projects at national and/or European level as appropriate in order to* ***accelerate innovation in research***.

We feel that TRAME is already a strong candidate to meet the above requirements for the upcoming COST IS1005 VCMS, but needs to be technologically improved - according to a clear development agenda, described later in this document - to allow **integrated access** to potentially unlimited digital **research resources** and **tools** across disciplines and user communities according to the Horizon2020 e-Infrastructures for virtual research environments (VRE) specific challenge.

The **TRAME Roadmap (2011-2014)** and **Future Workplan (2014-)** paragraphs are describing the development agenda for TRAME to move towards the Horizon2020 requirements as described above. The **TRAME development agenda implementation** and **Expected costs** paragraphs are focusing on the resources needed in order to improve TRAME to become a suitable candidate for building an Horizon2020 VRE Call, such as the above EINFRA9.

[4] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2144-einfra-9-2015.html

TRAME history and development

TRAME's first step was funded by the italian Ministry of Education, University and Research (FIRB program)[5], it then becomes an international project supported by COST Action IS 1005. Furthermore, TRAME has been recently evaluated as a relevant tool for the development of the medieval section of the CENDARI Digital Research Infrastructure[6]

The actual development of TRAME (started in 2011) is seeking to face the lack of standardisation of the Medieval Digital Ecosystem by:

- connecting research data around medieval manuscripts coming from authoritative sources (on the web)
- fostering the dialogue between research and other communities (libraries, archives etc.)
- providing a tool for complex scholarly research needs, seeking aggregation, discoverability and reuse of scholarly research data

The TRAME Roadmap (2011-2014)

TRAME started in 2011 as a meta-search tool focused on the medieval western manuscript tradition, but gradually evolved in complex set of tools and services to address other specific needs coming from the international research community. The TRAME development roadmap for the years 2011-14 was:

step 1: to build a metasearch engine capable to search across the databases of the italian partners that joined the starting phase of the project (achieved, 2011) and of the international partners that joined since 2012 (achieved, 2013)

step 2: to extend the metasearch approach to other web resources (libraries, portals, individual research projects), using various tools and technologies (achieved, 2014):

- complete TRAME code refactoring, using a Model-View-Controller approach and Object Oriented programming environment for extensibility and sustainability. The main difference between the two versions of TRAME is that we can reuse the most part of the application's logic when implementing new sites and functionalities;
- OAI harvester and mapper implementation;
- authority lists for shelfmarks establishment and integration;
- improved advanced search;
- geographic shelfmark search;

[5] cfr. E. Degl'Innocenti «Trame: Building a Meta-Search Tool for the Study of Medieval Literary Traditions» in EVA 2011 - Proceedings. Vito Cappellini - James Hemsley (eds.), Bologna, Pitagora, 2011, pp. 94-9
[6] http://www.cendari.eu

FUTURE WORKPLAN (2014-)

TRAME is designed to be a digital research tool for scholars doing research in medieval western culture. Its aim is to support users in answering actual research questions, rather than simply offering traditional searching and browsing features. As described it currently offers advanced meta-search tools working on a vast number of scholarly resources (DBs, Digital libraries, etc.) coming from libraries, archives and research institutes in Italy, EU and the USA, covering various disciplines among medieval studies (literary studies, paleography, codicology, philology, philosophy etc.). The main aim of the future TRAME development is to:

- implement a Semantic Knowledge Base, built using hi-quality data (coming from SISMEL/FEF and other leading international research institutes);
- growing thanks to advanced Named Entity Recognition (NER) and Extraction techniques + semantic annotation and User Generated Content (UGC);
- consolidating due to domain experts validation processes over extracted knowledge and extending with rule-based reasoning techniques, to make hidden information emerge.

step 3: to extend the metasearch approach to virtually any relevant web resource devoted to medieval manuscripts, texts and authors, through an extended partnership program (in progress, 2014)

- authority list for medieval authors' names establishment and integration
- TRAME extension for authors and texts

step 4: to implement semantic technologies and tools to create a medieval digital knowledge space for scholarly research. Parallel development of the metasearch and the semantic approach (*from data to knowledge*, forthcoming).

In addition to the on-the-fly metasearch feature, the TRAME(S) evolution will implement:

- a semantic knowledge base with trusted data coming from SISMEL/FEF and other international partners
- a workflow to gradually increase the KB without needing extensive human interaction with the content holding institutions, using Named Entity Recognition (NER) processes and domain experts validation (see diagram *TRAME Development towards Horizon2020 and VCMS*): For each TRAME query an indipendent (a-synchronous) process will be activated in order to crawl all the relevant resources matching the user queries and extract knowledge (names, roles, facts, etc.), according to a given set of rules. The NER extraction process will create a new layer of knowledge, stored in a temporary area of the Medieval Knowledge Base, waiting for human validation (i.e.: add, edit or remove a fact) to be added to the KB.
- a set of rule-based reasoning tools, to make implicit (i.e.: not expressed) knowledge emerge from hi-quality (i.e.: validated) data

TRAME development agenda implementation

According to the above development agenda, in order to achieve the goals stated in step **3** and **4** we'll leverage on the competencies and ICT infrastructures already present on the SISMEL and FEF Digital Humanities, Dept. under the coordination of Dr. Emiliano Degl'Innocenti[7].

The correspondent TRAME development workflow could be represented by the following activities:

1. sites integration into TRAME:
    a. building a comprehensive directory of Medieval scholarly resources candidates for integration in TRAME
    b. technical analisys of each site
    c. data and metadata structures mapping
    d. software implementation;
2. authors and shelfmarks authority lists further development and integration;
3. semantic knowledge base further implementation:
    a. research and build an ontological framework to model and link the dispersed resources for Medieval scholarship, allow semantically-rich and -disambiguated ways to organize knowledge, allow accurate and intelligent ways for the **access of knowledge**, through ontology query languages;
    b. complete data triplification from existing SISMEL/FEF DBs;
4. NER and Rule based reasoning implementation;
    a. use Natural Language Processing (NLP) functionality for entity recognition, extraction and disambiguation within the content
    b. implement rule based reasoners in order to reveal potential implicit relations and/or find inconsistent knowledge within the content;

[7] Full CV available at: http://www.linkedin.com/in/degli